Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews ••• (••••) •••–•••

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Review

Creativity, information, and consciousness: The information dynamics of thinking

Geraint A. Wiggins [a,b,*]

[a] *Computational Creativity Lab, AI Lab, Vrije Universiteit Brussel, Belgium*
[b] *School of Electronic Engineering and Computer Science, Queen Mary University of London, UK*

Received 8 January 2018; received in revised form 5 March 2018; accepted 4 May 2018

Communicated by L. Perlovsky

Abstract

This paper presents a theory of the basic operation of mind, Information Dynamics of Thinking, which is intended for computational implementation and thence empirical testing. It is based on the information theory of Shannon, and treats the mind/brain as an information processing organ that aims to be information-efficient, in that it predicts its world, so as to use information efficiently, and regularly re-represents it, so as to store information efficiently. The theory is presented in context of a background review of various research areas that impinge upon its development. Consequences of the theory and testable hypotheses arising from it are discussed.

Keywords: Cognitive architecture; Computational creativity; Information theory; Machine consciousness; Music cognition; Cognitive representation

1. Introduction

This paper overviews a unified account of several aspects of non-conscious cognitive processing, memory, and their relationship with conscious awareness. The account is at a mathematical level, completely removed from biological considerations, but its components and processes are motivated by empirical observation, and, in most cases, at least indirectly empirically supported. Where direct extant support is lacking, it is the focus of current research.

The aim is to cut away the foliage of detail that obscures the trunk of the tree of cognition, and to study broad, observable, fundamental cognitive behaviours, deferring study of detail until after the main framework is understood. Two aspects of the methodology go beyond modelling of straightforwardly observed specific cognitive process: one is the inclusion of life-long learning (that is, an account of how learning over a long period affects cognition, and *vice versa*); the other is the attempt to explain *why* the proposed cognitive functions might have evolved. This latter point

* Correspondence to: Computational Creativity Lab, AI Lab, Vrije Universiteit Brussel, Verdieping 3, Pleinlaan 9, 1050 Brussel, Belgium.
E-mail address: geraint.wiggins@vub.be.

https://doi.org/10.1016/j.plrev.2018.05.001
1571-0645/

must, of course, be taken with the salt of caution, because just-so stories [101] are untestable, unscientific argument. However, a cognitive account whose evolutionary development *lacks* potential explanation is no less unsatisfying. An equally strong second tenet is not merely to build a model which *describes* a behaviour, but to build one which *explains* that behaviour at a well-defined level of abstraction; in other words, the aim is to describe *mechanism*, and not merely *effect*, but not necessarily at the level of neurons. Thus, an *explanatory* model at a particular *level of abstraction*[1] is developed [224].

The following sections knit together theories from artificial intelligence and cognitive science covering creativity studies, cognitive musicology, information theory and cognitive architecture, drawing on formal theories of knowledge representation, learning, statistical linguistics, distributional semantics, and audio processing. Visual processing is omitted because it would require a complete review of its own, and because auditory processing is the original inspiration for the model; however, a rich seam of research remains to be mined in the visual area.

Andy Clark [37, p. 200] writes,

> . . . an action-oriented predictive processing framework is not so much revolutionary as it is reassuringly integrative. Its greatest value lies in suggesting a set of deep unifying principles for understanding multiple aspects of neural function and organization. It does this by describing an architecture capable of combining high-level knowledge and low-level (sensory) information in ways that systematically deal with uncertainty, ambiguity, and noise.

This paper lays out the theory of exactly such an integrative predictive-processing model, reassuring or otherwise. The model is currently the subject of experimental programming, in an implementation series called *Information Dynamics of Thinking*,[2] or IDyOT for short [224,215,65,203]. The word "thinking" in the acronym relates directly to Turing's use of the word "think," in his seminal paper "Can machines think?" [197], and carries with it the same rich and expensive baggage of questions. This review lays out the context for the theory of IDyOT, and then summarises the model, giving novel information about its memory mechanism, and identifying points where empirical studies might support or refute the proposal along the way.

The exposition begins with creativity studies, which provided the original motivation for the current research. Next, the primary source of inspiration, the study of the cognition of music, is introduced, both from the perspective of the current work and as a powerful general tool for the study of cognition. Some crucial concepts from information theory are then presented, and a summary of cognitive architecture research is given. The conceptual spaces theory of Gärdenfors [72,73] is introduced. Finally, these components are brought together to describe a cognitive-architectural framework that is focused on the problem of identifying and processing sequences of perceived events, and the regulation of attention thereto. The phenomenon of chunking, which is fundamental to the cognitive processing of complex signals [76], and its relationship with memory are considered. The chunks produced by the model are implicated in the formation of cognitive representations of signals and sequences, on a statistical basis, which in turn provide a statistical model to inform the chunking mechanism, in a mechanism that perhaps instantiates the *chunk-and-pass* approach of Christiansen and Chater [36]. This statistical model extends to multi-layer, hierarchical representations of semiotic form, which provide an alternative explanation for language from the traditional Chomskian view [35] and its more modern counterparts [61]: trees top out in semantic structures, not in syntactic categories, and certainly not in sentential forms. Thus, the theory affords an alternative kind of deep learning [cf., e.g., 121], in which sequence and information efficiency are explicit drivers.

The mechanism proposed here is entirely neutral with respect to the stimuli to which the model might be exposed: its operation is expressed in terms only of the statistics of those stimuli and their internal representations. Thus, it constitutes a parsimonious approach to cognitive science; the suggestion is not that all cognition is driven by one neural assembly implementing this mechanism, but by one key neural process, so that multiple neural assemblies, armed with modality-specific inputs supplied by modality-specific sensory organs, process different data in the same way, but with modality-sensitive outcomes. For the avoidance of doubt: this is certainly not to claim that there are no other ancillary cognitive processes; however, the aim is to identify the core processes at the chosen level of abstraction that other processes underpin or arise from. Furthermore, the representations used by the model are themselves inferred

[1] Related to, but not to be confused with *levels of description* [133,224].

[2] This implementation, and the reasoning behind it, rests heavily on my long-term collaborations with Marcus Pearce and Jamie Forth, and more recent ones with Matt Purver and Frank van der Velde, to whom I am deeply indebted.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

from the data [cf., e.g., 121], according to well-defined principles of *information-theoretic efficiency*; this means that it is essentially a *predictive coding* mechanism [71]. The proposed mechanism constitutes a testable hypothetical model of perceptual sequence memory and processing in humans and other higher animals: certain observable features of humans are potentially explainable as epiphenomena of the model. Finally, the theory affords an account of creative ideation, returning to the search for a model of creativity that initially motivated the work.

2. Creative cognition and computational creativity

The source of inspiration for the IDyOT model was not the attempt to design a cognitive architecture, but the search for computational creativity [25,41]. This route to fruition has consequences for the philosophy of the work, and therefore computational creativity is the first topic in this review. This research field, which began its work in earnest in the late 1990s, is a branch of artificial intelligence (AI) in which the focus is:

> The philosophy, science and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative. [41, §1]

It is important to understand that such behaviours are deemed to include creative endeavours that are not traditionally artistic and musical, such as mathematics: one of the early successes of the field was the inclusion of several mathematical structures, created by a program, in the encyclopaedic Journal of Integer Sequences [40,39]. As in wider AI, there is a broad multidimensional spectrum of interest. As in wider AI, most researchers are focused on specific tasks (for example, the generation of music [e.g., 154,142], paintings [e.g., 136], poetry [e.g., 75], jokes [e.g., 14] or narrative [e.g., 162]), and some are focused on broader creative mechanisms (for example, creative language use [e.g., 206] or new reasoning methods [e.g., 11]). Some, such as the current author, are interested in the generic frameworks within which such tasks take place, and how to evaluate them [214,219,168,161]. As, again, in wider AI, some computational creativity researchers are interested in human-like creativity, and some aim to build machines that are creative in their own terms; the current author is in the former group.

One way to view computational creativity is as an attempt to break free from the problem-solving approaches that became the mainstream of AI in the 1980s and '90s, in which the problem of intelligence is mostly studied in a strongly reductionist way, by focus on very specific problems or on very specific methods, and then subsequently from the tendency to reduce the field to optimisation, in machine learning. In this, computational creativity is similar to the field of Artificial General Intelligence (AGI), and, indeed, an annual workshop[3] and a recent publication unite the two [12]. In computational creativity work, often there is no "solution" and even no "problem"—which requires a very different methodology from problem-solving AI. It is natural, therefore, that the search for a model of human creativity would lead, as here, in the direction of cognitive architecture, since cognitive architecture figures prominently in AGI [e.g., 53,54].

The psychological study of creativity is not particularly rich in proposals that lend themselves to computational modelling, notwithstanding their value in other respects; a fuller summary than the following is given by Wiggins et al. [231]. Guilford [79] proposes a two-stage process, first of divergence, and then of convergence, but the theory is too general to be implemented without substantial elaboration. Csikszentmihalyi [47] proposes a notion of "creative flow", which describes mental states at a level not immediately amenable to computational study.

The theory of Wallas [210], however, is somewhat more specific. It identifies four steps in creativity:

Preparation: The creative goal is identified and considered;
Incubation: Conscious attempts at creativity are not made, but non-conscious effort may be applied;
Illumination: The moment when an idea appears in conscious awareness—often called "the 'Aha!' moment";
Verification: The new idea is applied.

Wallas' theory entails an important distinction, between conscious, or deliberate, creativity and non-conscious, or spontaneous, creativity. The former of these is the creativity where, for example, a professional designer must produce

[3] Details of the Computational Creativity, Concept Invention, and General Intelligence (C3GI) workshop may be found at http://cogsci.uni-osnabrueck.de/~c3gi/.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

a product in too short a time to wait for inspiration: she consciously applies rules of her craft as necessary. The latter is the creativity where an idea or concept appears in one's awareness, apparently without bidding, effort or intention. Wallas' ideas correspond directly with different parts of the IDyOT model, and will feature below.

Boden [19,18,17] proposes a much more computationally tractable approach. While her notion of *conceptual space*,[4] a theoretical space inhabited by concepts which may or may not have been encountered by a creator, was not intended to be the same as an AI search space, it is possible to use the ideas in this way, to helpful ends, and Wiggins [214,219] has done so. Boden makes a distinction between *exploratory* and *transformational* creativity, the former being, in her terms, everyday creativity, as found in, perhaps, designing a chair, and the latter being, respectively, the act of defining an entirely new category of furniture. Wiggins [214] demonstrates that transformational creativity is in fact exploratory creativity in the (Boden) conceptual space of (Boden) conceptual spaces, demonstrating the importance of reflection in creative work.

Boden [18] also introduces *combinatorial* creativity,[5] in which concepts from more than one conceptual space are brought together. Notwithstanding different terminology, this idea is indistinguishable from two other notions in the creativity literature: *bisociation*, due to Koestler [105], and *conceptual blending* of Turner and Fauconnier [198].

Schmidhuber [e.g., 55,179,178] gives an implementable mathematical theory of learning which affords creativity. However, the work founders on an emphasis on value judgement, perhaps borrowed from the philosophy of the Romantics. In essence, Schmidhuber argues that compressibility, in terms of an observer's compression-based learning algorithm, is the key factor that creates value in so-called "great" art; but this "greatness" is a problematic notion in the post-modern era. The IDyOT model accords with Schmidhuber's in several ways, however: in particular, his notion of *curiosity* (a concept also studied by Saunders [175]) related to information content, in the sense of Shannon [181], and *compression* as a fundamental motivation for cognitive process, are related to IDyOT's modus operandi, as will be seen below.

This issue of "great" creativity arises in Boden's work, too, in two ways. First, Boden argues that her transformational creativity corresponds with that of great creators, while exploratory creativity is exhibited by more prosaic workers. However, things are not this simple [219]: while it is certainly the case that some great creators transform the space in which they work (for example: Schoenberg's twelve-note music; Picasso's cubism; Duchamp's anti-art, which became dada; Utzon's Sydney Opera House), some others continue existing traditions (for example, Mozart continued and perfected the style that Haydn introduced, and Rennie MacIntosh developed the British Art Nouveau from Morris' Arts and Crafts movement via Horta and the Jugendstil). In the current work, these subjective distinctions are not useful. Second, Boden makes a distinction between *H-creativity* and *P-creativity*: creativity in historical (or wider, social) and psychological (or narrower, personal) terms, respectively. This distinction is about context: a drawing by a small child may be deemed very P-creative, but would not often be deemed H-creative enough to hang in an art gallery. However, Boden's simple dichotomy is not enough: Wiggins et al. [231] argue that creative value is function of four arguments: the creator, the observer, the created artefact, and the context in which it was created and observed. Fortunately, this complicated method of evaluation will be unnecessary for the current paper, though it will be important in engineering future versions of IDyOT. For the moment, we are interested only in IDyOT's personal creativity, which is readily identified by reference to its encoded knowledge [168].

The aim of the work presented here is to identify a detailed, testable, computational account of creativity in the human mind. Since the work is founded in music, it is appropriate to end this section on creativity with a quote attributed to one of humanity's lasting creators, Wolfgang Amadeus Mozart.

> 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.
>
> All this fires my soul, and provided I am not disturbed, my subject enlarges itself, becomes methodized and defined, and the whole, though it be long, stands almost completed and finished in my mind, so that I can survey

[4] Confusingly, this notion is different from the Conceptual Spaces of Gärdenfors [72] which are also important to the current proposal: see §5. Since the Gärdenfors theory is more central here, unqualified use of the term "conceptual space" should be taken to refer to his theory.

[5] This is often misquoted as "combinational".

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

> it, like a fine picture or a beautiful statue, at a glance. Nor do I hear in my imagination the parts successively, but I hear them, as it were, all at once. What a delight this is I cannot tell! All this inventing, this producing takes place in a pleasing lively dream. ... What has been thus produced I do not easily forget, and this is perhaps the best gift I have my Divine Maker to thank for. [90, pp. 317–8]

Mozart's extraordinary memory for music is attested elsewhere, and he is arguably right that this is his most important gift: contemporaneous others were capable of similar harmony, but the ability to hold entire compositions in memory and manipulate them as a whole is a considerable advantage to a composer. IDyOT theory is intended to account for exactly the creative experience that Mozart described. The next section continues the current musical theme, into music cognition research

3. Music in cognitive science research

3.1. Five properties that make music unique in psychology

The current proposal for a cognitive architecture is inspired, and arguably could only have been inspired, by the study of music cognition, and musical creativity in particular. Therefore, it is appropriate briefly to discuss relevant work in this research area.

Music is, in some sense, a Cinderella in cognitive science research. For example, Fournié [66], quoted by Lashley [119], writes

> Speech is the only window through which the physiologist can view the cerebral life. [119, p. 113]

Fournié's claim is false: music is an alternative, and a good one, at that. While music is self-evidently a cognitive entity (as well as a social, physiological, historical and fundamentally human one) [221,217,225,227,228,48], cognitive science often fails to engage with it as the extraordinary source of cognitive insight that it can be. While there is a healthy community studying the cognitive science of music,[6] crossover from musical to non-musical cognitive science is regrettably small.[7]

Music is universal in human society, though, in some cultures, the word used to name it is also synonymous with "dance". It serves diverse and often very important societal purposes; in the West, only the most austere rituals, associated with law, do not include it. Further, every known society has music. Some cultures even have elements that wish to suppress music for religious reasons. This is possibly one of the strongest arguments for the importance of music to humans that there is: it is so enchanting as to be seen as the work of evil.

Music (no matter what kind) is unique in human endeavour, for the following reasons:

1. It is *ephemeral*: music does not exist as an object in the world,[8] and is therefore *entirely* dependent on the mechanisms of memory for its cognitive effect [221,217];
2. It is *anepistemic*: except in very unusual circumstances, essentially equating with onomatopoeia in language, and notwithstanding evidence of association [104], music is without denotational meaning, being incapable of making statements that are truth-functional [213,221];
3. It is *autoanaphoric*: music can refer, but in a different way from language, because, except in the same unusual circumstances as above, it always refers to itself, and usually only within a single piece [173,145,146,220,222];
4. It is *cultural*: music (like language) is a cultural artefact and requires enculturation to be understood [225,227, 228]: in consequence (like language), it is heavily dependent on learning;

[6] See http://icmpc.org; also the journals *Music Perception*, *Psychology of Music* and *Musicae Scientiae*.

[7] It is common for music cognition papers to be rejected from general psychology journals on the grounds that they are "about music" and therefore ungeneral. Nothing could be further from the truth, because the same ears and brain are used for music as are used for language, a key research domain in cognitive science, and there is evidence of shared resource [151]. To dismiss this relationship is miss the point of cognitive generality.

[8] Here, it is important not to confuse *music* with *music notation*: while the latter evidently exists as part of a world object, it constitutes only instructions on how to produce music, and not the music itself. A similar usage is common when referring to recordings, and the corresponding response applies.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

5. It is *enchanting*: music engages humans in strong, non-conscious and conscious motor and/or affective responses, often to the extent that they react both physically, and often involuntarily [e.g., 20,57,203], and also by spending their hard-earned cash on recordings.

Of course, spoken language shares some of these properties, but not all of them. In particular, the study of language is burdened with the still-insuperable problems of real-world reference and societal semantic context, whose richness render the whole extremely complicated, and rather hard to address via reductionist science. Music, on the other hand, is anepistemic, and only autoanaphoric, making it a closed system, aside from the onomatopoeic effects mentioned above, which can be safely isolated as unusual outliers; in most cases, music's cultural meaning (e.g., association with wedding or funeral rite) is sufficiently non-specific to be unproblematic in analysis. Therefore, from a scientific perspective, music can be studied purely from the perspective of perception, representation, cognition and memory, free of semantic context and the overbearing burden of every-day reasoning. In these senses, it is more straightforward to access some aspects of music cognition empirically, than other cognitive phenomena—and many of these aspects share behaviour, and therefore possibly mechanism and resource, with language [151].

3.2. Music demands better models

Conversely, in a different and interesting sense, music also has the benefit of being *more* complicated to study than language, because it is inherently and *necessarily* multidimensional, in a way that language is not. While a speech communication stream does contain several dimensions (phoneme, pitch, loudness, timing, etc.), the phonemic element, and the structures built thereon (words, etc., up to semantics), are so dominant that both communication and research can, and often do, proceed through it alone: were it not so, this paper would be unreadable. This dominance has resulted, in some research areas, in technologies which are restrictive, such as the unidimensional, low-order Markov models that are often used in linguistic cognitive science. In music, such simplicity simply will not work, because a significant amount of the cognitive effect of the music resides in the interaction between the various dimensions, and so overly reductive approaches invalidate the stimulus. Sometimes it is necessary to develop special evaluation techniques to enable this study: for example, where it is impossible to construct stimuli that are simple enough to test a particular hypothesis, while still free of ambiguity [160].

An important example of how this multidimensional challenge has improved modelling technology is to be found in the seminal work of Conklin and Witten [42], who extended the unidimensional, variable order models of Witten's earlier linguistics work into multidimensional predictors, primarily for the purpose of automated composition of musical melody. Subsequently, Pearce [157] has shown that Conklin's multi-dimensional method, with some adjustments [155], models human music perception (specifically, of melody) extremely well [158], and it has been taken back into linguistics, demonstrating that the same class of multidimensional models may benefit linguistics more than their simpler ancestors [226]. The broad class of multidimensional musical models has been found effective in various creative musical tasks [149,85,211,88]. It lies in strong contrast with approaches to musical structure borrowed from Chomskian linguistics, where top-down structural governance is deemed central [124] and where the statistical regularities captured by Markov models are deemed external constraints [170]. This relationship is considered further in an evolutionary context by Rohrmeier et al. [171].

IDyOT theory draws inspiration from Conklin's multi-dimensional Markovian approach, but extends it to long-term dependency (a shortcoming of non-hierarchical Markov models), and transcends it by providing a hypothetical account of why it works as a cognitive model. While the thinking originally arose from music, the hypothesis is that the ideas can transfer directly to language or any other sequential cognitive process [215]. To be accessible to readers unfamiliar with musical taxonomy and vocabulary, language will be the primary source of examples when introducing IDyOT in §7; nevertheless, the argument in the current section needs to be motivated from music: every attempt has been made to render it accessible to readers who are not musically expert.

