The Information Dynamics of Thinking: a cognitive architecture for human creative cognition

Geraint A. Wiggins, Kat Agres, Jamie Forth, Matt Purver
Computational Creativity Lab, Queen Mary University of London

1 Introduction

We present IDyOT (Wiggins, 2012b), the Information Dynamics of Thinking, a novel cognitive architecture that instantiates the (much more abstract) Global Workspace Theory of Baars (1988), in a way motivated by a principle of efficiency in information processing. IDyOT simulates a hypothetical fundamental perceptual/cognitive loop, explicating bottom-up learning of concepts, the hierarchical structure in them, and the correlations between them from percepts. Even though the proposal is basically cognitive, embodiment is not ruled out, and the tight connection between perception and cognition potentially affords a future extension of the account to an embodied view.

2 Founding Principles of IDyOT

The architecture presented here is founded on several strict principles, which are presented as hypothetical bases for cognitive function. They are as follows. Some are in common with other theories. Some are more radical than others.

- A key function of mind/brain is to process information in the environment to maximise chances of survival.
- Because of the biological expense of nervous tissue and its use, energetic efficiency, and hence efficiency of use, is a key driver in the development of mind/brain functions.
- A particular advantage is conferred on organisms that can perceive the structure in perceived information.
- A particular advantage is conferred on organisms that can predict the next state of the environment from the current one.
- A particular advantage is conferred on organisms that can learn the relations between the current state of the environment and the next one.
- A mind/brain records sequences of percepts and uses them to anticipate the next state of the environment.
- An efficient mind/brain uses its expectations to assist perception in terms both of what is expected and when it is expected. It may sometimes be confused by them.
- An efficient mind/brain uses the most efficient representation possible for the data it has