3.3. What music similarity tells us about memory

Music's ephemerality enforces its sequentiality, and this has consequences for more than just the local structure captured by Markov models. Comparison of one musical item with another—that is, the perception of *music similarity*—is necessarily sequential, because two stimuli played at once simply become one (probably cacophonous and

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

incomprehensible) stimulus. This entails the development of specific techniques to study music similarity [2]. More importantly, it means that a human listener can *never* directly compare two pieces of music: a piece of music can only ever be compared with the *memory* of another. This is a fundamentally different situation from, for example, the visual comparison of two shapes, side by side, where it is possible to flick back and forth over the detail. Of course, visual stimuli could be presented sequentially to create the same effect, but the enforced sequential nature of music means that the stimuli are still very different. The first immediate consequence of this enforced sequentiality is the effect of *priming*, in which the perceptual and cognitive processing of a first stimulus affect the perception and cognition of a second; this leads to an *unavoidable* case of the asymmetry of similarity judgements noted by Tversky [199]. The second consequence is that, in a strong sense, the study of musical similarity corresponds with the study of musical memory [220]. Adding to this the anepistemic and autoanaphoric nature of music, one may reasonably start from the hypothesis that music similarity corresponds directly with structural similarity in memory (whatever that may mean), because there is no denotation or reference to add anything else; individual associations (such as two songs reminding an individual of a particular concert, and thus being "similar" in that weak sense) can safely be ignored, because they will be statistically irrelevant in the behaviour of groups that can nevertheless agree on the structure and function of music. Thus, there is reason to hypothesise that the experience of music similarity lays bare significantly more of the function of memory than would similarity in other domains. Furthermore, in many societies, quasi-formal music theory[9] exists to describe what is agreed about the style [225], giving a ready-made folk-psychology toolkit for constructing hypotheses and stimuli. One could, of course, construct visual equivalents, perhaps as abstract animated movies. However, to enable the degree of detailed empirical probe that is possible in music, observers would need an equivalent degree of enculturation in abstract animation. This is not available; even if the necessary materials did exist, abstract visual stimuli do not have the same enchanting quality as music, and therefore it is not inevitable that comparable learning, and, hence, comparable memory effects, would arise.

One section, phrase, or entire piece of music can refer to another, preceding one, by "sounding like" it—that is, by being in some perceptual or cognitive respect, similar. Much of the time, this autoanaphoric structure will not be consciously noticed as such: instead, a listener will understand the reference as, for example, "the second verse", or say that "the melody comes back", or "this reminds me of that". Many studies have looked at similarity on various levels of scale, from individual phrases to whole songs, with a view to helping prospective listeners find music that they will want to spend money on.[10] However, many such engineering studies omit the cognitive element of the problem, reducing it to audio signal processing, with the top-down, predictive element of music perception and cognition omitted from the models, leading to what is sometimes problematised—problematically—as a "semantic gap" [221]. Wiggins [220] argues that similarity in music (and other perceptual domains) should be studied as a function of memory, rather than as a function of the domain.

Deliège's theory of Cue Abstraction [49], closely related to Ruwet's (musicological) method of paradigmatic analysis [173], addresses the cognitive element and its contribution to musical structure directly, arguing that the fundamental process of music listening is one of identifying *cues* which constitute salient[11] units: themes, musical gestures, even down to just a few notes. Cues may be indexed, by repetition, or by the appearance of similar structures later in the piece. As a listener hears each cue and each index, she constructs a mental data representation[12] called

[9] It is important to understand that music *theory* is not the same kind of construct as a scientific theory. Rather, it is a set of vocabulary and rules that describe atomic constructs in music (e.g., note) and how to assemble them into structures (e.g., chord, phrase) that accord with whatever musical style is under consideration [225].

[10] See http://ismir.net for many high-quality publications on this topic from an engineering perspective. The music cognition journal *Musicae Scientiae* has also published a special issue on this topic [193], covering mostly empirical research.

[11] The word *salient* is problematic in psychology. If one asks for a definition, one is often told it means "important in context". On pursuing the question, by asking what "importance in context" is, one often hears that it is the same as "salience". IDyOT theory aims to explicate at least part of what perceptual "salience" means.

[12] The word *representation* is problematic in the current context, because its meaning to computer scientists is different from its meaning to psychologists. In both fields, a representation (cognitive or computational) is a description of a thing, concept or state of affairs; but in computer science, the term is also refers to the scheme or language used to express that description—for example, a consistent set of objects and predicates capable of describing some set of things is termed "a representation", even before anyone has attempted to describe anything specific using it. Because, in computational terms, such a representation is given semantics by the inference rules associated with it, these rules are usually included in the meaning of the word as well [21,22]. The current author is primarily a computer scientist, and therefore uses the term in the latter sense. Where it is necessary to make a distinction, "data representation", "represented data", or similar, is written.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

an *imprint*, which in some sense summarises the set of structures (cue plus indices) involved; a consequence of this is that each index can inform later indices about what would be a recognisable similarity, leading to evolution of the cue through the piece [146,44]. To account for such strong context-dependence, priming must be dynamically present within the model, and any symbolic representation must be expressive enough to represent the formation of the imprint as time progresses.

In music cognition research, as elsewhere in psychology, success in understanding perceptual similarity has been achieved by modelling the behaviour of musical percepts and structures by means of low-dimensional geometrical spaces, such as that of pitch [183], and numerous related proposals (implicitly or explicity geometrical) exist to explicate Western tonal harmony [130,131,10,126,127,125,33]. A geometrical space has even been proposed to capture the full complexity of musical metre and rhythm [63]. These mathematical structures, theorised as Conceptual Spaces by Gärdenfors [72,73] (see §5) contain discrete regions which afford symbolic labelling, similarly to the more widely-known spaces of colour [e.g., 51] and vowel sounds [e.g., 108], and share with them the interesting property that, given at least one definitive prototype in a region, other points in the same region can be described relative to it; thus, effectively two representations are simultaneously active: a gross discrete one, modulated by a detailed continuous one. In music, however, the detailed modulation is particularly useful. The example of musical pitch is especially interesting in this context. Modern Western keyboard instruments are usually tuned in such a way that the perceptual distance (*interval*) between any two contiguous pitches is the same, a pitch difference of one *semitone*, equivalent to a frequency factor of $\sqrt[12]{2}$, because there are twelve possible pitch divisions in the commonest Western scales. This tuning is called *equal temperament*. It does not reflect the more natural tuning used by competent players of fretless stringed instruments, or wind instruments, which is relative to the central tone (*tonic*) of the piece being played,[13] and in which the semitones between contiguous notes are not all the same size. So in this sense,[14] a freshly tuned piano, tuned in equal temperament, is slightly out of tune—yet most listeners tolerate the compromise without even noticing it, even though the difference in sound between chords played in different tunings is substantial and easily heard when they are juxtaposed. All this is due to categorical perception, which affords two advantages to a listener: first, variance from the norm can be tolerated, and thus approximate similarity can be recognised; and, second, categorical, discrete representations are easier to store and access (in the particular sense of information theory, explained below) than continuous ones [134]. The first advantage is also afforded in phoneme space, when listening to speech by different people or in different accents [83,82]; and the same effect is used in video compression, to reduce the colour palette of the signal in tolerable ways.

3.4. What musical pitch tells us about cognitive representation

An important aspect of musical memory research is the diachronic development of music perception in individual humans. Evidence [174] suggests that Western neonatal infants have the ability to recall *absolute pitch* (a percept logarithmically related to the frequency of a musical tone), but lose conscious access to it in the first few months of life, in favour of *relative pitch*, defined between successive pitches; this is perhaps accounted for by the need for pre-verbal infants to understand the emotional content of verbal gestures of both parents (the same patterns of up and down, but in adult male and female pitch ranges) and infant siblings (in a different pitch range again). Nevertheless, some people retain absolute pitch hearing throughout life, with the ability to name notes, and to produce specific pitches; furthermore, evidence [128] suggests that adults who cannot name absolute pitches retain absolute pitch memory implicitly, and can reproduce memorised pitch, though they do not have deliberate access to labels. Why should it be so?

Given constant amplitude, musical pitch may be thought of as a straight line between the pitch corresponding with around 44 Hz, and that corresponding with around 22 kHz. The pitch percept changes linearly with exponential change in frequency. This change gives us the idea of *pitch height*: a note is *higher* if it is nearer the high frequency end, which corresponds with the right hand end of a piano keyboard, or the sound of a small animal, while *lower* tones are leftward on the keyboard, or more like the sound of most large animals (whales being a notable exception

[13] Some pieces do not have tonics, and in this context, the ideal is often equal temperament.

[14] In fact, the octaves on a piano are tuned slightly large, because this gives a more harmonious sound; the effect is called “stretch”. This, however, is a different issue from temperament.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

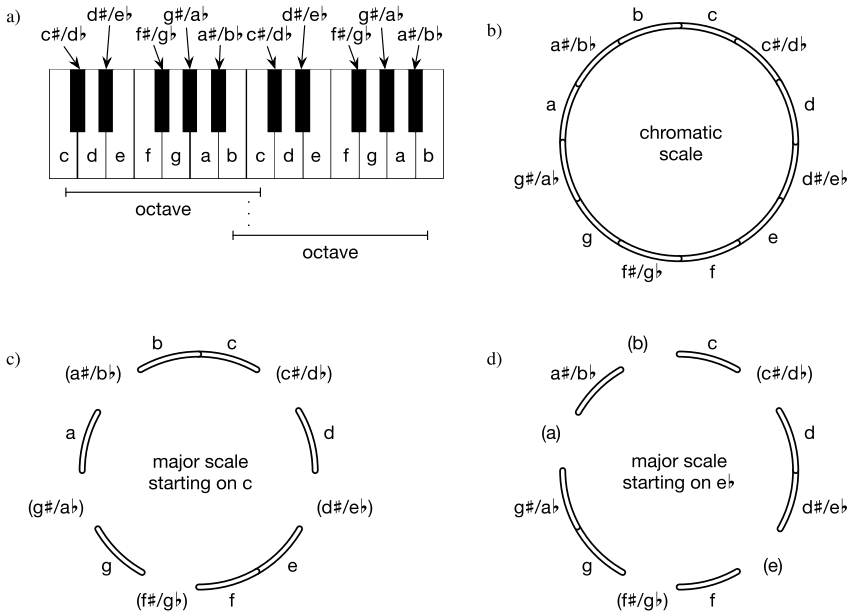

Fig. 1. Three musical scales, represented in the pitch spiral [183], projected onto the $x-z$ *chroma* plane, losing the octave dimension. a) The equivalent piano keyboard, for two octaves (i.e., going twice round the circle, from c). b) The chromatic scale. c) The major scale starting on c. d) The major scale starting on e♭. Notes associated with white keys are labelled with letters a…g, and the black keys in between are associated with notes named after the white note below, *sharp*, ♯, or the one above, *flat*, ♭. The reason for this alternative nomenclature is beyond the scope of the current discussion. The chromatic scale includes all twelve semitones; the other two scales contain the subset $\{0, 2, 4, 5, 7, 9, 11\}$. Observe the asymmetry of the pattern of permitted semitones in the major scales, and note that one is de facto a rotation of the other; any major scale may be generated in this way. The initial note is the *tonic*. The asymmetry of the scale allows a competent listener to infer the tonic relative to the rest of the scale.

because of the way they use pitch to communicate [231]). This pitch line affords absolute pitch perception with about 109 pitches at equally tempered semitone intervals; this can be represented in a minimum of 7 bits.

However, there are other ways to compare pitches. The first is most likely afforded by the physiological structure of the ear and its logarithmic match to the harmonic series that makes up all but a vanishingly small percentage of pitched natural sounds (and, of paramount importance, the human voice) [143]: the percept of an *octave* occurs when the frequency ratio of two distinct tones is an integer power of two, so the fundamental frequency of one pitch matches either the second harmonic of the other, or that same relation applies recursively more than once, but within audible range. This means that there is a cyclic structure implicit in the harmonic series that ears seem to have evolved to hear, and the one-dimensional line of pitch space can now be looped into a vertical helix, embedded in a three-dimensional space [183] where the y axis is the number of octaves between pitches. Moving round the helix moves from one pitch category to the next. Thus, there is a perceptual sense, *octave equivalence*, in which the first and last notes of the sequence (or *chromatic scale*) formed by playing the 12 contiguous notes in either direction from a given pitch (see the keyboard in Fig. 1) are substantially more similar to each other than they are to any other note in the scale. However, this perception of similarity increases with musical sophistication, and the percept is therefore, to some (as yet unknown) degree learned [182,183,106,107]. Some researchers view this spiral model as a line in 3-d space, but this is not strictly correct, because the other points in the 3-d space that contains it are undefined in terms of pitch: the pitch space is actually the spiral itself, not the points off it, and that it is *embedded* in a 3-d space affords the metric described above. Tenenbaum et al. [97,96,192] suggest that cognitive structures such as lines, spirals and trees may be innate, the best one for representing a particular range of concepts being statistically selected by the process of learning from the data at hand. The learning of octave equivalence perception fits this proposal extremely well.

Because of the asymmetry induced by the logarithmic relationship between frequency and pitch, the most useful musical scales in most musical cultures are themselves asymmetrical: they are defined as subsets of the chromatic scale introduced above. Perhaps the most common in the West is the *major* scale, probably because of the way it matches well with the natural harmonic series. It selects 7 of the 12 chromatic pitches, numbered relatively from zero

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

(and we can start at any pitch), as follows: {0, 2, 4, 5, 7, 9, 11}. This is illustrated, in comparison with the chromatic scale, in Fig. 1.

The structure of these scales, the pitch spiral, and relative pitch hearing affords an opportunity for efficiency, as follows. Because relative pitch allows us to hear any piece of music relative to its tonic note, and the asymmetry of the scale allows us to identify that note, then any major scale can be represented by just its tonic, and any note in the scale can be identified by an offset from the tonic between 0 and 6 inclusive. McAdams and Saariaho [134] argue that a small number of discrete distinguishable perceptual categories and a small number of elements to be remembered are probably important for any perceptual dimension that is to carry some aspect of form [135]: two dimensions of 7 categories each seems to conform to this, in comparison with one dimension of 109 categories. The vast majority of pitch systems known to musicologists use octave equivalence.

The point of this argument is that there is every reason to suppose that the nature of the pitch percept is defined *ab initio* by our basic auditory capacity, but that further, in some sense better, representations may be inferred, given data from a musical practice that conforms to, and is presumably partly driven by, the requirements of the representations. Thus, a naïve listener is stimulated to develop more sophisticated representations, better suited to the data, by listening; any composer (presumably a sophisticated listener) who creates new music is also likely to been enculturated with the same representations. So a tight feedback loop is hypothesised, between the affordances of cognitive representations and the musical structures that are created to use them.

This phenomenon is not merely theoretical. A musician with relatively little experience, particularly if she sings, or plays instruments such as orchestral strings, that require the player to take responsibility for tuning each note, will quickly learn to hear the position in the scale (*scale degree*) as a quality of tone,[15] assuming she knows what the tonic is: it is not necessary to hear both pitches at once. A competent musician can imagine a tonic pitch (even without knowledge of the corresponding frequency) and then hear subsequent pitches as scale degrees relative to it. Having made this representational change, where the relative pitch is primary, the representation is reduced to just 7 categories, representing an information saving of more than 57% on the absolute representation for an individual pitch percept. Finally, the most advanced representation is one that combines the interval with the scale degree from which it started: to a very advanced listener, each given interval will have a different subjective sound, depending on where it starts in the scale, even in an equally tempered performance, *where there is objectively no difference*.

The principles that seem to apply in individual musical pitch representation seem also to apply in musical harmony, in which we consider groups of simultaneous pitches (*chords*), as opposed to single ones. It is important to understand that a chord is usually perceived as a perceptual atom, and that picking out the pitches that make a chord is a skill that must be learned, though there is a clear and regular relationship between the combinations of notes and the perceived quality of the chords. Western musical harmony, as mentioned above, can be wrapped into a spiral embedded in 3-space [33], or around the surface of a torus [126,127], in a way exactly analogous with the pitch space, with exactly the same effect of increased efficiency of representation, and perceptual equivalence under pitch change (*transposition*). Its perception must, too, be learned. Thus, invariances (thought of as aspects of a stimulus that make no difference to musical similarity) are fundamental in determining the nature of cognitive musical representations [123]: the cognitive process appears to seek out the best way of representing the data, given its structure. All this neatly follows the sequence of learning that we expect and actively reinforce in our musical children.

The key feature of Pearce's work [157,1] on IDyOM (Information Dynamics of Music) is the system's ability to predict the next pitch in a melody, taking into account probabilities entailed not only by the preceding pitch sequence, but also by dimensions other than pitch [159]. These are combined into a single distribution, weighted inversely by the entropy of the respective source distributions—thus, distributions containing more information contribute more [157,155]. The system learns, unsupervised, from a corpus of music represented symbolically, in basic dimensions of pitch and time; some more advanced musical information, such as the tonic of the scale being used, is also supplied [157,158]. From this data, further derived dimensions are added to the list from which the system can choose: for example, successive pitches can be subtracted to yield a representation in terms of musical *intervals* (jumps between pitches), and pitch can also be expressed in terms of scale degree, as above in Fig. 1. Dimensions can be combined, too, producing new dimensions whose alphabet is the cross product of the originals; this allows dimensions that are (perhaps partially) correlated to be appropriately expressed [155,84]. Having produced these various dimensions, it

[15] Regrettably, as with other similar learned percepts, the *qualia* (nature of the experience) of hearing this is indescribable to someone who has not learned to do it.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

is possible to choose the ones amongst them that produce the optimal model according to an information criterion: the average number of bits per symbol required to represent the corpus is minimised, by a simple search process—in other words, the representation is chosen that gives the best statistical fit of the model to the corpus. Any dimensions that do not improve the representation in this sense are removed. The best fit was achieved with a representation space including not absolute or relative pitch, but a combination of scale degree and interval [157,158]. What is more, the best fitting model is the one that best predicts human expectations. Crucially, the scale degree/interval combination is the representation that accords best with music theory, corresponding with the perception of the advanced listener, described above. So the program picks the representation that matches with an advanced musician. This result is achieved bottom-up from the statistics of the data, via the single heuristic of *information efficiency*, with no programming of the preference whatsoever.

3.5. What music tells us about how we understand sequence in the world

Another important property of music's sequentiality, and one that it shares with language, is that it affords *perceptual chunking* in time [76]. Indeed, even relatively simple music contains hierarchical grouping, which has informed music analysis techniques[16] over the past 100 years [177,124], though the debate between top-down and bottom-up approaches continues. It is possible to give an analysis of a piece of music as a tree [124], which corresponds closely in form and intent with the parse tree produced by analysis of language—with the major omission of syntactic categories [15]. Despite this very significant difference, exactly the same bottom-up statistical techniques predict human judgements of musical segmentation as do so for language segmentation [160,113,77], which probably tells us something about a general process implicated in both. Since both the phenomena share segmentation, but only one has syntactic categories, it is reasonable to hypothesise that segmentation is the more general property.

The cues of Deliège's Cue Abstraction theory [50] and the salient items of Ruwet's Paradigmatic Analysis technique [173], are rarely single notes, but usually sequences. It is reasonable to suppose that these sequences must be contained within perceptual chunks, since memory for sequences that cross chunk boundaries is impaired [191,34], and therefore recognition would not work well. This implies that representation of chunks works in such a way as to facilitate local similarity judgements. Musical pitch and time for individual notes can be represented straightforwardly using the spatial approaches mentioned above [e.g., 216], but sequences are more problematic: they can be represented as (mathematical) sequences [e.g., 184], or as multi-point trajectories through the pitch/time space.

However the representation is implemented, it must be accessible hierarchically, and the elements of the hierarchy must be accessible as units, simply because this is what musicians do all the time: zooming effortlessly in and out of levels of detail, in discussing and rehearsing music. Further, the similarity relation must be defined between them, and a high degree of similarity must be taken as equivalence and lead to unification. It is this property that leads to the bottom-up, hierarchical learning procedure that is proposed for IDyOT by Wiggins and Forth [215] and van der Velde et al. [203] and summarised in §7.

3.6. What music tells us about prediction: what to expect; when to expect it

Huron [92] argues that expectation is a key driving force of musical experience, and Pearce's work adds evidence for his argument; and these ideas broadly concur with wider cognitive-scientific models [e.g., 192]. Huron [92] makes a convincing case that the feeling of uncertainty (at a meta-level with respect to expectation), which can vary in a very short time, accounts for much of the detailed aesthetic experience of music: the increase and decrease of tension due to changes in uncertainty resolving into expected certainty, or denial of expectation, is what musicians sometimes call the "ebb and flow" of the music.

Anticipation of what is coming next (followed by confirmation or denial and the concomitant affect) is only one aspect of this. Another key aspect is the synchronisation that allows groups of humans to perform music together, in perfect time, in ways which have never been demonstrated in other species. Humans are able to pick

[16] In which music analysts attempt to explicate in music-theoretic terms how a piece of music works.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

up and anticipate a beat, in a phenomenon known as *entrainment* [60,65], and many find it difficult *not* to do so, when presented with music that they find engaging. Entrainment is studied extensively in the music cognition literature, along with timing and rhythm [e.g., 46,45,60,152,138,167,129]. While other species, particularly those capable of vocal imitation, exhibit temporary entrainment to music when encouraged to do so [153,176], and others, such as crickets and fireflies, exhibit synchronisation via reflex response [e.g., 81], spontaneous and sustained self-motivated active entrainment seems to be reserved to humans. Grahn [78] gives a thorough survey of related research.

IDyOM [157] is very good at modelling human predictions of *what* will happen, and does take into account information about timing of its input sequences; and it can predict *when* the next event will happen in a statistically satisfactory way [204], but there is no real-time element to the model. To play music requires the ability to judge precisely where in time an action should be placed, anticipating the exact moment with motor preparation so that the sound is synchronised. The synchronisation involved can be extremely complicated, with irregular rhythmic structures spanning cycles of several seconds, as in, for example, Greek folk or Indian classical music, or with simultaneous multiple levels of synchronisation at different speeds and with very subtle deviations from the beat which are highly musically significant, as in, for example, funk and rap. The potential for entrainment, then, is extremely advanced in humans, though it needs development to reach these levels of precision.

As mentioned above, entrainment capacity in animals seems to correlate with the capacity for vocal imitation [176]. From this, one might reasonably hypothesise that entrainment is related to the process of vocal imitation. Since vocal imitation is implicated in learning to speak [187], and cannot be done without some kind of speech perception (even if without semantic association), maybe entrainment can be related to all these things. A reason for it to be so would, again, be cognitive efficiency. Attending to speech, as to anything else, is energetically expensive. If periods of attention can be appropriately timed, by predicting when the next unit of information from an interlocutor will appear, the efficiency of attending is optimised [118]. There is a debate as to whether such attentional control is achieved by oscillators or by interval timers; Grahn [78] presents evidence for timer-based control, while Large and Jones [118] argue for oscillators. Further, from this perspective, shared entrainment would be a necessary feature of effective sustained conversation. Finally, a positive hedonic response would be selected for by evolution, because it would promote conversational bonding, and perhaps cooperative musical behaviour, in the early human organism, which needed a strong social group to survive. Thus, the human tendency to enjoy musical entrainment may be explained; a similar hierarchical system of *beats* for linguistic synchronisation is a given in phonology [e.g., 83]. A new definition of entrainment, improved over others in the literature, and a hypothetical underpinning mechanism, it is proposed by Forth et al. [65].

3.7. Music and emotion

For many people, the primary function of music is to stimulate internal affective experience. Juslin and Sloboda [95] supply encyclopaedic coverage of research in this area, and further approaches are reported in a special issue [112]. However, an evolutionary understanding of the affective correlates of music processing is the subject of open debate: Pinker [164] argues that it is without explanation, Huron [92] argues that music generates affective responses by manipulating listeners' expectations (in a proposal very much in line with the current one), and Perlovsky [163] argues that it supplies an emotionally healing alternative to the more cognitively dissonant semantics of language. Fortunately, this difficult debate is not relevant to IDyOT, which operates without simulation of emotional response, at a rather more fundamental perceptual level.

3.8. Summary

The richly endowed mine of music research, coupled with some reasoning from statistical language processing, has motivated the majority of IDyOT theory, supplying components for a process based on statistical anticipation (entailing sequential, hierarchical memory), perceptual similarity (enabling category formation and motivating the synthesis of the representation of meaning) and hence entailing concept formation and affording sequence generation. These last two are both aspects of creativity. Below, these components will be assembled into the design for the IDyOT cognitive architecture. First, however, it is necessary to outline Information Theory.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

4. Information theory

4.1. Information efficiency

The notion of *information efficiency* introduced above is borrowed from the Information Theory of Shannon [181]. According to Alan Turing,

> Shannon wants to feed not just data to a Brain, but cultural things! He wants to play music to it!
>
> (Alan Turing, quoted by Hodges [89, p. 251])

Presumably, Shannon was aware of the pattern element of music, and so hypothesised, correctly, that its structure could be captured by a Turing machine (a "Brain" in Turing's terminology). It is satisfying, therefore, to be using Shannon Information Theory in IDyOT, since IDyOT grew out of research on music cognition.

A fuller summary of the Information Theory background to IDyOT (and IDyOM) is given by Pearce and Wiggins [159], and full detail is given by Pearce [157]. Two applications of the theory are relevant here, the first in context of sequential behaviour, and the second in context of categorisation.

IDyOT theory takes the stance, uncontroversial in cognitive science [91], that a key function of a mind/brain is to process information so as to assist the organism that surrounds it in surviving, and that a successful mind/brain will do so as efficiently as possible. Shannon's mathematics, originally intended to measure the information efficiency of digital signals in wires, supplies the tools that allow this principle to be programmed into the model [181].

4.2. Information theory applied to sequential structures

Information may be passed as a message from a source to a sink encoded in sequences of *bits*, or binary integers. The bit is used because it is the smallest possible quantum of information: it simply says "on" or "off", or "yes" or "no". However, if some information is common to the source and the sink prior to transmission, the number of bits required to express it may be reduced. For example, if George and Gillian have both read Tolstoy's *War and Peace*, and know it well, they may transmit the ideas in it between them with just the words, "War and Peace." However, if George has not read the book, then Gillian will need to read it to him to get the ideas across in full detail, which will take somewhat longer, and use many more symbols, and hence more bits. It is easy to see, therefore, that taking advantage of shared knowledge will *improve communication efficiency* as measured by the number of bits required to send the message.

Shannon's key insight is that the number of bits needed to transmit a signal may be estimated from the probability of that symbol's occurrence, according to the formula

$$H = -\sum_{i \in A} p_i \log_2 p_i \tag{1}$$

where H denotes *entropy* (by analogy with thermodynamics: it is in a sense a measure of disorder, or the inverse of structure, because it increases as the distribution approaches uniformity, when everything is equally likely), A is an alphabet of symbols, i is a symbol in the alphabet, and p_i is the probability of that symbol's occurrence; the $\log_2$ term is present because it estimates how many bits will be required: there are *two* possible values. It is easy to see that if one symbol is definitely expected, with probably 1, the distribution contains no information (the symbol tells us nothing we did not already know); conversely, if all symbols are equally likely, entropy—uncertainty—is maximised, its value increasing with the size of the alphabet.

This explanation discusses only zeroth-order distributions, but the same principle applies to arbitrarily complex dynamic distributions generated by the application of higher-order statistical models to a given context [132]: armed with a distribution across the alphabet at any point in a sequence, one can estimate the likelihood or probability of the next symbol, and the entropy of that distribution. It is important to understand that this estimate changes as the sequence proceeds: hence, the term *information dynamics*. Mathematically, the entropy value is an estimate of the amount of information in the distribution; there is some evidence, again from music research, that entropy values predict human judgements of *uncertainty*, as we might expect if human expectations are represented as distributions [80].

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

MacKay [132] points out that a corresponding formula can be used to estimate the number of bits required to transmit a *given* symbol, once it is known, as opposed to estimating the average number of bits required across the whole distribution:

$$h_i = -p_i \log_2 p_i \tag{2}$$

where h_i is the *information content* of the symbol i, and the other symbols are as above. This quantity is relevant when considering rates of information that can reasonably be transmitted in human speech. It also serves, in music, with the multi-dimensional, variable-order IDyOM (see above), to predict the human experience of *unexpectedness* very well ($r = .91$, multiple studies) [157,158,156,80]. There is also some evidence that h directly predicts particular neural activity [156], though this has yet to be replicated. Mean information content ($\bar{h}$; sometimes called "cross-entropy" in the literature, because it is determined by n-fold cross-calculation) is the quantity that is minimised in order to optimise the IDyOM model's representation, as described in §3.4, above.

The two information-theoretic values above can be used to chunk sequences of music [160] and language [188] into syntactically similar patterns, with slightly different results. The intuition behind this is as follows. Immediately after the start of a chunk whose sequence has been memorised, the entropy (uncertainty) of the distribution over possible continuations is comparatively low, because the chunk's trajectory in time is known, and this remains the case until the end of the chunk. However, after the end of chunk, there is less evidence to predict what happens next, and therefore entropy rises, as does the information content of the subsequent symbol. Therefore, segments can be identified by detecting rises in entropy and/or information content.

Both information content and entropy serve key functions in IDyOT. Information content is a significant determinant of conscious attention [224]. It and entropy contribute to driving perceptual segmentation, which forms the basis of categorical and sequential perception [215,65]. Finally, Shannon's information estimates serve as a heuristic for choosing and redefining representations that are information-efficient [230].

4.3. Information theory applied to categorisation

Quinlan [165] presents the ID3 (Iterative Dichotomiser) algorithm, which computes an optimal binary decision tree for given data and a given categorisation. It does so by exhaustively computing all possible binary partitions of the data points, and then choosing the partition that most reduces the entropy of the resulting data with respect to the desired categorisation. The key idea is that the biggest reduction in entropy corresponds with the maximum information gain, and so the most efficient decision tree is the one that repeatedly makes the biggest possible information gain first. ID3 is different from IDyOT in two respects: first, it is intended to run over a complete set of data, once only, and therefore its exponential computation cost is not prohibitive, because the resulting decision tree is very efficient to run; and, second, it requires to be given a set of classes. IDyOT, on the other hand, is intended to learn from data over a long period, and to update its knowledge as it goes. And, more problematically, it must learn the categories it uses.

IDyOT theory proposes an idea similar to Quinlan's to partition spaces of perceptual and conceptual values (as discussed in §2 and by van der Velde et al. [203] and Wiggins [230]) into regions that correspond with perceptual categories. This is summarised in §7. First, however, it is necessary to review the theory of Conceptual Spaces [72], on which IDyOT's semantic memory is based.

5. Conceptual spaces

Gärdenfors [72,73] proposes to bridge the scientific gap between symbolic, discrete (e.g., logical) and non-symbolic, continuous, high-dimensional (e.g., neural net) representations. To do so, he introduces an intermediate layer of continuous, low-dimensional, geometric representations. The middle layer forms a means by which a mapping from symbolic concepts to the inscrutable structure of a neural network is managed. Gärdenfors [72] gives many convincing examples of how this might work, from colour perception, through to perceiving the movements of humans and animals. Conceptual spaces have been proposed for musical pitch [183] and musical rhythm and meter [64,65]. At this level of detail, Gärdenfors' conceptual level is very like Boden's conceptual space, as formulated by Wiggins [219], with the addition of geometrical structure. A key point for IDyOT is the fact that the conceptual level of these representations is *continuous*, but that there is a reasonably well-defined theory of meaning associated with it, and a notion of similarity, which underpins notions of similarity at the symbolic level.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

Xiao et al. [233] discuss network (and other) representations of meaning, and how they can promote creativity. It is impossible for a network representation with a fixed vocabulary of node and arc types to introduce new basic meanings: it can only produce combinations thereof. It is possible to extend such a vocabulary in machine learning techniques such as Inductive Logic Programming [144], but this is conceptually and computationally challenging. It is here that the conceptual layer becomes useful in creativity simulation. The geometrical nature of the conceptual layer means that it is literally possible to ask questions like "what concept is one third of the way between these two others?" and to get a range of answers that are meaningful[17]; while this level of precision is probably not required by IDyOT, the flexibility of directly representing "betweenness" [72] is fundamentally important. Indeed, the necessity to differentiate between perceived objects, and then place others on the dimension so created, is a key motivation for the theory of IDyOT memory.

However, Gärdenfors' approach is much richer than just the above. His spaces are of different kinds. Some, such as the space of colour, are delineated by *integral* sets of dimensions: these dimensions are not meaningful unless all are present. Conversely, *separable* dimensions are not necessarily tied together, and may be used independently. Some conceptual spaces have similarity (or betweenness) defined as a Euclidean metric, and some have a Manhattan metric. Some spaces may refer to other spaces, giving, for example, the relative positions (one space) of objects (other spaces). A collection of spaces may be co-originally superimposed to create a manifold of dimensions which are capable of describing, in principle, any structure.

Meanings, directly perceptual or otherwise, are encoded as regions in these spaces, with convex regions being the representation of what Gärdenfors calls *natural* concepts: these are concepts which convex, in the sense that they admit no non-members between any two members. The use of regions affords natural descriptions of relations such as homonymy and hypernymy (equality and inclusion) as well as the possibility of identifying areas of the space that are currently labelled, and thus creating a new concept that is well-defined in terms of the manifold, even without a label. It is also possible to model prototype theory, using geometrical constructs such as the centroid of a region [72].

If follows that the pitch spiral described in §3.4 is a conceptual space, its metric being the Euclidean one of the space in which the spiral is embedded, restricted to points only on the spiral, while pitch height alone is modelled by the one-dimensional metric of the line itself, or by the isomorphic value of the pitch-height dimension of the enclosing space.

Gärdenfors [72] suggests that temporally dynamic quantities may be modelled as trajectories in these spaces; however, this is limiting, as it is hard to compare such structures, undermining the otherwise elegant uniformity of the theory. Chella and colleagues [30–32,29] propose using spectral representations (e.g., Fourier Transforms) of the trajectories, to render them as points in higher-level conceptual spaces, tucking them elegantly back into the theory, as long as appropriate metrics can be supplied for these spaces in turn.

The semantic memory proposed for IDyOT below is a conservative extension of Conceptual Space theory, in the sense that a different mathematical basis is proposed, which will preserve the properties of Gärdenfors' original theory, but enrich it with new possibilities and justifications. This is explained in §7.2.3. Before moving on to describe IDyOT, the final research background to review is that of Cognitive Architecture.

6. Modelling cognitive architecture

6.1. Aims of cognitive architecture research: a unified theory of cognition

Cognitive architecture research has a different aim from the more common phenomenon-focused research in cognitive science, where mechanisms are sought to explicate specific observations of specific behavioural phenomena. Instead, it attempts to provide a general overarching functional structure of, and framework for, cognitive process, sometimes motivated by observations of biological brain structure [115]. The panoptic goal, therefore, is to understand cognition in the large: how the various findings of phenomenon-focused research co-exist, and, in some cases, how they may be explained by fewer, more general mechanisms when taken together.

[17] An example is the percept "purple" which lies between "red" and "blue"; various different shades of purple exist depending on exactly where in the conceptual space the point is placed.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

Allen Newell, designer of Soar,[18] one of the earliest cognitive architectures, gives a list of desiderata for unified theories of cognition [147], also summarised by Lehman et al. [122]:

Goal-orientation Newell subscribes to the idea that humans are goal-oriented. More recent evidence [e.g., 186,16] questions exactly what is the relationship between goals in the sense of conscious, intentional aims, and outcomes of reasoning in a pre-conscious subsystem: it now seems equally credible that goals experienced as conscious decisions are in fact generated non-consciously, and then enter conscious awareness, as the more intuitive converse. If this counter-intuitive state of affairs is actually the case, then it raises the possibility that some goals are in fact the end-states of plans that are made non-consciously, the process of planning being more or (mostly) less available to attention. This would mean that the status of goals as driving forces is somewhat reduced. The assumption of goal-orientation will be examined further below, in particular in context of prediction.

Rich environment Cognition is fundamentally situated, and is arguably selected for (in the evolutionary sense) by the complexity of the environment in which biology developed.

Rich knowledge Cognition requires a large amount of knowledge to address even simple problems in the rich environment.

Symbols and abstractions A conscious agent must be capable of representing knowledge at multiple levels of abstraction: for example, distinguishing individuals from classes, and properties from objects. The exact nature and origin of the symbols in question, however, is a moot point, to which we return below. A key advantage of abstraction in this context is that it affords the ability to reason about generalities, which is, again, particularly important in a predictive system.

Flexibility and sensitivity to the environment The more cognitively developed a system is, the better it should cope in situations where the environment changes. This bears comparison with the taxonomy of agents specified in §6.2.4, below, and, in the current work, the idea must be extended to include prediction.

Learning The more cognitively developed a system is, the more it should be able to learn from experience and apply what it has learned. This is the point where prediction becomes especially valuable: predicting from what has been learned yields the most flexible and adaptive possible agent.

These features all appear to some degree in the selected systems outlined below. Newell's desiderata serve usefully to sharpen our perspective on these ideas; some are self-evidently correct, and some, as noted above, will be questioned or extended in the current discussion. Langley [114] notes that

> Most cognitive architectures draw heavily on results from the study of human problem solving. [114, p. 2]

and, though it is less generally true at the time of current writing than it was in 2005, this is an important point from the perspective of the current paper. The majority of the architectures surveyed here are in the tradition of symbolic Artificial Intelligence, and this was a research field with a strong tendency to treat cognition exclusively as a problem-solving activity. This tendency leads to some implicit assumptions, particularly about the nature of representations, which are not always desirable; these assumptions are eschewed in IDyOT theory, and also in current research in computational creativity, whence that theory came [41,233]. Currently, cognitive architecture research may be grouped into two primary approaches: *cognitivist*, where there is an assumption of predefined symbols that represent meaning in the system; and, *emergent*, where the system generates its own meaning and internal representation language—generally, these languages are not readable by humans [209]. IDyOT is in a third category: a *hybrid*, in that it is symbolic, but *both* its symbols and its rules are emergent from its perception, being given semantics by low-dimensional representational spaces [72]; thus, it has a flavour of the cognitivist tradition, with the particular advantage that its reasoning is explicable in terms of the explicit symbols that it invents and the perceptual signals that ground them, but also that it can be adaptable and sensitive to its environment, in the style of emergent cognitive architectures.

[18] Originally, SOAR was an acronym for "State, Operator and Result", but it is now regarded simply as a name, and the upper case letters have been abandoned. See http://ai.eecs.umich.edu/soar/.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

The most recent and exhaustive survey of cognitive architectures of which this author is aware is due to Vernon[19] [209]. An important contribution of this work is to produce a more detailed range of desiderata for cognitive architectures, to a level which is in fact more granular than will benefit the current outline survey. At a higher level, Vernon argues that there are seven primary groups of criteria that must be considered for inclusion in a cognitive architecture: Embodiment; Perception; Action; Anticipation; Adaptation; Motivation; and Autonomy. Other surveys are given by Vernon et al. [207], Duch et al. [54] and Vernon et al. [208]. The historical connections with Newell's seminal work, above, are clear.

Varma [205] proposes criteria for the design and evaluation of cognitive architectures, which are divisible into two categories. The most prominent, *empirical coverage* and *parsimony*, are properties of the architectures themselves, and will apply here. The others are more methodological: possession of *subjective* and *intersubjective meaning*,[20] leveraging social effect on the cognitive science community; provision of *idioms* that inform subsequent research; and *strangeness*, which may be interpreted as a mixture of originality, unexpectedness and validity. These latter three stand as relations between an architecture and the cognitive science research community, and are best thought of from the perspective of Lakatos's [111] philosophy of science. While they will, indeed, serve as evaluation criteria for the current proposal over time, they cannot yet apply because it has not existed for long enough. However, the longer-standing cognitive architectures surveyed in this section have been selected on this basis.

Parsimony is the key property of the IDyOT model, in the sense that as few mechanisms as possible are proposed to explain as much *theoretical coverage* as possible. In the longer term, the aim is also to maximise *empirical coverage* so as to support and improve the theoretical coverage.

6.2. Computational cognitive architectures

A large number of computational cognitive architectures pass Varma's tests [207,54,208,209]. The aim of these programs is to model cognition in terms of programmed and/or learned rules, within certain constraints that are intended to model human cognitive constraints: for example, that of attentional focus. The tendency is towards essentially reactive, agent-like systems, which respond to perceptual inputs, and sometimes learn from them. Here, only relatively long-standing proposals are reviewed, following Varma's criteria of *meaning* and *idioms*, mainly because they suffice to raise the necessary issues and supply examples for discussion; *strangeness* is hard to quantify, and is therefore not considered here.

6.2.1. Basic operational components

Anderson [5,4] motivates his seminal and life-long work on ACT-R[21] with the following slogan.

> All that there is to intelligence is the simple accrual and tuning of many small units of knowledge that in total produce complex cognition. The whole is no more than the sum of its parts, but it has a lot of parts. [4, p. 365]

This very strong claim will be contested below: while IDyOT indeed relies on the accrual and tuning of many small units of knowledge, the assembly of such parts into more progressively more complex and meaningful hierarchical wholes is paramount; the contention here is that this progressive organisation and reorganisation is a fundamental part of cognition. One interpretation of Anderson's slogan denies dualism, and that interpretation is agreed here.

The same broad principle seems to be assumed in Soar [109,122,110], in LIDA[22] [67,70,43,68,166,69] and in EPIC[23] [99,100]: respectively, these four systems are founded on *rules* (`if-then` statements, or production rules), *chunks* (which are rules, and not perceptual chunks), *codelets* (which are simple pieces of program), and, again, *production rules*. These atomic components are programmed by human programmers, using symbolic languages, to implement theories that they wish to test. The components are then treated (explicitly or otherwise) as *agents*,

[19] Extensive notes and course materials relating to the book cited here can be found on the author's website: http://vernon.eu.

[20] To clarify: these are effects of the theory on the scientific community, not properties of a system that conforms to the theory. An implemented IDyOT does itself have these properties, however.

[21] Adaptive Control of Thought–Rational: http://act-r.psy.cmu.edu/.

[22] Learning Intelligent Distribution Agent: http://ccrg.cs.memphis.edu/.

[23] Executive-Process/Interactive Control: http://ai.eecs.umich.edu/people/kieras/epic.html.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

or autonomous communicating programs [52], that respond to input data that is relevant to them (i.e., which is of a type that they can process); individual agent responses to specific stimuli are determined by the content of the rules, notwithstanding Newell's desideratum of learning, with which they all comply in some way. Thus, complex cognition is broadly modelled as a (large and detailed) collection of small reflex responses, which may cascade, forming more complex, compound operations, as in Anderson's description, above. Further, methodologically, this level of the architectures does not (and is not intended to) constitute a principled set of mechanisms accounting for cognition as a whole. Rather, the mechanisms implemented are designed and built to reflect theory developed elsewhere about individual aspects of cognition, the advantages of this approach being the ability to make testable predictions from the computational implementation and the unified framework within which different theories can be combined together; it is the responsibility of the designers to maintain theoretical and technical coherence and compatibility between agents. In this sense, then, these architectures serve as test-beds for various different theories, and as a unifying context in which they may cooperate. Newel [147] sums up these distinctions in a helpful equation:

$$\text{BEHAVIOR} = \text{ARCHITECTURE} + \text{CONTENT},$$

though it must be noted that CONTENT here includes rules, and not just static knowledge. This approach has a consequence for the current discussion: in this review, the claim that a property does not hold one of these architectures does not entail that it *cannot* be programmed in (or maybe learned) by whatever means is appropriate; rather, it is not a property of the framework *itself*. John Laird makes the same point in context of extension work on Soar [110].

At this level, then, these systems do not appear very different. The differences become more apparent on consideration of the way components are selected, what happens when one or more is applicable, and on the overall control cycle within which their programs are executed. These matters are covered in later sections.

An alternative approach to the design of cognitive architectures is to embody theoretical principles at all levels, and thus to produce a single, designed computational whole. The aim here is different from the work covered above: instead of providing a framework to explore different theories and their mutual interactions, these systems primarily test theories that define the architecture itself. IDyOT is one such. Another, and the only other computational cognitive architecture of which this author is aware that is intended explicitly to study creativity, is CLARION[24] [189,87]. Both of these systems differ from the rule-based architectures in that a learning system supplies their core function.

CLARION is, first, a connectionist system, in contrast with the symbolic frameworks, above.[25] There is a particular focus on the distinction between *explicit* and *implicit* knowledge and meta-cognition, lacking in the frameworks above, and covered in §6.2.3. Explicit knowledge can be given, encoded as directional associative rules that link chunks (of declarative knowledge) together. Implicit knowledge is learned by an algorithm related to standard backpropagation [190], using both implicit and explicit networks: thus, the given explicit knowledge affects the learned implicit knowledge, and the converse is also true. In as far as a CLARION chunk may be considered as a symbol, therefore, symbolic rules (asymmetrical associations between chunks) are, to a degree, common currency with non-symbolic ones.

Necessarily, CLARION's authors do not unfurl the same strong slogan as Anderson, above; nor would this be desirable. In such a network architecture, it is highly likely that some form of generalisation (Newell's abstraction) is going on, and this entails the probability that something much more integrated than a large set of independent productions is being learned—this is the whole point of such an architecture. However, the representational inscrutability of artificial neural networks renders questions about this difficult to ask. Methodologically, it is for this reason that IDyOT uses specifically statistical learning and Bayesian inference for its modelling: diagnostic analysis of such systems is significantly more tractable.

6.2.2. *Sensing, representation, understanding the environment, and embodiment*

For any computational system, artificial or biological, that is embodied in an environment, sensing the environment requires transduction. All the architectures reviewed here are at least implicitly *sensationalist* in intent [3]: that is to say, at least some knowledge is derived from (encodings of) the environment, and not innate; innate knowledge would be simulated by hard-coded chunks, rules or associations, and is not a property of the architectures themselves. There

[24] Connectionist Learning with Adaptive Rule Induction ON-line: http://www.clarioncognitivearchitecture.com.

[25] For completeness, the existence of a connectionist version of ACT-R, ACT-RN [120], is acknowledged; however, detailed discussion will not assist the current exposition.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

is an extensive debate on the nature of embodied cognition, not least as to where the boundary of the cognition lies [209, Ch. 5]; this philosophical debate is avoided here, because it essentially concerns the definition of the word "cognition", and does not do much to elucidate the nature of the systems themselves. IDyOT is separated from the environment via a notional boundary, over which transduction (in either direction, via microphone, loudspeaker, data file, whatever) takes place; the present topic is what goes on inside this boundary. The boundary need not be at the lowest level of continuous real-world representations, but a higher level of perceptual unit (e.g., phonemes or musical notes), but as will be seen below, the perceptual hierarchy inherent in IDyOT's way of working means that it can be moved without loss of generality.

Of the architectures covered here, only EPIC takes a strong position on the nature of sensory transduction, explicitly including processor paths for auditory and visual processing in its design. However, the EPIC architecture includes no actual audio or video processing, and symbolic representations are used, instead presupposing mechanisms deduced by empirical observation [23], and representing their outputs directly [98].

The decision not to include principled transduction, or a theoretical substitute, means that the relationships between representation, learning, and learning of representation cannot be directly explored, even in those frameworks which do admit learning. The sensationalist position entails that knowledge is affected by sensory representation; and, as explained above, there is evidence in music and language that representations change with increasing knowledge: in music, successively more information-efficient representations are learned as exposure proceeds [174]; in language, pre-extant linguistic knowledge (e.g., the existence of words) predetermines perceptual categorisation to some degree [169]. Thus, there is evidence for on-going co-evolution of representation and represented data in humans (see §3.4), but it cannot be captured by these architectures. This is regrettable, as there is a debate to be had regarding the degree to which top-down influences can affect perception [59], and a theoretical framework in which to make the questions precise is desirable. This capacity is fundamental to IDyOT theory, and a novel mechanism, amenable to empirical study, is summarised below.

Anderson [4, p. 359] notes

> The most striking thing about the ACT-R theory of knowledge acquisition is how simple it is. One encodes chunks from the environment and makes modest inferences about the rules underlying the transformations involved in examples of problem solving.

Of course, the devil here is in the detail, because questions of *how* the encoding is done, what features contribute, and concomitant questions such as whether the encoder has biases based on biology or previous experience, are not answered. Necessarily, all cognitive architectures assume some kind of transduction at one level or another, encoding the data therefrom as feature vectors, property-value pair lists, semantic networks, sparse distributed memory, and other well-established representations. LIDA explicitly includes the assignment of meaning—labelling of individuals, categories, etc.—as part of *perception*, explicitly ruling it out of cognition; even qualia themselves are relegated to the pre-cognitive level. Thus, these architectures do not attempt to explain the relationship between sensory transduction and cognition.

Another important aspect of human cognition is *expectation* (that is, *prediction*) [37]. Of the architectures reviewed here, only LIDA encodes perceptual expectation explicitly (though the priming of semantic or neural networks by prior stimuli is a potential source of anticipatory reasoning in any of them). The expectation mechanism is implemented as part of LIDA's perceptual filtering, which as noted above, is explicitly distinct from cognition, and therefore cognitive outcomes are not able to affect the filters (at least, not without circumvention of the architecture's own rules). In contrast, IDyOT's guiding principle is that of prediction and expectation, and it is, we hypothesise, prediction and expectation that give rise to creativity.

6.2.3. Representation and memory

The position on transduction, described above, has consequences for memory and its representation of knowledge. In the conventionally symbolic architectures surveyed here, there are various approaches, all of which consist in representations and processes designed by human programmers according to empirically motivated theories of memory, the most prominent of which is Baddeley's [8]. These theories are based on observed behaviours, and divide memory into various modules based on that observation: long term or short term, working, episodic, semantic, procedural, and so on. It is important to understand that these descriptive theories of function do not entail a corresponding biological

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

modularity, but such a modularity is often assumed, and, in some cases, it is imposed in these cognitive architectures. This approach to modelling denies the cognitive architectures the opportunity to *explain* the different behaviours in terms of underlying function, because they are imposed *ab initio*.

CLARION, again, is different from the other architectures. Explicit knowledge is represented by feature vectors and weights, but they are supplied to an artificial neural network (ANN), and not processed as discrete symbols. A key feature of CLARION's design is its double-layer, implicit/explicit architecture,[26] where explicit knowledge, including rules, clamps part of the implicit network to particular values, thus biasing its learning. The implicit/explicit distinction bears more than a passing resemblance to Gärdenfors' [72,73]. Thus, symbols are given meaning in relation to each other and to perceptual context, but co-exist with nameless regions. The training of CLARION's explicit network with vectors mapped to symbols seems broadly analogous to the mapping between Gärdenfors' symbolic and conceptual levels, and it seems to serve the same semantic function. IDyOT embodies a symbolic method of achieving the same effect, and extends Gärdenfors' theory at the same time.

An important aspect of the IDyOT memory model that is different from the other models (though not in contradiction of them) is the emphasis on *activation* of structure, as opposed to the (albeit metaphorical) *moving* of structure: it is more explicitly distributed, supposing what is sometimes called *in situ* processing [202]. Thus, for example, the phrase "in working memory" is meaningful in IDyOT only in the sense that current matching with sensory and other information is engaged with particular parts of the memory structure. This is a key point in modelling of memory [203]. Sun [190] rules out such an approach in CLARION, on the grounds of an argument due to Jackendoff [93], that working memory cannot emerge from the mere activation of long-term memory nodes:

> in the phrase "the big star's beside the little star", the word "star" occurs twice. These two instances cannot be kept distinct if each consists simply of an activation of the entry for "star" in the long-term memory.
>
> [190, p. 7, footnote 2]

While the statement might be true in a simplistic symbolic system, the argument is not general, because it conflates static, declarative knowledge (the "entry" for star) with knowledge about the discourse (which indeed has two occurrences of "star" in its surface form) and the situation described (which also has two such objects); this need not be the case with all forms of memory. Thus, there is a confusion here which exactly matches a confusion between type and token in the surface form: the type *star* can appear but once; nevertheless, there are two tokens of type *star*. van der Velde [201] proposes a neural-level architecture which addresses this issue: concepts are represented by neural assemblies, which are activated when the concept is in use, and a comparison with IDyOT is given by van der Velde et al. [203]. But they are not discrete, as is perhaps presupposed in Sun's reasoning, above: the two instances of the *star* concept are both part of the same assembly, but each contains structure that distinguishes it from the other, and from the generic. (Of course, similar reasoning applies to the various other concepts activated by the ambiguous word, "star".) If nothing else distinguishes the two stars in any interpretation of this sentence, the mere fact that one is mentioned before the other is enough to do so.

In this view of the world, data is not *loaded into* working memory. Rather, working memory is simply the part of memory that is working: Wiggins [224] draws an analogy with a follow-spotlight in a darkened opera house, moving round to highlight the singer that is currently active. It is important to add that there needs to be some kind of coordination of this (in the spotlight analogy, we need an operator to move the follow-spot[27]) where (or by whom) the activity of the roaming working memory may be summarised and made available to the rest of the mind/brain. Merker [139,140] proposes the dorsal pulvinar as a potential locus of this function.

6.2.4. The cognitive cycle

Any cognitive architecture must have an intrinsic function to make it do anything. This is generally called the *cognitive cycle*, though the implication that there is mere serial rotation through a sequence of events does not hold in all cases. The most basic kind of AI agent repeatedly executes a sense-act cycle, in which a stimulus triggers a

[26] Not to be confused with two-layer ANNs: this is two separate ANNs modules, working together, though necessarily the boundary between them may be ill-defined and inscrutable for the usual reasons.

[27] This is not a dualist perspective: the "operator" is just another part of the mind/brain.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

response that is enacted [172,52]. Russell and Norvig [172] give a taxonomy of agents, depending on their capacity to choose actions, summarised here:

Stateless Agents have no memory and merely respond reflexively. Example: a thermostatic heater.

Agents with Memory have internal state, and can therefore reason from information about past states of the world; they are still essentially reflexive, but better informed. Example: a TiVo unit that automatically records a favourite series, once its user has recorded two episodes.

Agents with Goals have aims beyond merely responding to individual stimuli; their responses to individual stimuli are conditioned by their longer-term considerations. Example: a driver-less car that has no map, but works by following the road and using a compass to determine which direction to turn at each corner.

Utility-based Agents use utility theory to act rationally, achieving their goals in ways that might be said to be more efficient according to a given measure. Example: a driver-less car that gets its rider home by the quickest and most fuel-efficient route by using maps and live traffic information.

Missing from this list is the idea of *predictive agents* [37]. While it is certainly the case that a Utility-based Agent, and possibly an Agent with Goals, will plan ahead to determine a route to its goal, this is a different kind of prediction from that intended here. Of the architectures reviewed here, all except LIDA fall into the Agents with Goals category; LIDA is different in that the cognitive cycle does not necessary begin with sensing, but can be initiated by one of its own actions. This difference is important, because it means that LIDA is capable, at least in principle, of initiating action sequences without external stimulus.

The idea of a predictive agent entails the production of expectations *prior* to the reception of sensory input; production of expectations in response to the receipt of a stimulus entails a processing bottleneck, which would reduce the efficiency benefit of the predictions. Therefore, the predictive agent would work on a different cycle from agents whose cycle is initiated by sensory input. Sensing should not be assumed to be error-free, but, instead, at each sensing step, predictions are made about what is expected in the relevant sensory modalities. In evolutionary terms, this has the benefit of allowing experience to be used effectively [37]: if a current situation is likely to lead to danger, or to the opportunity to feed or mate, then appropriate physical measures (arousal, defence, flight) can be prepared in advance. On an everyday level for modern humans, knowledge about the world, a speaker, and a discourse can be used to overcome the problems of a noisy communication channel, by filling gaps, or helping with corrections. Multiple modalities can be used for this purpose, so, for example, mouth shape can be used as a prompt for phoneme identification, as in the McGurk effect [137]. Further, and probably the most important evolutionary benefit, internally, the likelihood of the various outcomes can be used to assign (biologically expensive) attention: unlikely outcomes will necessarily need more processing than likely, familiar ones.

Given the ability to predict, further evolutionary pressure can apply. An organism that predicts and makes informed choices between its predictions must have something akin to a statistical distribution over the possibilities considered. The properties of this (quasi-)distribution can supply information that supports survival: if the predicted likelihoods are more or less similar, outcomes are uncertain, and this is a reason to be cautious, if any potential outcomes are threatening. Thus, what turns out to be a fairly simple calculation in terms of counting likelihoods and comparing them [181] admits powerful reflection over the quality of knowledge in a content-independent way (for the content of the outcomes does not change the nature of the calculation). This corresponds precisely with the idea of uncertainty that Huron [92] describes and Hansen and Pearce [80] support with their experiments.

The addition of prediction into the agent taxonomy unravels the tight reactive loop assumed in its simpler members. Now there is a freer cycle: agent predicts; agent perceives; agent matches prediction to perception, deciding which is correct (or merging them, or some other solution); agent acts if required. Within the freer cycle, the internal state of the agent is more influential in determining outcomes because, at least in principle, prediction can override perception, and that prediction can be affected by context [215]. A concomitant requirement is cooperative timing: the predictive agent does not merely react to input; instead, to be efficient and effective it must time its predictions to roughly coincide with input to which attention is being paid; otherwise, resource would be wasted in predictions that are too late, or in memory storage for predictions that are too early [65]. In non-human animals, it is self-evident that anticipatory mechanisms can predict aspects of individual physical events (such as the precise moment when and angle at which a dog jumps to catch a ball).

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

Another, more difficult, question entailed by the step up to predictive agents is whether it is appropriate to presuppose goals as a distinguished part of a cognitive architecture. Ron Sun, in explaining CLARION [190, p. 8], writes

> the goal structure is a necessary part of a cognitive architecture.

While there may well be goals, it does not follow that a classically computational data structure, such as a goal stack, is required—which is the kind of approach taken in the older cognitive architectures. While Sun's argument [190, p. 8] that

> In a way, we may view the goal structure as part of the working memory.

is quite defensible, given the assumption of a goal structure, it is important also to consider the possibility that the distinguished status of goals be *emergent* from the overall behaviour of the architecture, and, indeed that working memory behaves similarly, as outlined above. In this case, no special data structure is necessary, and no *a priori* assignation of goals is required: goals are merely derived consequences of the current state that happen to be the focus of attention. This is the position taken in IDyOT theory.

The part of any cognitive architecture that most determines its behaviour is the method used for selecting behaviours. In all of the review subjects, the current state of the architecture, including sensory input where appropriate, is used to select rules encoded as knowledge, which then generate outcomes. In ACT-R and Soar, the rules are symbolic and very much in the vein of traditional AI, for example, being focused on really quite high-level cognitive activities such as performing mathematics, or playing baseball. At this point, all the systems learn, in various ways: ACT-R uses reinforcement learning to increase the likelihood of using a rule in a given context if it is used successfully in that context, for example.

In IDyOT, the motivation is somewhat different. Input matches (exactly or otherwise) with known structure, and gives rise to predictions, which may or may not entail action. Thus, goals are not given: they arise automatically through prediction, and the hierarchical structure of IDyOT memory supplies plans where plans are needed, all at a non-conscious level; only when the properties of the predictions are as required (see below), do they enter conscious awareness.

The anticipatory behaviour of the predictive agent entails the question of what happens when such a prediction is made, but no corresponding sensory input is received. This is the situation described by Mozart [90], and quoted in §2: ideas subjectively appear while he is in the absence of informative sensory input. In the absence of contentful external stimulus, his music-perceptual agents free-wheel, and generate fragments of music. The contention encoded in IDyOT theory is that these fragments are generated by the same mechanism that would allow Mozart to understand music, were he listening to it instead of imagining it, and that this mechanism is freewheeling in the absence of other processing. Wiggins and Bhattacharya [218] suggest that neuroscientific evidence for this proposal lies in the activity of the so-called "resting" brain, often thought of as noise, and therefore discarded from analysis. These options are not available to cognitive architectures that presuppose an action cycle exclusively triggered by sensory input.

6.3. *Global workspace theory: a general theory of consciousness*

An important theory that qualifies as a cognitive architecture without being *a priori* implemented is the Global Workspace Theory (GWT) of Bernard Baars [7]. The LIDA architecture, introduced above, explicitly implements GWT [166].

First, it is necessary to remove the obscuring philosophy of dualism. In a Cartesian, dualist view, a homunculus is required, to be the conscious entity. However, a non-dualist position is taken here; following Shanahan [180], GWT avoids Chalmers' "hard question" of "what is consciousness?" [27,28] and instead asks "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 (c.f. *Society of Mind* [141], *Pandemonium* [94]), operating in parallel over sensory and memory data. The agents *compete for access* to a *Global Workspace* via which, exclusively, information is exchanged; access is restricted according to rather vaguely specified criteria [224]. The Global Workspace can be read by all the agents, and models consciousness. It can contain exactly

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

one "thing" at a time, though that "thing" can be anything. Meaning in the Global Workspace is context sensitive and structured; contexts can contain goals, desires, etc., of the kind familiar from the discussion above. 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 of Wallas [210]; this equation carries through into IDyOT. However, he does not develop this idea further beyond noting that a process of creative refinement may be implemented as cycling of information into the Workspace and out again. To the best of the current author's knowledge, creativity in the Global Workspace has not been directly addressed elsewhere.

An important aspect of GWT is the idea of *information integration*, where sensory input is progressively integrated as it approaches the Workspace. Tononi and Edelman [194,195], propose information-theoretic measures of information integration as a measure of consciousness. Baars has embraced information theory, too, and the three authors have proposed a "conscious"[28] machine [56] based on these ideas. The contention of IDyOT theory is that information integration can be modelled by a statistical process of inference analogous to Pearce's [155] methods for combining Conklin's viewpoints (as outlined in §3.2).

This concludes the review section of the current paper. The next section summarises the theory of the Information Dynamics of Thinking, which aims to draw all the previous ideas together in one minimally-complicated, principled mechanism.

7. The information dynamics of thinking

7.1. Aims and inspirations

The aim of the current work is to test how far a minimal cognitive architecture (IDyOT) based on some very simple but powerful mathematical principles can go in accounting for human cognition and creativity. The goal is the closest possible shave of Ockham's razor.

There is an evolutionary and biological inspiration behind the work: the aim is to propose an explanation of why the cognitive architecture should do what it does in terms of the evolutionary development of humanity. This entails that it is not only a learning system, but that it accounts for life-long learning, in a human-like way. This feature complicates assessment of IDyOT's performance, because it has to be treated as a discrete, incremental dynamical system, rather than being evaluated by more familiar static, summative methods from machine learning. In keeping with the minimalist approach, there are no innate capacities beyond the essential mechanisms of the cognitive cycle [203, summarised below]; therefore, IDyOT must learn everything it knows and does.

The minimal evolutionary position also entails, literally, a *raison d'être*: why should a human mind function this way? The principle behind IDyOT is of *information efficiency*. Information processing is very expensive, in evolutionary, developmental, and energetic terms: a successful cognitive entity must therefore jealously conserve this precious resource.

Information efficiency determines three key factors of IDyOT's operation: first, it is predictive: it uses its information to assist its perception and hence operation in its environment; second, it processes input from its environment so as to construct the most efficient possible model of that environment; third, it applies the most expensive resource of all, active attention, only to those structures that contain sufficient information to warrant it. This third factor entails that IDyOT is, in the terms of Biederman and Vessel [13], an *infovore*: that it, it is *curious* [178,175]. Biederman and Vessel provide a biological account of the developments necessary to make this the evolutionary norm in humans: the process of learning causes mu-opioid release, and thus becomes a desirable end in its own right.

The minimalist nature of IDyOT entails the exclusion of some self-evident factors of human cognition, which have been mentioned in the foregoing review: IDyOT is different from other cognitive architectures in not having explicit goals and goal management, and, more importantly, in being agnostic as to the origin of what humans experience as their goals. In particular, IDyOT excludes affect, though there is nothing to prevent a simulation of affect being added later. The concomitant question of what are IDyOT's intrinsic drives is answered thus: IDyOT is curious; it seeks new information, and it continually predicts and generates. Any goals it may have emerge from this behaviour.

In terms of Vernon's [209] seven properties of cognitive architectures (see §6.1), IDyOT is positioned as follows.

[28] To be clear: this particular definition of "conscious" is not used here.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

Embodiment IDyOT is conceived purely as a computational theory, and so the core model does not include special facilities for embodiment. However, unlike many cognitive architectures (and all those surveyed above) IDyOT does have an account of transduction which will admit the addition of continuous-valued sensors in future. This is illustrated in Fig. 2 and explained by van der Velde et al. [203]. Thus, IDyOT can in principle conform to most non-extreme notions of embodiment by means of appropriately rich and, crucially, bidirectional connection to a robot.

Perception IDyOT is capable of perceiving its environment by means of the same transduction, or by higher-level and therefore less realistic discrete simulations. A crucial factor is the incremental transition from continuous through progressively more discrete domains of sensing afforded by the transduction and related learning mechanism; this means that IDyOT's behaviour can be simultaneously studied at multiple levels of representational abstraction, affording a powerful research tool. A consequence of this approach is that the boundary between perception and cognition in IDyOT is blurred: there is an incremental sequence of representations what are progressively more abstract until representations top out in simulations of experienced meaning.

Action Again, IDyOT can be given the capability to act via continuous transducers driving a robot. In the initial versions, its actions are limited to emitting streams of symbols that correspond with meanings in its memory, or distributions across them.

Anticipation IDyOT's driving principle is anticipation, not just in the weak sense of planning what to do next, but in the stronger sense of prediction of the upcoming events in the environment, and beyond the next contiguous event. Anticipation drives everything that IDyOT does.

Adaptation IDyOT learns continuously, and continually revises its memory according to what it learns. Thus, not only does it learn new facts, but it adapts its representation to account for new patterns in the data to which it is exposed. Since its memory model directly drives its behaviour, its behaviour also adapts.

Motivation IDyOT has no predefined motivation; instead, it behaves as determined by its cognitive cycle. Motivations may arise once it has learned something to be motivated by, because of the prediction mechanism. Thus, motivations may arise as a result of external stimulus, or, if external stimulus is lacking, because of the information content of the distributions that it has learned. However, the current research avoids the approach of built-in emotions and goals, or similar, on grounds of its minimalist approach: it is appropriate to study the emergence of a system without such drivers, in order to understand how they might best be implemented. There are philosophical problems associated with "giving" a cognitive architecture affective response, because, at best, it can only be a simulation of human (or other) response, in the absence of qualia within the architecture; it is not clear to the current author that the addition of such facilities is meaningful.

Autonomy Once started, an IDyOT is completely autonomous: it senses, and if sensory input is available, it is processed. If no sensory input is available, IDyOT generates internal data to process.

The mechanisms that afford these properties are described in more detail by Wiggins [224], Wiggins and Forth [215], Forth et al. [65], van der Velde et al. [203], Wiggins [230]; the current paper aims to provide a synoptic overview and to add novel components and connections.

7.2. IDyOT memory

As with any cognitive architecture, IDyOT's operation is heavily dependent on its memory representation. There is not space here to review the extensive history of knowledge representation in cognitive science and AI: Xiao et al. [233] provide a review specifically relating to creativity. This subsection summarises the operation of IDyOT memory, including a novel contribution on memory consolidation, which is further explored by Wiggins [230].

7.2.1. IDyOT's memory structure

IDyOT's memory model is founded on a multidimensional temporal sequence of perceptual inputs, quantised at a resolution of around 40 Hz, as this is a candidate for the boundary between the sensation of pitch and that of rhythm [116,150]. Each dimension has its own conceptual space: for example, auditory input is represented as a time-variant spectrum (with phase) modelling the input from the Organ of Corti [143]. A parameter determines the size of conceptual regions in the spaces: points in the same region are considered to be the same. Boundary entropy (§4.2) is used to identify boundaries in this sequence, and each resulting segment is represented by a symbol at the next

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

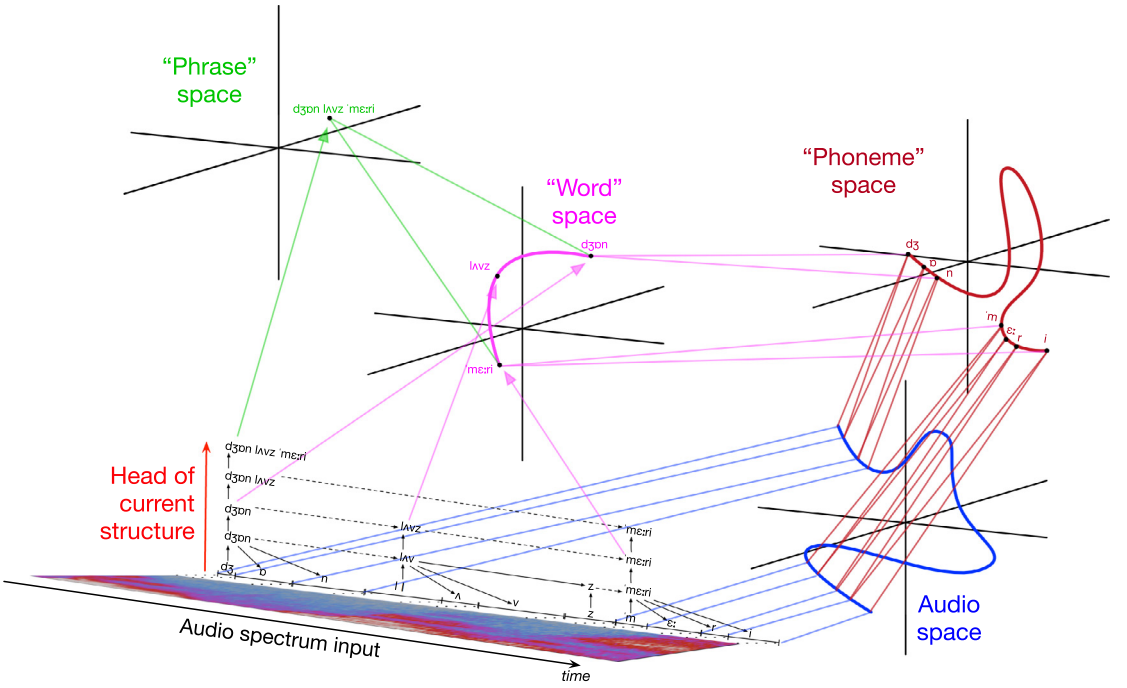

Fig. 2. Schematic illustration of IDyOT memory generated by one perceptual input: the spoken sentence, "John loves Mary". Detail of "loves" is omitted, to avoid clutter; it is treated exactly as the other words. The audio spectrum (approximating output from the Organ of Corti) is chunked into phonemes, then morphemes, then words, presented here in idealised form: reality is more messy. As chunking proceeds, new layers are built above the lower ones: the head symbols in the leading edge of the structure subtends short-term memory. As sequences grow at each level, they are also chunked, building a structural representation of the input. The resulting episodic memory structure is shown here in black. As each layer is built, Markov models are updated; here, implications are illustrated as arrows: solid for low-entropy predictions; dotted for high-entropy. Each slice of the perceptual input corresponds with a point in an appropriate conceptual space, and trajectories of such points are represented spectrally (and therefore independently of time) in the conceptual space of the next layer up. The spaces in the diagram are colour-coded: the episodic chunk subtended is shown by coloured lines between spaces. Each symbol in each sequence is associated with a representation of the duration of the sequence that it subtends [65]. Symbols in the episodic memory (notated in IPA) correspond with points in the corresponding space, shown by light arrows in the corresponding colour. The illustration supposes that this IDyOT has been exposed to enough English speech that it has learned the components of the example sentence. Because this example uses only one dimension, it is limited to the representation of regularities in sound; van der Velde et al. [203] give a multidimensional example with the same sentence. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

level up in the hierarchy (terminology: the upper symbol *subtends* the lower sequence). The conceptual space of the upper level is populated by points which are spectral representations of the temporal trajectory described by the points in the segment. Each of these conceptual spaces is a vector space, generalising Gärdenfors' [72] Cartesian conceptual spaces (see §5), and has an inner product that yields a norm that models similarity in the perceptual dimension. Fig. 2 schematically illustrates this structure. Hedges [86] has built a preliminary simplified implementation of the IDyOT memory model, and found it to discover structure in musical harmony in a human-like way.

In computational implementations, the lowest layer, which corresponds with, for example, echoic memory in audition, must be discarded, as it is in the human case, because there is simply too much data to store. Thus, abstraction leads to information efficiency.

7.2.2. *Declarative vs. implicit memory*

A distinction is made between *declarative* (or *explicit*) and *implicit* memory [62]. The former is memory acquired via conscious awareness, while the latter is acquired without conscious attention. Here, it is useful to note that this distinction is not made in terms of the information, or the means by which it is stored, but in terms of the way it is acquired. In IDyOT, there is only the following distinction: information transduced from the environment may be stored without attracting attention, and hence without being processed via IDyOT's Global Workspace: this constitutes implicit learning in these terms. This is the kind of learning that occurs when humans are exposed to language,

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

where structures are learned that model the surface statistics of the language experienced, even though the meaning is the focus of conscious attention. Thus, humans and IDyOTs are able to make predictions about the world without conscious intervention. Therefore, there is no top-down, systemic distinction between implicit and declarative learning in IDyOT, beyond the question of whether an item passes through the Global Workspace or not. The representation of the memory is the same in both cases.

7.2.3. Episodic vs. semantic memory

Declarative memory is generally described as falling into two kinds, the *episodic* and the *semantic* [196]; IDyOT supports this distinction, though refining the notion of "semantic." IDyOT's episodic memory consists of a multi-dimensional, hierarchical statistical model that records low-level sequences of perceptual inputs, and builds on top further layers that result from their segmentation by boundary entropy (§4.2, Fig. 2). Thus, IDyOT's episodic memory is a literal representation of everything it has sensed explicitly related to everything it has inferred from that information. Aspects of human memory such as forgetting are not explicitly modelled, though, as mentioned above, IDyOT will need to forget its most literal sensory input, as there would be too much of it (arguably as for human echoic memory).

IDyOT's semantic memory, on the other hand, is a hierarchy of conceptual spaces, again grounded on the literal perceptual input, but relating to the symbols used to label those inputs and the segments cut out of them, and independent from the sequences in which they appear. Each dimension of each layer has its own space, and each layer's space is a spectral representation of the one below it. The spectral nature of the representations is fully motivated in §7.2.5; its most convenient feature in the current context is that it abstracts away time. Thus, for example, a slower speaker's language may still be understood, even though IDyOT has learned from a faster one; and short and long sequences may be compared.

IDyOT's notion of "semantics" is very general. It covers logical meanings, or sense, at the higher levels of the hierarchy, though it does not include reference, since that is encoded explicitly in the relation between the episodic and semantic memory (illustrated in Fig. 2 by coloured arrows). Furthermore, it covers sensory memory, right down to perceptual input. Although IDyOT's representation does afford a propositional interpretation (by taking Gärdenfors' step from conceptual space to logical representation), it is primarily a representation of relations between meanings, giving them sense in terms of each other and of the stimuli from which they arose.

So, while remaining deeply skeptical as to the ability of silicon to experience qualia, the semantic memory of IDyOT is perhaps best described as modelling *the feeling of perceived and abstract meaning*.

7.2.4. Short-term vs. long-term memory

A distinction is made between Short-term and Long-Term memory [6,9]. Often, cognitive scientists write as though the computational metaphor for mind were literally true, with data being moved in and out of memory for processing. The *in situ* approach to processing [203] undermines this metaphor; in IDyOT, where processing is thought of as *in situ*, the distinction between short-term and long-term memory (in Baddeley's terms) is that "short-term memory" includes the memory structure subtended by the topmost leading symbol in the memory, while the boundary of long-term memory does not include this structure until it is complete, in the sense that it has no further expectations for immediate continuation.

7.2.5. The spectral nature of IDyOT conceptual spaces

At several points in the current discourse, the suggestion has been made that, in order to implement IDyOT's semantic memory, the geometrical theory of Gärdenfors [72] should be extended from its low-dimensional, real nature, to a more powerful mathematics involving generalised vector spaces and complex arithmetic: specifically Hilbert space. This step violates the principle of Ockham's Razor, espoused at the beginning, and therefore demands justification. There are several pieces of evidence, more or less circumstantial, as to how spectral representations of meaning may be justified, arising from several different aspects of IDyOT and of the system it models.

The first is the oscillatory nature of the brain [e.g., 26]. In order to represent the behaviour of biological neural networks directly, oscillatory behaviours must be captured. One way to do this is with time-domain differential equations [e.g., 117], but this is prohibitively computationally expensive to implement on current computing devices at scale. Another is to use vector spaces whose geometry can represent the oscillations independently of time, and/or represent time in a static way.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

Secondly, it is a property of IDyOT memory that temporal trajectories of different lengths, in both episodic and semantic memory, will need to be compared. This is necessary, for example, for the formation of structures analogous to syntactic categories, and for the representation of classes of movement which may take place at varying speed. As mentioned above, Chella et al. [31,29] proposed the use of Fourier transforms to perform this representational mapping; IDyOT theory uses the same idea as an initial hypothesis. To represent a Fourier transform as a point in space calls for a spectral mathematical structure of variable dimension. This proposal affords a rich array of empirical possibilities, because it means that perceived similarity should be modelled by norms in these (hierarchically structured) spaces. It is necessary to accommodate time-invariant representations of classes of percepts such as perceived movements [31] and sounds in such a way as to afford the correct notion of similarity between them: a word spoken slowly remains the same word if spoken quickly, notwithstanding the subtleties of timing within the word, and this ought to be implicit in the representation. The representation illustrated in Fig. 3 for sound shows how dynamic detail may be represented within a hierarchy. However, that representation requires the underpinning conceptual space to represent chunks which are different speed versions of a common prototype, and this is afforded by spectral representations.

Third, an argument from a different direction is the successful application of so-called quantum logic in linguistics [212]. This theory is not directly to do with quantum physics, as might be supposed from the name, but it does use the same mathematics: that of Hilbert spaces.

A fourth piece of circumstantial evidence comes from speech processing. In this area of audio signal analysis, the most successful representation, for the purposes of computational understanding of human speech, is the Mel-Frequency Cepstral Coefficient (MFCC) vector [24]. These coefficients are calculated by applying two spectral transformations, in turn, to a speech signal: very similar to the upshot of the first two layers of the IDyOT memory. Thus, the success of MFCCs as a speech representation supports the method of operation of IDyOT memory.

Finally, Gärdenfors' conceptual spaces are a special case of the more powerful mathematics proposed here, in that they use real numbers (a subset of complex numbers) and use a restricted range of distance metrics (relatively simple inner products that induce Euclidean and Manhattan norms). So this proposal constitutes a conservative extension of Gärdenfors' theory.

7.2.6. Memory consolidation

This section presents a novel contribution to IDyOT theory. It is self-evident that IDyOT's incremental learning process will create chunks which are later proven to be incorrect by further data, in a way that one-shot learning will not. In this way, it is more human-like than many machine learning algorithms. Behaving thus has the same drawback that humans experience: it is necessary periodically to clean up—to revise and consolidate memory. Eysenck and Keane [58, Ch. 6] present a helpful summary of human learning and memory research, including this process, which is not currently well-understood, in cognitive terms. The aspect of memory theory on which IDyOT draws is the idea that *memory consolidation* is a process whereby memories that are insignificant (for whatever reason) are pruned away [232], thus reducing the representational load. IDyOT supplies a new proposal for the mechanism by which consolidation takes place: namely, the re-representation of learned data to improve *both* the expressive power of the representation and the predictive power of the model that is used to express it. Necessarily, to be effective, consolidation needs to take place mainly when the memory is not receiving new information, corresponding with sleep in humans.

The aim of IDyOT memory consolidation is to find the optimal categorisation, in the sense that $\bar{h}$ (see §4) is minimised both within the perceptual space (as with ID3), and *also* in the sequential model (as with IDyOM). Thus, the sequential element of the model interacts with the existing representation(s) to choose alternative representations to better suit the data space in context of the sequences. The essential principle is that perceptual values are more likely to be grouped into perceptual categories if they predict and/or are predicted by the same kinds of things. The flavour, then, is of the kind of statistical linguistics based on cooccurrence, but there is an additional layer of categorisation, implied by the sequential statistics, working in parallel with and sometimes in opposition to the cooccurrence statistics. The tendency is towards fewer symbols, because more symbols cause more uncertainty, and hence higher entropy, but this is counterbalanced by the loss in prediction accuracy if too many symbols are merged; thus, equilibrium is reached, and adjusted as more data arrives. Consolidation involves three different, interlinked effects.

Symbol definitions The regions in conceptual space corresponding with symbols may be adjusted, merged or split, like the constructs in the language learning process proposed by Kirby [102]; indeed, IDyOT supplies a more

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

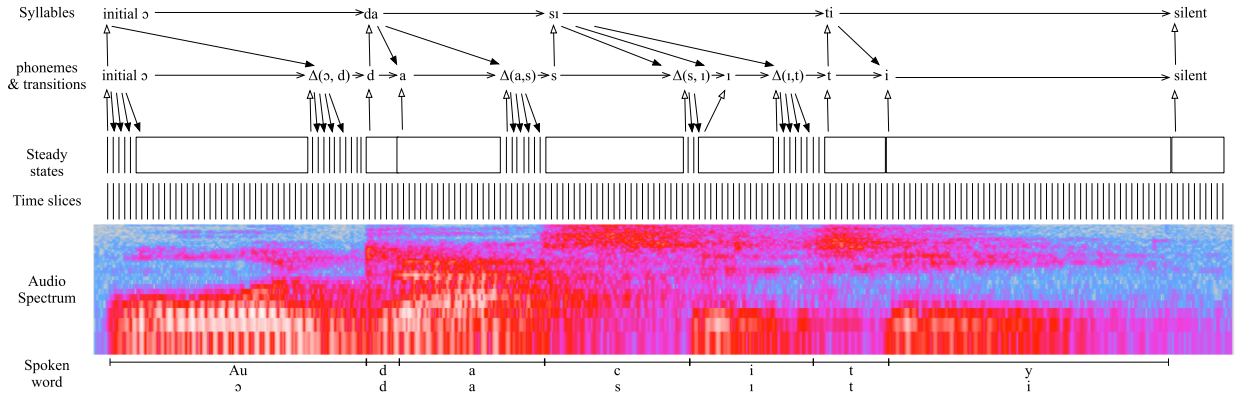

Fig. 3. The mel scale audio spectrum for the word "audacity", spoken by the author in UK standard English, in a somewhat noisy environment, and rendered by the program of that name. Below the spectrum, the sound categories of the spoken word are annotated for reference. Immediately above the spectrum, regular vertical lines represent the frames of the fixed-time Fourier transform. Above this, steady state sections of spectrum are segmented into single blocks; this is the most immediate form of segmentation. Further still above, following more exposure to data, the regularities of phonemes and transition sections between phonemes are detected, and, finally, syllables are identified. Beyond the scope of this diagram, progressively larger groupings appear as illustrated in Fig. 2.

general mechanism that would account for Kirby's rules. In the two latter cases, symbols must be removed or added, respectively.

Conceptual spaces The conceptual spaces themselves may need to change their geometry in order to re-represent the new data efficiently, as it appears. This can be achieved by search through a range of inner product functions for the relevant vector space, allowing only candidates that are cognitively plausible, analogously to the work of Kemp and colleagues [97,96], who identify candidate mathematical structures on the basis of statistical properties.

Hierarchy change As a result of the above changes, the IDyOT memory hierarchy may change, with the addition of new derived representations (in music, for example, the derived representation of pitch interval from pitch). Therefore, the associated predictions may also change.

The process of consolidation, illustrated for one example in Fig. 4, is based on the underlying principle of information efficiency: that prediction should be as good as possible, but that the representation used should yield as compact a model as possible. This is the principle, introduced in §3.2, used by Conklin and Witten [42] and Pearce [157] under the name of *viewpoint selection*, but extended to include potential live inference to create new representations. In these earlier statistical models, programmers provided multiple representations of input data, some of which were basic (meaning, essentially, atomic in terms of the level of representation considered: they could not be broken down into smaller sections). A search process was then used to select the subset of these representations that gave the model that best fit the data, in terms of 10-fold cross-validation. The intuition behind this is that the model is trying to represent its data in such a way that everything it knows is, on average, as expected as possible, which corresponds, on average, with accurate predictions. The music-cognitive motivation for this idea was explained in §3.4.

The example of the speech domain gives further insight. As the low-level, time-sliced perceptual input stream is analysed, it will initially produce segments which delineate areas of no change: the transition matrix for the layer quickly biases towards repeated symbols and creates segment boundaries at their ends. Thus, a more time-variant representation is produced, one level above, in which steady states are represented by one variable-length symbol instead of multiple fixed-length symbols. This, and the two layers above it, are illustrated in Fig. 3, which is a more detailed illustration of the structures developed in Fig. 2. It can be seen that the low-level, time-slice representation serves no useful predictive purpose once the layer above it is determined, because it is much more likely that a variable-length segment will match a new input than the fixed-length bottom level. Similarly, the change-points can be learned with only a little more difficulty, and the Fourier-type representation will afford a similarity space to identify them. Once these points have been established, and enough data received to create a reliable general model, the statistics will naturally lead to the identification of segments that are more general still, and new symbols to label them. Thus, information that is unhelpful in predicting (in this case, the exact length of the steady state) is

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

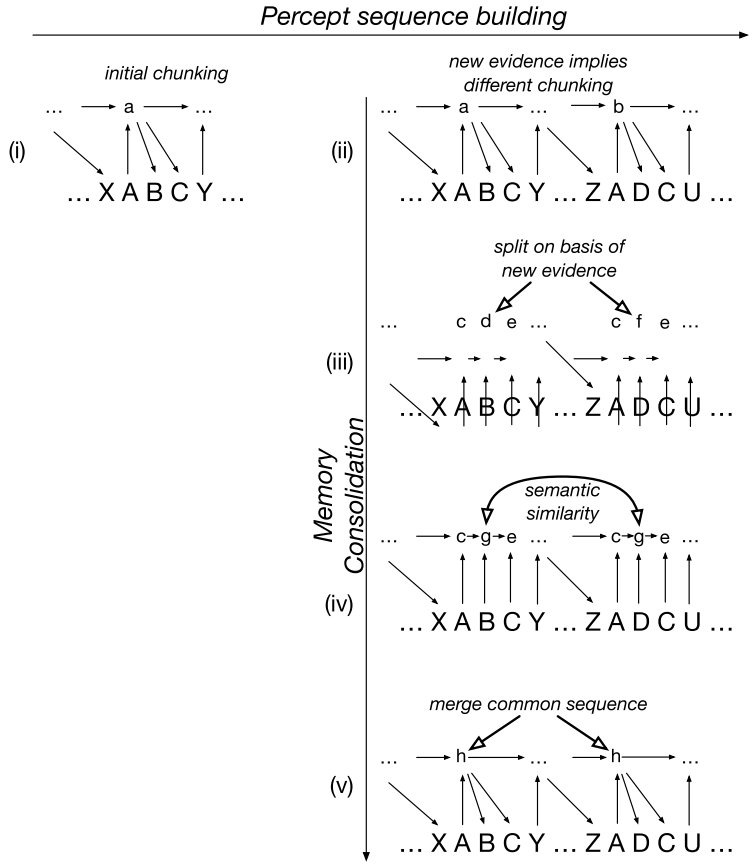

Fig. 4. Schematic illustration of one effect of memory consolidation. (i) A sequence is chunked as it is received, but (ii) later data means that a different chunking is preferable. (iii) Memory consolidation first applies the new chunking, adjusting the local memory accordingly (iv) and then examines the possibility of merging the new higher-level symbols on grounds of similarity, which succeeds in this example. (v) The result is a new higher-level symbol which behaves like a syntactic category, subtending more than one possible sequence below. It is important to understand that these steps will only be taken if the information efficiency of the model is improved: this would depend on the rest of the model, not shown here.

ignored. What is needed, therefore, is a process whereby such representations can fall out of use, or perhaps be deleted altogether, to improve predictive efficiency. For this reason, the different levels of representation must be weighted, much as Gärdenfors' quality dimensions, so that irrelevant information does not reduce the quality of the prediction. This approach means that some uncommon details may be sacrificed to the quality of the overall model, which may be expected to give rise to some of the forgetting effects predicted by interference theory [e.g., 200].

Alongside this, it is necessary to optimise the IDyOT-generated symbols themselves, and to identify which should be considered to be equivalent. Without this process, IDyOT would simply create more and more labels, with no indication of association where necessary, which would be particularly problematic if early-learned structures happened not to be statistically general. The process admits a kind of grouping in its own right: new points appearing inside an extant labelled region of a conceptual space may be considered to have the same label. However, the incremental nature of IDyOT learning means that it is possible for memory traces to be recorded in ways that are subsequently determined not to be information-efficient: for example, with inadequate categorisation schemes. In this instance, the memory consolidation process, though not guaranteed to optimise, will help.

In consolidation, IDyOT takes each conceptual space learned, moving from the lowest to the highest, and considers possible re-labellings of the relevant chunks that are close to the current state of the memory. Success is an improvement in the predictive power of the overall model with respect to the known data, measured by $\bar{h}$. The search space of re-labellings grows exponentially with the number of percepts (at whatever level) encountered, so exhaustive search is intractable, and heuristic methods must be used, with a limit to the amount of change allowed. The approach

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

proposed here, which is future work for implementation, is that semantic regions of episodic symbols be adjusted to include points in the space that are nearer to its centroid than they are to the centroid of the (different) region in which they are currently contained. The effect of this is to adjust regions towards convexity, a required feature of concepts in Gärdenfors conceptual spaces—but only if overall $\bar{h}$ is improved. Essentially, the idea is to correct misclassifications. In the limit, it is possible for a region to lose all its points to one or more others, in which case its label simply disappears. Thus, the process can extend or reduce the alphabet for each layer, and knock-on effects may propagate up the memory structure. Therefore, a stopping point is needed: there is a brake on this effect, because the entropy of each prediction increases with the size of the alphabet, until the alphabet becomes too small for prediction to be accurate; thus, alphabet size and prediction accuracy mutually countervail to produce a good representation and model.

7.2.7. Prediction

In order to make predictions at any level of the memory hierarchy, IDyOT maintains a first order transition matrix for each layer, in each of three directions: forward, along the sequence of the layer, upward, to the more abstract layer above, and downward, to the less abstract layer below. Currently, the hypothesis is that these models are first-order, with interactions between layers supplying the more subtle statistics calculated by the PPM* algorithm [38] successfully used in IDyOM and elsewhere. Furthermore, as in Conklin's viewpoint system and IDyOM [42,157], predictions are combined between statistical models from different input features, depending on the level of statistical correlation between the features [155,84].

7.3. Global workspace and attention

IDyOT was partly inspired by, and is, to a degree, an implementation of, Baars' [7] Global Workspace Theory (GWT: §6.3); it solves an important problem in the mechanism of the Workspace, the so-called Threshold Paradox [229], and instantiates the idea of information integration in an empirically testable way [224,215,65,203]. A key aspect of GWT is competition between Baars' agents-generators for access to the GW. The proposal here is that this is controlled by information content, and, crucially, that it drives chunking. Current and planned research on IDyOT includes empirical assessment of several different proposals for the detail of this work, motivated by work, outlined above, in music cognition and linguistics that shows a good correlation between information content and/or entropy, and perceptual segmentation in sequential stimuli [188,160]; the current IDyOT work, therefore, aims to find an explanatory [223] mechanism for this process. This said, nothing presented here is at odds with the view of Oakley and Halligan [148], that conscious experience is an epiphenomenon of non-conscious processing: the current author does not see Baars' proposals, particularly the later work that admits distributed, hierarchically structured, global workspace, as being at odds with the epiphenomenal view, but rather as a more mechanistic framework in which it might be implemented.

The key aspect of GWT that is relevant to IDyOT is the regulation of attention: the theory is agnostic as to answers to the "hard problem" [27,28]. In the current view, attention is a property associated with information processing: where there is high information in IDyOT memory, there is a focus of attention. In particular, attention can move up and down the memory hierarchy, as well as along it. This position affords testable hypotheses on the processing of sentences like that in Fig. 2: the information content at different points in the memory structure will change as input is received, and thus testable predictions can be made about the dynamic focus of attention as memory population proceeds.

7.4. Timing and the cognitive cycle

As noted above, IDyOT operates in a quantised version of real time, which is different from many cognitive architectures, whose timing is not related to real time at all. IDyOT symbols are recorded in episodic memory alongside a representation of their duration. This allows the prediction of temporal expectations, as discussed in §3.6 and detailed by Forth [63], Forth et al. [65]. It also affords a new and better definition and explanation of the capacity for entrainment than was previously available [63,65].

Another difference between IDyOT and other cognitive architectures is that IDyOT is not a reactive agent, but that it predicts continually by sampling from its existing memory. If perceptual input is present, then the prediction is of what is next [203] and when it will occur [65]. Otherwise, IDyOT generates freely from its memory, and it is this

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

that forms the basis of the creativity described in §9. Wiggins and Bhattacharya [218] propose that this process is the reason for the so-called "noise" detected by neuroscientists in the "resting brain."

Continual prediction also raises the question of when the statistical babbling it produces should be switched off. A top-down rule would be antithetical to the minimalist principles of the approach; so the solution should be sought in the Global Workspace. As IDyOT learns more, information content in context will be a progressively better guide as to what is important. Thus, the threshold approach mentioned in §7.3 serves as a choke on unnecessary predictions, which might have been manifest as movements, utterances, or other effects, depending on transduction.

8. Validation, predictions and future work

In this section, behaviours expected to be emergent from an IDyOT are considered. These, again, constitute ways in which the theory might be evaluated as a cognitive model, by empirical comparison with human behaviour. This affords the opportunity to propose testable hypotheses and/or identify future work. Once validated, IDyOT will be used to make testable predictions about human behaviour, in a methodology similar to that used in the IDyOM work (§3, [159]), based on its observed behaviour in situations such as those listed below.

8.1. Validating the model

The first step to validating a computational model of this nature is to compare specific aspects of its behaviour with the human behaviour they are claimed to simulate. Music will be a key domain, again, though language is relevant to IDyOT in a way that it was not to IDyOM (§3), because of the semantic elements of IDyOT, which will afford verifiable outcomes. Given this, comparison with both human behaviour and the outputs of IDyOM (which has been shown to be a good model of human behaviour) will be key methods of evaluation.

8.1.1. Sequential memory model and consolidation

As mentioned earlier, the IDyOT memory model takes the place of the PPM* algorithm that was used in IDyOM for prediction. It is hypothesised that IDyOT's behaviour would be approximately equivalent to PPM*, for any given dataset. This hypothesis may be tested empirically, but also by means of formal comparative analysis of the two algorithms.

8.1.2. Prediction accuracy

IDyOT should predict human melodic expectation in the same way as IDyOM, and to a comparable or better accuracy. (Because of ambiguity inherent in isolated musical stimuli, one cannot necessarily expect universally better performance on stimuli drawn from real music [160].) This is easily testable, because the empirical data required for comparison has already been gathered and published.

8.1.3. Validation of conceptual spaces

IDyOT should chunk speech in a human-like way: multiple standard test sets for chunking exist. It should also develop conceptual spaces that correctly model the vowel and consonant spaces established in phonology.

8.1.4. Validation of quasi-grammatical structure

As discussed by Wiggins and Forth [215], IDyOT's memory will produce structures which are reminiscent of grammar. In a simple IDyOT, based on speech or text input only (that is, without associative semantics), conceptual spaces should learn to group syntactic categories in recognisable ways.

8.2. Epiphenomena

If the model is correct, then some epiphenomena may be predicted to arise from its behaviour. This, in essence, is the point of the project: the desired epiphenomena are observable features of human cognition, but the claim here is that they arise from the interactions between the various components of IDyOT during its learning and prediction process. If so, they no longer need to be modelled descriptively, but are explained by the underlying IDyOT model [223]. It is in the nature of these proposals that they are more speculative than the detailed validations mentioned above; they are therefore the subject of more long-term research.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

8.3. Compositionality and unbounded dependency from association and consolidation

An open question in the cognitive science of language is that of how compositionality can arise from processes which appear to be fundamentally associative [e.g., 102,185,103]. Kirby [102] proposes a rule set that admits the emergence of grammar by a process of sequence association and then revision, crucially allowing the production of two symbols by dividing the sequence associated with what previously had been one, on the basis of evidence from segmentation patterns.

IDyOT memory consolidation generalises this idea, and provides a hypothetical mechanism by which it may take place, as illustrated in Fig. 4: the rules of memory consolidation admit such splitting, and the consequence of doing so, in combination with the hierarchical nature of the memory model, lends Markov chains the capacity to learn more powerful probabilistic languages than finite state. The nature of IDyOT's semantic memory is based on the notion of composition, in the mapping from time-variant sequence to pointwise spectral representation. However, the process is general, and not just linguistic. Thus, linguistic learning emerges from general cognition, so long as there are enough layers in the memory to support these operations, which raises interesting questions about the relation between cognitive capacity and cortical volume. These issues are explored further by Wiggins [230].

These capacities afford a range of potential comparison with human language, featuring long-term dependencies. IDyOT's ability to deal with such phenomena can be assessed by noting the information content of the appropriate sentences. If it has learned correctly, then correct dependencies should not be surprising, and the converse would also hold.

8.4. Imitation in early learning

An important area of robotics is that of imitative learning. IDyOT's operating loop, if connected to appropriate sensors and actuators, provides a potential mechanism for bottom up imitative learning, providing a cognitive-architectural motivation for systems such as that of Gaussier et al. [74]. Crucially, an IDyOT requires no specific motivation to imitate. If an otherwise-untrained IDyOT is shown how to do something, it will imitate that thing, because its statistics will make doing that thing maximally likely. As it learns a wider range of possible actions, the IDyOT will become less likely to imitate more familiar things, and perceptual context will become more important in selecting from the learned actions. However, it will remain more likely to reproduce newer actions, because they are more novel and therefore have higher information content (see §9). The lack of need for motivation arises from the system's continual production of predictions: it will always try to do something, and so if it knows little, then it will appear to imitate, because it is choosing from a limited range of possibilities.

Similarly, an IDyOT with appropriate sense and action may be expected to exhibit motor babbling, because of random production in the absence of other dynamic input. The motor babbling can be observed by proprioceptors. All this information can be recorded in memory, allowing the IDyOT to learn the relation between motor signals and their proprioceptive or visual correspondents, and thus build a model of its own physical capabilities.

This raises the interesting question of when an IDyOT should stop motor babbling. A top-down rule would be antithetical to the minimalist principles of the approach; so the solution should be sought in the Global Workspace. This is essentially a mechanism that should shut off inconsequential predictions from memory, much as children learn not to speak their inconsequential predictions as they get older, unless they are actually speaking.

8.5. A different kind of deep learning

An interesting relationship is that between IDyOT memory formation/consolidation and deep learning. IDyOT differs from deep learning as espoused by LeCun et al. [121], because it is essentially time-based, whereas more conventional deep networks do not have to be. Furthermore, the dual hierarchy of IDyOT, formed by distinction between episodic and semantic memory lends a symbolic aspect to the model: it is possible to point to a particular structure or symbol in the memory and ask for its meaning, in terms of the perceptual input from which it was derived. This is not generally possible with neural networks, though research proceeds in this direction. In the case of RBM-based networks, there is a very clear similarity: each layer of such a network identifies statistical correlations, just like IDyOT. However, these networks lack an explicit chunking mechanism. To understand this relationship fully would shed light in both directions: on IDyOT and on the networks with which it is compared.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

9. Conclusion: creativity in the information dynamics of thinking

IDyOT was originally motivated by the search for a computational cognitive model of human creativity. Perhaps disappointingly, once the theory is laid out as a whole, as here, creativity becomes a somewhat more everyday activity than sometimes assumed [215]. Wiggins et al. [231] argue for the importance of addressing small, everyday creativity in seeking a satisfactory account of the evolution of creative behaviour, and IDyOT theory takes the same line. In IDyOT, creativity is equated with prediction: the ability to predict a sequence that is novel with respect to an IDyOT's memory (as can many Markov models, [e.g., 154]), but also to predict a previously unvisited point in conceptual space that corresponds with a class of novel sequences, at any level of abstraction in the memory. Thus, unusually, IDyOT affords an account of the creation of abstract, vague or generic ideas, where points in relatively high-level, abstract conceptual spaces subtend sequences in lower level, more concrete ones; furthermore, a process can be proposed for instantiating those more detailed sequences based on traversal of the episodic memory structure. Perhaps this can be the beginning of an account of the creative experience described by Mozart (§2).

Much more work, both theoretical and practical, is required before the theory of the Information Dynamics of Thinking is complete. This paper has laid out a summary of the key points of the theory, drawing them together in a way that has not previously been possible. It has also outlined some novel aspects of the theory. The next steps will be to proceed with implementation experiments, and to being to fill in the remaining gaps in the IDyOT account of cognition.

Acknowledgements

The thinking presented here would not have taken place without the continual intellectual stimulation of collaboration with my research labs in the Department of Artificial Intelligence, University of Edinburgh, the Departments of Computing, City University, London, and Goldsmiths, University of London, and the School of Electronic Engineering and Computer Science, Queen Mary University of London. Collaboration with Kat Agres, Roger Dean, Irène Deliège, Jamie Forth, Tom Hedges, Stephen McGregor, Daniel Müllensiefen, Marcus Pearce, Dan Ponsford, Matthew Purver and Raymond Whorley has been particularly important in framing the current thinking. The UK Engineering and Physical Sciences Research Council made two grants on the Information and Neural Dynamics of Music (IDyOM; GR/S82213 & EP-H01294X), which allowed broader collaboration with Samer Abdallah, Joydeep Bhattacharya, Mark Plumbley, Keith Potter and Suzie Wilkins. The European Union made two grants, Concept Creation Technology and Learning to Create, which have allowed the unification of the ideas into the current whole, and which acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant numbers 611733 and 610859, respectively. These grants have enabled collaboration with many valued colleagues, of whom Darrell Conklin, François Pachet and Frank van der Velde have made particular contributions to the current work. Finally, Antonio Chella helped me to see that all this is about cognitive architecture and not just information processing.

References

[1] Agres Kat, Abdallah Samer, Pearce Marcus. Information-theoretic properties of auditory sequences dynamically influence expectation and memory. Cogn Sci 2017:1–34. https://doi.org/10.1111/cogs.12477.
[2] Allan RH, Wiggins GA, Müllensiefen D. Methodological considerations in studies of musical similarity. In: Proceedings of ISMIR; 2007.
[3] Anderson JR, Bower GH. Human associative memory. Washington, DC: Winston and Sons; 1973.
[4] Anderson John R. ACT: a simple theory of complex cognition. Am Psychol April 1996;51(4):355–65.
[5] Anderson JR. The architecture of cognition. Cambridge, MA: Harvard University Press; 1983.
[6] Atkinson RC, Shiffrin RM. Human memory: a proposed system and its control processes. In: Spence KW, Spence JT, editors. The psychology of learning and motivation, vol. 2. New York: Academic Press; 1968. p. 89–195.
[7] Baars Bernard J. A cognitive theory of consciousness. Cambridge University Press; 1988.
[8] Baddeley AD, Eysenck M, Anderson MC. Memory. 2nd edition. Hove, UK: Psychology Press; 2014.
[9] Baddeley Alan D. The psychology of memory. New York: Basic Books, Inc; 1976.
[10] Balzano GJ. The group-theoretic description of 12-fold and microtonal pitch systems. Comput Music J 1980;4(4):66–84.
[11] Berthold Michael R. Bisociative knowledge discovery. LNCS/LNAI, vol. 7250. Springer; 2012.
[12] Besold TR, Schorlemmer M, Smaill A, editors. Computational creativity research: towards creative machines. Atlantis thinking machines. Atlantis/Springer; 2015.
[13] Biederman Irving, Vessel Edward A. Perceptual pleasure and the brain. Am Sci 2006;94:247–53. May–June.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

[14] Binsted K, Pain H, Ritchie G. Children's evaluation of computer-generated punning riddles. Pragmat Cogn 1997;5(2):309–58.
[15] Bod Rens. Memory-based models of melodic analysis: challenging the gestalt principles. J New Music Res 2001;30(1):27–37.
[16] Bode S, He AH, Soon CS, Trampel R, Turner R, Haynes J-D. Tracking the unconscious generation of free decisions using ultra-high field fMRI. PLoS ONE 2011;6(6):e21612. https://doi.org/10.1371/journal.pone.0021612.
[17] Boden MA. Creativity and artificial intelligence. Artif Intell J 1998;103:347–56.
[18] Boden MA. The creative mind: myths and mechanisms. 2nd edition. London, UK: Routledge. ISBN 0-349-10469-7, 2004.
[19] Boden Margaret. Artificial intelligence and natural man. Harvester Press. ISBN 0-85527-700-9, 1977.
[20] Bown Ollie, Wiggins Geraint A. From maladaptation to competition to cooperation in the evolution of musical behaviour. In: Special issue on evolution of music. Mus Sci 2009;13:387–411. https://doi.org/10.1177/1029864909013002171.
[21] Brachman RJ, Levesque HJ, editors. Readings in knowledge representation. Morgan Kaufmann; 1985.
[22] Brachman RJ, Levesque H. Knowledge representation. London: MIT Press; 1992.
[23] Bregman AS. Auditory scene analysis. Cambridge, MA: The MIT Press; 1990.
[24] Bridle JS, Brown MD. An experimental automatic word-recognition system. JSRU report 1003, Ruislip, England: Joint Speech Research Unit; 1974.
[25] Cardoso Amílcar, Veale Tony, Wiggins Geraint A. Converging on the divergent: the history (and future) of the international joint workshops in computational creativity. AI Mag 2010;30(3):15–22.
[26] Cariani Peter. Temporal coding of periodicity pitch in the auditory system: an overview. Neural Plast 1999;6(4).
[27] Chalmers DJ. On implementing a computation. Minds Mach 1994;4:391–402.
[28] Chalmers DJ. The conscious mind: in search of a fundamental theory. OUP; 1996.
[29] Chella Antonio. A cognitive architecture for music perception exploiting conceptual spaces. In: Applications of conceptual spaces: the case for geometric knowledge representation. Synthese library, vol. 359. Springer; 2015.
[30] Chella Antonio, Coradeschi Silvia, Frixione Marcello, Saffiotti Alessandro. Perceptual anchoring via conceptual spaces. In: Proceedings of the AAAI-04 workshop on anchoring symbols to sensor data; 2004.
[31] Chella Antonio, Dindo Haris, Infantino Ignazio. Imitation learning and anchoring through conceptual spaces. Appl Artif Intell 2007;21(4):343–59.
[32] Chella Antonio, Frixione Marcello, Gaglio Salvatore. A cognitive architecture for robot self-consciousness. Artif Intell Med 2008;44(2):147–54. https://doi.org/10.1016/j.artmed.2008.07.003.
[33] Chew E. Mathematical and computational modeling of tonality: theory and applications. International series on operations research and management science, vol. 204. New York, NY: Springer; 2014.
[34] Chiappe Penny, Schmuckler Mark A. Phrasing influences the recognition of melodies. Psychon Bull Rev 1997;4(2):254–9.
[35] Chomsky N. Syntactic structures. The Hague: Mouton; 1957.
[36] Christiansen Morten H, Chater Nick. The now-or-never bottleneck: a fundamental constraint on language. Behav Brain Sci 2016;39. https://doi.org/10.1017/S0140525X1500031X.
[37] Clark Andy. Whatever next? predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 2013;36(6):181–204. https://doi.org/10.1017/S0140525X12000477. http://journals.cambridge.org/article_S0140525X12000477.
[38] Cleary J, Witten I. Data compression using adaptive coding and partial string matching. IEEE Trans Commun Apr 1984;32(4):396–402. https://doi.org/10.1109/TCOM.1984.1096090.
[39] Colton S, Bundy A, Walsh T. Automatic invention of integer sequences. In: Proceedings of AAAI 2000. AAAI Press/MIT Press; 2000. p. 558–63.
[40] Colton Simon. Refactorable numbers—a machine invention. J Integer Seq 1999;2:99.1.2.
[41] Colton Simon, Wiggins Geraint A. Computational creativity: the final frontier? In: de Raedt L, Bessiere C, Dubois D, Doherty P, editors. Proceedings of ECAI frontiers; 2012.
[42] Conklin D, Witten IH. Multiple viewpoint systems for music prediction. J New Music Res 1995;24:51–73.
[43] Conscious Software Research Group. How minds work: a cognitive theory of everything. Web publication. http://csrg.cs.memphis.edu/tutorial/, 2006.
[44] Crawford T, Iliopoulos CS, Winder R, Yu H-F. Approximate musical evolution. In: Wiggins GA, editor. Proceedings of the AISB'99 symposium on musical creativity. ISBN 1-902956-00-1, 1999.
[45] Cross Ian. The evolutionary nature of musical meaning. Music Sci 2009;13(2 suppl):179–200.
[46] Cross Ian, Woodruff Ghofur Eliot. Music as a communicative medium. In: Botha Rudie, Knight Chris, editors. The prehistory of language. Oxford: Oxford University Press; 2008. p. 77–98.
[47] Csikszentmihalyi M. Creativity: flow and the psychology of discovery and invention. New York: HarperCollins; 1996.
[48] Currie Adrian, Killin Anton. Musical pluralism and the science of music. Eur J Philos Sci 2015:1–22. https://doi.org/10.1007/s13194-015-0123-z.
[49] Deliège I, Mélen M. Cue abstraction in the representation of musical form. In: Perception and cognition of music. Hove, England: Psychology Press; 1997.
[50] Deliège Irene. Grouping conditions in listening to music: an approach to Lerdahl and Jackendoff's grouping preference rules. Music Percept 1987;4:325–60.
[51] Derefeldt Gunilla, Swartling Tiina, Berggrund Ulf, Bodrogi Peter. Cognitive color. Color Res Appl 2004;29(1):7–19. https://doi.org/10.1002/col.10209.
[52] d'Inverno M, Luck M. Understanding agent systems. Springer series on agent technology. Berlin, Heidelberg: Springer. ISBN 9783662046074, 2013. https://books.google.co.uk/books?id=E9ioCAAAQBAJ.
[53] Duch W. Brain-inspired conscious computing architecture. J Mind Behav 2005;26(1–2):1–22.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

[54] Duch W, Oentaryo RJ, Pasquier M. Cognitive architectures: where do we go from here? In: Proceedings of the conference on artificial general intelligence; 2008. p. 122–36.
[55] Eck Douglas, Schmidhuber Jürgen. Finding temporal structure in music: blues improvisation with LSTM recurrent networks. In: Bourlard H, editor. Proceedings of IEEE workshop on neural networks for signal processing XII. New York: IEEE; 2002. p. 747–56.
[56] Edelman Gerald M, Gally Joseph A, Baars Bernard J. Biology of consciousness. Front Psychol 2011;2. https://doi.org/10.3389/fpsyg.2011.00004.
[57] Egermann Hauke, Pearce Marcus T, Wiggins Geraint A, McAdams Stephen. Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. Cogn Affect Behav Neurosci 2013;13(3):533–53. https://doi.org/10.3758/s13415-013-0161-y.
[58] Eysenck M, Keane MT. Cognitive psychology: a student's handbook. 3 edition. Psychology Press; 1995.
[59] Firestone Chaz, Scholl Brian J. Cognition does not affect perception: evaluating the evidence for 'top-down' effects. Behav Brain Sci, FirstView 2015;1–72:7. https://doi.org/10.1017/S0140525X15000965. http://journals.cambridge.org/article_S0140525X15000965.
[60] Fitch W Tecumseh. Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. Front Syst Neurosci 2013;7:68. https://doi.org/10.3389/fnsys.2013.00068. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3813894/.
[61] Tecumseh Fitch W, Hauser Marc D, Chomsky Noam. The evolution of the language faculty: clarifications and implications. Cognition 2005;97:179–210.
[62] Foerde K, Poldrack RA. Procedural learning in humans. In: Squire LR, editor. The new encyclopedia of neuroscience, vol. 7. Oxford, UK: Academic Press; 2009. p. 1083–91.
[63] Forth James Christopher. Cognitively-motivated geometric methods of pattern discovery and models of similarity in music. PhD thesis, Goldsmiths, University of London; 2012.
[64] Forth Jamie, Wiggins Geraint, McLean Alex. Unifying conceptual spaces: concept formation in musical creative systems. Minds Mach 2010;20:503–32. http://dx.doi.org/10.1007/s11023-010-9207-x.
[65] Forth Jamie, Agres Kat, Purver Matthew, Wiggins Geraint A. Entraining IDyOT: timing in the information dynamics of thinking. Front Psychol 2016;7:1575. https://doi.org/10.3389/fpsyg.2016.01575. http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01575.
[66] Fournié Edouard. Essai de Psychologie: La Bête et L'homme. France: Nabu Press; 1887. Originally published 1887; this edition 2010.
[67] Franklin A, amd Kelemen SP, McCauley L. IDA: a cognitive agent architecture. In: DiCesare F, Jafari MA, editors. Proceedings of the 1998 IEEE international conference on systems, man, and cybernetics: intelligent systems for humans in a cyberworld. Los Alamitos, CA: IEEE Computer Society Press; 1998.
[68] Franklin S, Strain S, Snaider J, McCall R, Faghihi U. Global workspace theory, its LIDA model and the underlying neuroscience. Biol Inspir Cognit Archit 2012;1:32–43. https://doi.org/10.1016/j.bica.2012.04.001.
[69] Franklin S, Madl T, D'Mello S, Snaider LIDA J. A systems-level architecture for cognition, emotion, and learning. IEEE Trans Auton Ment Dev 2014;6(1):19–41. https://doi.org/10.1109/TAMD.2013.2277589.
[70] Franklin Stan, Patterson FG Jr. The LIDA architecture: adding new modes of learning to an intelligent, autonomous, software agent. In: Integrated design and process technology. Society for Design and Process Science; 2006. http://ccrg.cs.memphis.edu/assets/papers/zo-1010-lida-060403.pdf.
[71] Friston K. The free-energy principle: a unified brain theory? Nat Rev Neurosci 2010;11(2):127–38. https://doi.org/10.1038/nrn2787.
[72] Gärdenfors Peter. Conceptual spaces: the geometry of thought. Cambridge, MA: MIT Press; 2000.
[73] Gärdenfors Peter. Geometry of meaning. Cambridge, MA: MIT Press; 2014.
[74] Gaussier P, Moga S, Quoy M, Banquet JP. From perception-action loops to imitation processes: a bottom-up approach of learning by imitation. Appl Artif Intell 1998;12(7–8):701–27. https://doi.org/10.1080/088395198117596.
[75] Gervàs Pablo. Computational approaches to storytelling and creativity. AI Mag 2015;30(3). https://doi.org/10.1609/aimag.v30i3.2250.
[76] Gobet Fernand, Lane Peter CR, Croker Steve, Cheng Peter C-H, Jones Gary, Oliver Iain, et al. Chunking mechanisms in human learning. Trends Cogn Sci 2001;5(6):236–43.
[77] Goldwater S, Griffiths TL, Johnson M. Contextual dependencies in unsupervised word segmentation. In: Calzolari N, Cardie C, Isabelle P, editors. Proceedings of the 21st international conference on computational linguistics and the 44th annual meeting of the ACL. East Stroudsburg, PA, USA: Association for Computational Linguistics; 2006. p. 673–80. http://www.stanford.edu/~sgwater/papers/acl06.pdf.
[78] Grahn Jessica A. Neural mechanisms of rhythm perception: current findings and future perspectives. Top Cogn Sci 2012;4:585–606.
[79] Guilford JP. The nature of human intelligence. New York: McGraw–Hill; 1967.
[80] Hansen Niels Chr, Pearce Marcus T. Predictive uncertainty in auditory sequence processing. Front Psychol 2014;5(1052). https://doi.org/10.3389/fpsyg.2014.01052. http://www.frontiersin.org/auditory_cognitive_neuroscience/10.3389/fpsyg.2014.01052/abstract.
[81] Hartbauer Manfred, Kratzer Silvia, Steiner Klaus, Römer Heiner. Mechanisms for synchrony and alternation in song interactions of the bushcricket Mecopoda elongata (Tettigoniidae: Orthoptera). J Comp Physiol A 2005;191(2):175–88. https://doi.org/10.1007/s00359-004-0586-4.
[82] Hawkins Sarah. Roles and representations of systematic fine phonetic detail in speech understanding. J Phon 2003;31(3–4):373–405. https://doi.org/10.1016/j.wocn.2003.09.006. http://www.sciencedirect.com/science/article/pii/S0095447003000597. Temporal Integration in the Perception of Speech.
[83] Hawkins Sarah, Smith Rachel. Polysp: a polysystemic, phonetically-rich approach to speech understanding. It J Linguist - Riv Linguist 2001;13:99–188. http://www.ling.cam.ac.uk/sarah/docs/hawkins-smith-01.pdf.
[84] Hedges Thomas, Wiggins Geraint A. The prediction of merged attributes with multiple viewpoint systems. J New Music Res 2016;45(4):314–32. https://doi.org/10.1080/09298215.2016.1205632.
[85] Hedges Thomas, Roy Pierre, Pachet François. Predicting the composer and style of jazz chord progressions. J New Music Res 2014;43(3):276–90. https://doi.org/10.1080/09298215.2014.925477.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

[86] Hedges Thomas W. Advances in multiple viewpoint systems and applications in modelling higher order musical structure. PhD thesis, Queen Mary University of London; 2017.
[87] Hélie Sébastien, Sun Ron. Incubation, insight, and creative problem solving: a unified theory and a connectionist model. Psychol Rev 2010;117(3):994–1024.
[88] Herremans D, Weisser S, Sörensen K, Conklin D. Generating structured music for bagana using quality metrics based on Markov models. Expert Syst Appl 2015;42:7424–35.
[89] Hodges Andrew. Alan Turing: the enigma. London: Vintage; 1992.
[90] Holmes Edward. The life of Mozart: including his correspondence. Cambridge library collection. Cambridge, UK: Cambridge University Press; 2009.
[91] Horst Steven. The computational theory of mind. In: Zalta Edward N, editor. The Stanford encyclopedia of philosophy. Center for the Study of Language and Information (CSLI), Stanford University; 2011.
[92] Huron David. Sweet anticipation: music and the psychology of expectation. Cambridge, MA: Bradford Books, MIT Press; 2006.
[93] Jackendoff Ray. Foundations of language: brain, meaning, grammar, evolution. Oxford, UK: OUP; 2002.
[94] Jackson John V. Perception, logic and action through structured motivated associative pandemonium. In: Butz MV, Sigaud O, Pezzulo G, Baldassarre G, editors. Anticipatory behavior in adaptive learning systems: from brains to individual and social behavior. LNCS/LNAI, vol. 4520. 2007.
[95] Juslin PN, Sloboda JA. Handbook of music and emotion: theory, research, applications. Affective science. Oxford University Press; 2010. http://books.google.co.uk/books?id=1N85AQAAIAAJ.
[96] Kemp Charles, Tenenbaum Joshua B. The discovery of structural form. Proc Natl Acad Sci USA 2008;105(31):10687–92. https://doi.org/10.1073/pnas.0802631105. http://www.pnas.org/content/105/31/10687.abstract.
[97] Kemp Charles, Perfors Amy, Tenenbaum Joshua B. Learning domain structures. In: Proceedings of the 26th annual conference of the cognitive science society; 2004. p. 672–7.
[98] Kieras David E. EPIC architecture principles of operation. Technical report, University of Michigan; 2004. http://ix.cs.uoregon.edu/~hornof/downloads/EPIC_Tutorial/EPICPrinOp.pdf.
[99] Kieras David E, Meyer David E. Predicting human performance in dual-task tracking and decision making with computational models using the epic architecture. In: Proceedings of the 1995 international symposium on command and control research and technology. National Defense University; 1995.
[100] Kieras David E, Meyer David E. An overview of the epic architecture for cognition: performance with application to human–computer interaction. Hum-Comput Interact 1997;12:391–438.
[101] Kipling Rudyard. Just so stories. New edition. Wordsworth Editions Ltd.; 1993.
[102] Kirby S. Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In: Knight C, editor. The evolutionary emergence of language: social function and the origins of linguistic form. Cambridge University Press; 2000. p. 303–23. http://groups.lis.illinois.edu/amag/langev/paper/kirby00syntaxWithout.html.
[103] Kirby S, Smith K, Cornish H. Language, learning and cultural evolution: how linguistic transmission leads to cumulative adaptation. In: Cooper R, Kempson R, editors. Language in flux: dialogue coordination, language variation, change and evolution. College Publications; 2008.
[104] Koelsch Stefan, Kasper Elisabeth, Sammler Daniela, Schulze Katrin, Gunter Thomas, Friederici Angela D. Music, language and meaning: brain signatures of semantic processing. Nat Neurosci 2004;7(3):302–7.
[105] Koestler Arthur. The act of creation. London, UK: Hutchinson; 1976.
[106] Krumhansl CL. The psychological representation of musical pitch in a tonal context. Cogn Psychol 1979;11:346–74.
[107] Krumhansl CL, Shepard RN. Quantification of the hierarchy of tonal functions within a diatonic context. J Exp Psychol Hum Percept Perform 1979;5(4):579–94.
[108] Kuhl Patricia K. A new view of language acquisition. Proc Natl Acad Sci USA 2000;97(22):11850–7. https://doi.org/10.1073/pnas.97.22.11850. http://www.pnas.org/content/97/22/11850.abstract.
[109] Laird JE, Newell A, Rosenbloom PS. Soar: an architecture for general intelligence. Artif Intell 1987;33(1):1–64.
[110] Laird John E. Extending the soar cognitive architecture. In: Proceedings of the 2008 conference on artificial general intelligence 2008: proceedings of the first AGI conference. Amsterdam, The Netherlands: IOS Press. ISBN 978-1-58603-833-5, 2008. p. 224–35. http://dl.acm.org/citation.cfm?id=1566174.1566195.
[111] Lakatos I. Falsification and the methodology of scientific research programmes. In: Lakatos I, Musgrave A, editors. Criticism and the growth of knowledge. Cambridge, UK: Cambridge University Press; 1970. p. 91–196.
[112] Lamont A, Eerola T, editors. Musicae scientiae, special issue on music and emotion, vol. 15:1. Sage; 2011.
[113] Lanchantin Pierre, Morris Andrew C, Rodet Xavier, Veaux Christophe. Automatic phoneme segmentation with relaxed textual constraints. In: Proceedings of language resources and evaluation conference (LREC); 2008.
[114] Langley P. An adaptive architecture for physical agents. In: Proceedings of the 2005 IEEE/WIC/ACM international conference on web intelligence; Sept 2005. p. 18–25.
[115] Langley P, Laird JE. Cognitive architectures: research issues and challenges. Technical report, Palo Alto, CA: Institute for the Study of Learning and Expertise; 2002.
[116] Large EW, Kolen JF. Resonance and the perception of musical meter. Connect Sci 1994;6(2–3):177–208. ftp://archive.cis.ohio-state.edu/pub/neuroprose/large.resonance.ps.Z.
[117] Large EW, Almonte F, Velasco M. A canonical model for gradient frequency neural networks. Physica D 2010.
[118] Large Edward W, Riess Jones Mari. The dynamics of attending: how people track time-varying events. Psychol Rev 1999;106(1):119–59.
[119] Lashley K. The problem of serial order in behavior. In: Jeffress LA, editor. Cerebral mechanisms in behavior. New York: Wiley; 1951. p. 123–47.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

[120] Lebiere C, Anderson JR. A connectionist implementation of the ACT-R production system. In: Proceedings of the fifteenth annual conference of the cognitive science society. Mahwah, NJ: Lawrence Erlbaum Associates; 1993. p. 635–40.
[121] LeCun Yann, Bengio Yoshua, Hinton Geoffrey. Deep learning. Nature 2015;521:436. http://dx.doi.org/10.1038/nature14539. EP–,05.
[122] Lehman Jill Fain, Laird John, Rosenbloom Paul. A gentle introduction to SOAR, an architecture for human cognition: 2006 update. http://ai.eecs.umich.edu/soar/sitemaker/docs/misc/GentleIntroduction-2006.pdf. Web publication, 2006.
[123] Lemström Kjell, Wiggins Geraint A. Formalizing invariances for content-based music retrieval. In: Tzanetakis George, Hirata Keiji, editors. Proceedings of ISMIR 2009; 2009.
[124] Lerdahl F, Jackendoff RS. A generative theory of tonal music. Cambridge, MA: MIT Press; 1983.
[125] Lerdahl F, Krumhansl CL. Modeling tonal tension. Music Percept 2007;24:329–66.
[126] Lerdahl Fred. Tonal pitch space. Music Percept 1988;5(3):315–50.
[127] Lerdahl Fred. Tonal pitch space. Oxford: Oxford University Press; 2001.
[128] Levitin Daniel J. Absolute memory for musical pitch: evidence from the production of learned melodies. Percept Psychophys 1994;56(6):414–23.
[129] London Justin. Hearing in time: psychological aspects of musical metre. Oxford, UK: Oxford University Press; 2004.
[130] Longuet-Higgins HC. Letter to a musical friend. Mus Rev 1962;23:244–827180,-.
[131] Longuet-Higgins HC. Second letter to a musical friend. Mus Rev 1962;23:271–80.
[132] MacKay David JC. Information theory, inference, and learning algorithms. Cambridge, UK: Cambridge University Press; 2003. http://www.inference.phy.cam.ac.uk/mackay/itila/.
[133] Marr David. Vision: a computational approach. San Francisco: Freeman & Co.; 1982.
[134] McAdams S, Saariaho K. Qualities and functions of musical timbre. In: Proceedings of the 1985 international computer music conference, Vancouver. Berkeley, CA: Computer Music Association; 1985. p. 367–74.
[135] McAdams Stephen. Music: a science of the mind? Contemp Mus Rev 1987;2:1–61.
[136] McCorduck Pamela. AARON'S CODE: meta-art, artificial intelligence and the work of Harold Cohen. Freeman; 1991.
[137] McGurk Harry, MacDonald John. Hearing lips and seeing voices. Nature 1976;264(5588):746–8. http://dx.doi.org/10.1038/264746a0. 12.
[138] Merchant H, Grahn JA, Trainor LJ, Rohrmeier M, Fitch WT. Finding the beat: a neural perspective across human and non-human primates. Philos Trans R Soc B 2015;370(20140093). https://doi.org/10.1098/rstb.2014.0093.
[139] Merker Björn. From probabilities to percepts a subcortical "global best estimate buffer" as locus of phenomenal experience. In: Edelman Shimon, Fekete Tomer, Zach Neta, editors. Being in time: dynamical models of phenomenal experience. John Benjamins Publishing Company; 2012.
[140] Merker Bjorn. The efference cascade, consciousness, and its self: naturalizing the first person pivot of action control. Front Psychol 2013;4(501). https://doi.org/10.3389/fpsyg.2013.00501.
[141] Minsky M. The society of mind. New York, NY: Simon and Schuster Inc.; 1985.
[142] Monteith Kristine, Brown Bruce, Ventura Dan, Martinez Tony. Automatic generation of music for inducing physiological response. In: Proceedings of the 35th annual meeting of the cognitive science society; 2013. p. 3098–103.
[143] Moore Brian CJ. An introduction to the psychology of hearing. 2nd edition. London: Academic Press; 1982.
[144] Muggleton Stephen. Inductive logic programming. New Gener Comput 1991;8(4):295–318. https://doi.org/10.1007/BF03037089.
[145] Nattiez J-J. Is a descriptive semiotics of music possible? Lang Sci 1972;23.
[146] Nattiez J-J. Fondements d'une sémiologie de la musique. Paris: Union Générale d'Editions; 1975.
[147] Newell Allen. Unified theories of cognition. Cambridge, MA: Harvard UP; 1994.
[148] Oakley David A, Halligan Peter W. Chasing the rainbow: the non-conscious nature of being. Front Psychol 2017;8:1924. https://doi.org/10.3389/fpsyg.2017.01924. https://www.frontiersin.org/article/10.3389/fpsyg.2017.01924.
[149] Pachet F. Playing with virtual musicians: the continuator in practice. IEEE Multimed 2002;9(3):77–82.
[150] Patel AD, Balaban E. Temporal patterns of human cortical activity reflect tone sequences structure. Nature 2000;404:80–4.
[151] Patel Aniruddh D. Music, language, and the brain. Oxford: Oxford University Press; 2008.
[152] Patel Aniruddh D, Daniele Joseph R. An empirical comparison of rhythm in language and music. Cognition 2003;87:B35–45.
[153] Patel Aniruddh D, Iversen John R, Bregman Micah R, Schulz Irena. Experimental evidence for synchronization to a musical beat in a nonhuman animal. Curr Biol 2009;19(10):827–30. 05, http://linkinghub.elsevier.com/retrieve/pii/S0960982209008902.
[154] Pearce MT, Wiggins GA. Evaluating cognitive models of musical composition. In: Cardoso A, Wiggins GA, editors. Proceedings of the 4th international joint workshop on computational creativity. London: Goldsmiths, University of London; 2007. p. 73–80.
[155] Pearce MT, Conklin D, Wiggins GA. Methods for combining statistical models of music. In: Wiil UK, editor. Computer music modelling and retrieval. Heidelberg, Germany: Springer Verlag; 2005. p. 295–312. http://www.doc.gold.ac.uk/~mas02gw/papers/cmmr04.pdf.
[156] Pearce MT, Herrojo Ruiz M, Kapasi S, Wiggins GA, Bhattacharya J. Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation. NeuroImage 2010;50(1):303–14. https://doi.org/10.1016/j.neuroimage.2009.12.019.
[157] Pearce Marcus T. The construction and evaluation of statistical models of melodic structure in music perception and composition. PhD thesis, London, UK: Department of Computing, City University, London; 2005.
[158] Pearce Marcus T, Wiggins Geraint A. Expectation in melody: the influence of context and learning. Music Percept 2006;23(5):377–405.
[159] Pearce Marcus T, Wiggins Geraint A. Auditory expectation: the information dynamics of music perception and cognition. Top Cogn Sci 2012;4(4):625–52. https://doi.org/10.1111/j.1756-8765.2012.01214.x.
[160] Pearce Marcus T, Müllensiefen Daniel, Wiggins Geraint A. The role of expectation and probabilistic learning in auditory boundary perception: a model comparison. Perception 2010;39(10):1367–91.
[161] Pease A, Colton S. Computational creativity theory: inspirations behind the face and idea models. In: Proceedings of the international conference on computational creativity; 2011.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

[162] Pérez Y, Pérez R. A computer-based model for collaborative narrative generation. Cogn Syst Res 2015. https://doi.org/10.1016/j.cogsys.2015.06.002.
[163] Perlovsky Leonid. Music: passions and cognitive functions. San Diego, CA: Academic Press; 2017.
[164] Pinker Steven. The language instinct. Perrenial; 1995.
[165] Quinlan JR. Induction of decision trees. Mach Learn 1986;1(1):81–106.
[166] Ramamurthy U, Baars B, D'Mello SK, Franklin S. LIDA: a working model of cognition. In: Del Missier Fabio, Fum Danilo, Stocco Andrea, editors. Proceedings of the 7th international conference on cognitive modeling. Trieste, Italy: Edizioni Goliardiche; 2006. p. 244–9.
[167] Repp Bruno H. Temporal evolution of the phase correction response in synchronization of taps with perturbed two-interval rhythms. Exp Brain Res 2011;208:89–101. https://doi.org/10.1007/s00221-010-2462-5.
[168] Ritchie Graeme. Some empirical criteria for attributing creativity to a computer program. Minds Mach 2007;17(1):67–99.
[169] Roberson D, Davidoff J, Davies IRL, Shapiro LR. Colour categories and category acquisition in Himba and English. In: Pitchford N, Bingham C, editors. Progress in colour studies. Amsterdam: John Benjamins; 2006. p. 159–72.
[170] Rohrmeier M, Cross I. Artificial grammar learning of melody is constrained by melodic inconsistency: Narmour's principles affect melodic learning. PLoS ONE 2013;8(7):e66174. https://doi.org/10.1371/journal.pone.0066174.
[171] Rohrmeier Martin, Zuidema Willem, Wiggins Geraint A, Scharff Constance. Principles of structure building in music, language and animal song. Philos Trans R Soc Lond B, Biol Sci 2015;370(1664). https://doi.org/10.1098/rstb.2014.0097.
[172] Russell S, Norvig P. Artificial intelligence—a modern approach. New Jersey: Prentice Hall; 1995.
[173] Ruwet N. Language, musique, poésie. Paris: Editions du Seuil; 1972.
[174] Saffran JR, Griepentrog GJ. Absolute pitch in infant auditory learning: evidence for developmental reorganization. Dev Psychol 2001;37(1):74–85. http://www.waisman.wisc.edu/infantlearning/publications/DevPsychAP.pdf.
[175] Saunders R. Curious design agents and artificial creativity. PhD thesis, Sydney, Australia: The University of Sydney; 2001.
[176] Schachner Adena, Brady Timothy F, Pepperberg Irene M, Hauser Marc D. Spontaneous motor entrainment to music in multiple vocal mimicking species. Curr Biol 2009;19(10):831–6. p. 05, http://linkinghub.elsevier.com/retrieve/pii/S0960982209009154.
[177] Schenker Heinrich. Das Meisterwerk in der Musik. Munich: Drei Maksen Verlag; 1930. In 3 volumes, published 1925, 1926, 1930.
[178] Schmidhuber J. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). IEEE Trans Auton Ment Dev sept. 2010;2(3):230–47. https://doi.org/10.1109/TAMD.2010.2056368.
[179] Schmidhuber Jürgen. Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. Connect Sci 2006;18(2):173–87.
[180] Shanahan Murray. Embodiment and the inner life: cognition and consciousness in the space of possible minds. OUP; 2010.
[181] Shannon CE. A mathematical theory of communication. Bell Syst Tech J July and October 1948;27:379–423. 623–56.
[182] Shepard RN. Circularity in judgments of relative pitch. J Acoust Soc Am 1964;36:2346–53.
[183] Shepard RN. Structural representations of musical pitch. In: Deutsch D, editor. Psychology of music. New York: Academic Press; 1982. p. 343–90.
[184] Smaill A, Wiggins GA, Miranda E. Music representation—between the musician and the computer. In: Smith M, Wiggins G, Smaill A, editors. Music education: an artificial intelligence perspective. London: Springer; 1993. p. 108–19. http://www.soi.city.ac.uk/~geraint/papers/WCAIEd93.pdf.
[185] Smith Kenny. Learning biases for the evolution of linguistic structure: an associative network model. Berlin, Heidelberg: Springer. ISBN 978-3-540-39432-7, 2003. p. 517–24.
[186] Soon Chun Siong, Brass Marcel, Heinze Hans-Jochen, Haynes John-Dylan. Unconscious determinants of free decisions in the human brain. Nat Neurosci 2008;11(5):543–5. http://dx.doi.org/10.1038/nn.2112. 05.
[187] Speidel Gisela E, Nelson Keith E, editors. The many faces of imitation in language learning. New York, NY: Springer Science & Business Media; 2012.
[188] Sproat R, Shih C, Gale W, Chang N. A stochastic finite-state word-segmentation algorithm for Chinese. In: Proceedings of the 32nd annual meeting of the association for computational linguistics; 1994. p. 66–73.
[189] Sun Ron. Duality of the mind: a bottom-up approach toward cognition. Mahwah, NJ: Erlbaum; 2002.
[190] Sun Ron. A detailed specification of CLARION 5.0. Technical report, Rensselaer Polytechnic Institute; 2003. http://www.cogsci.rpi.edu/~rsun/sun.tutorial.pdf. [Accessed 24 July 2015].
[191] Tan N, Aiello R, Bever TG. Harmonic structure as a determinant of melodic organization. Mem Cogn 1981;9(5):533–9.
[192] Tenenbaum Joshua B, Kemp Charles, Griffiths Thomas L, Goodman Noah D. How to grow a mind: statistics, structure, and abstraction. Science 2011;331(6022):1279–85. https://doi.org/10.1126/science.1192788. http://www.sciencemag.org/content/331/6022/1279.abstract.
[193] Toiviainen Petri, editor. Musicae scientiae. Special issue on musical similarity. ESCOM, vol. 11. 2007. p. 2.
[194] Tononi G, Edelman GM. Consciousness and complexity. Science 1998;282(5395):1846–51.
[195] Tononi Giulio. An information integration theory of consciousness. BMC Neurosci 2004;5(42). https://doi.org/10.1186/1471-2202-5-42.
[196] Tulving E. Episodic and semantic memory. In: Tulving E, Donaldson W, editors. Organization of memory. New York: Academic Press; 1972. p. 381–403.
[197] Turing Alan. Computing machinery and intelligence. Mind 1950;LIX(236):433–60.
[198] Turner Mark, Fauconnier Gilles. Conceptual integration and formal expression. Metaphor Symb Act 1995;10(3):183–203.
[199] Tversky A. Features of similarity. Psychol Rev 1977;84:327–52.
[200] Underwood BJ, Postman L. Extra-experimental sources of interference in forgetting. Psychol Rev 1960;67:73–95.
[201] van der Velde Frank. Communication, concepts and grounding. Neural Netw 2015;62:112–7. http://doc.utwente.nl/92568/. Special issue on communication and brain: emergent functions through inter-neuron and inter-brain communications.
[202] van der Velde Frank. In situ representations and access consciousness in neural blackboard or workspace architectures. Front Robot AI 2018;5(32). https://doi.org/10.3389/frobt.2018.00032. https://www.frontiersin.org/article/10.3389/frobt.2018.00032.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001

[203] van der Velde Frank, Forth Jamie, Nazareth Deniece S, Wiggins Geraint A. Linking neural and symbolic representation and processing of conceptual structures. Front Psychol 2017;8:1297. https://doi.org/10.3389/fpsyg.2017.01297. http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01297.
[204] van der Weij B, Pearce MT, Honing H. A probabilistic model of meter perception: simulating enculturation. Front Psychol 2017;8: 824. https://doi.org/10.3389/fpsyg.2017.00824.
[205] Varma Sashank. Criteria for the design and evaluation of cognitive architectures. Cogn Sci 2011;35:1329–51.
[206] Veale Tony. Exploding the creativity myth. New York, NY: Bloomsbury Academic; 2012.
[207] Vernon D, Metta G, Sandini G. A survey of artificial cognitive systems: implications for the autonomous development of mental capabilities in computational agents. IEEE Trans Evol Comput 2007;11(2):151–80.
[208] Vernon D, von Hofsten C, Fadiga L. A roadmap for cognitive development in humanoid robots. Cognitive systems monographs (COSMOS), vol. 11. New York, NY: Springer; 2011. chapter 5.
[209] Vernon David. Artificial cognitive systems: a primer. Cambridge, MA: MIT Press; 2014.
[210] Wallas Graham. The art of thought. New York: Harcourt Brace; 1926.
[211] Whorley Raymond P, Wiggins Geraint A, Rhodes Christophe, Pearce Marcus T. Multiple viewpoint systems: time complexity and the construction of domains for complex musical viewpoints in the harmonization problem. J New Music Res 2013;42(3):237–66. https://doi.org/10.1080/09298215.2013.831457. http://www.tandfonline.com/doi/abs/10.1080/09298215.2013.831457.
[212] Widdows D. Geometry and meaning. Stanford, CA: Center for the Study of Language and Information; 2004.
[213] Wiggins GA. Music, syntax, and the meaning of "meaning". In: Proceedings of the first symposium on music and computers, Corfu, Greece. Ionian University; 1998. p. 18–23.
[214] Wiggins GA. A preliminary framework for description, analysis and comparison of creative systems. J Knowl-Based Syst 2006;19(7):449–58. http://dx.doi.org/10.1016/j.knosys.2006.04.009.
[215] Wiggins GA, Forth JC. IDyOT: a computational theory of creativity as everyday reasoning from learned information. In: Besold Tarek Richard, Schorlemmer Marco, Smaill Alan, editors. Computational creativity research: towards creative machines. Atlantis Thinking Machines. Atlantis/Springer; 2015. p. 127–50.
[216] Wiggins GA, Harris M, Smaill A. Representing music for analysis and composition. In: Balaban M, Ebcioğlu K, Laske O, Lischka C, Soriso L, editors. Proceedings of the second workshop on AI and music. Menlo Park, CA: AAAI; 1989. p. 63–71. http://www.soi.city.ac.uk/~geraint/papers/EWAIM89.pdf.
[217] Wiggins GA, Müllensiefen D, Pearce MT. On the non-existence of music: why music theory is a figment of the imagination. Mus Sci, Discuss Forum 2010;5:231–55.
[218] Wiggins Geraint, Bhattacharya Joydeep. Mind the gap: an attempt to bridge computational and neuroscientific approaches to study creativity. Front Human Neurosci 2014;8(540). https://doi.org/10.3389/fnhum.2014.00540. http://www.frontiersin.org/human_neuroscience/10.3389/fnhum.2014.00540/abstract.
[219] Wiggins Geraint A. Searching for computational creativity. New Gener Comput 2006;24(3):209–22.
[220] Wiggins Geraint A. Models of musical similarity. Music Sci 2007;11:315–38. https://doi.org/10.1177/102986490701100112.
[221] Wiggins Geraint A. Semantic Gap?? Schemantic Schmap!! Methodological considerations in the scientific study of music. In: Proceedings of 11th IEEE international symposium on multimedia. ISBN 978-1-4244-5231-6, 2009. p. 477–82.
[222] Wiggins GA. Cue abstraction, paradigmatic analysis and information dynamics: towards music analysis by cognitive model. Mus Sci 2010:307–22. Special issue: understanding musical structure and form: papers in honour of Irène Deliège.
[223] Wiggins Geraint A. Computer models of (music) cognition. In: Rebuschat P, Rohrmeier M, Cross I, Hawkins J, editors. Language and music as cognitive systems. Oxford: Oxford University Press; 2011. p. 169–88.
[224] Wiggins Geraint A. The mind's chorus: creativity before consciousness. Cogn Comput 2012;4(3):306–19. https://doi.org/10.1007/s12559-012-9151-6.
[225] Wiggins Geraint A. Music, mind and mathematics: theory, reality and formality. J Math Mus 2012;6(2):111–23.
[226] Wiggins Geraint A. "I let the music speak": cross-domain application of a cognitive model of musical learning. In: Rebuschat Patrick, Williams John, editors. Statistical learning and language acquisition. Amsterdam, NL: Mouton De Gruyter; 2012. p. 463–95.
[227] Wiggins Geraint A. On the correctness of imprecision and the existential fallacy of absolute music. J Math Mus 2012;6(2):93–101.
[228] Wiggins Geraint A. The future of (mathematical) music theory. J Math Mus 2012;6(2):135–44.
[229] Wiggins Geraint A. Crossing the threshold paradox: modelling creative cognition in the global workspace. In: Proceedings of the international conference on computational creativity; 2012. p. 180. http://computationalcreativity.net/iccc2012/wp-content/uploads/2012/05/180-wiggins.pdf.
[230] Wiggins Geraint A. Consolidation as re-representation: revising the meaning of memory. 2018. In preparation.
[231] Wiggins Geraint A, Tyack Peter, Scharff Constance, Rohrmeier Martin. The evolutionary roots of creativity: mechanisms and motivations. Philos Trans R Soc Lond B, Biol Sci 2015;370(1664). https://doi.org/10.1098/rstb.2014.0099.
[232] Wixted John T. The psychology and neuroscience of forgetting. Annu Rev Psychol 2004;55:235–69.
[233] Ping Xiao, Toivonen Hannu, Gross Oskar, Cardoso Amílcar, Correia João, Machado Penousal, et al. Conceptual representations for computational concept generation. ACM Comput Surv 2018.

Please cite this article in press as: Wiggins GA. Creativity, information, and consciousness: The information dynamics of thinking. Phys Life Rev (2018), https://doi.org/10.1016/j.plrev.2018.05.001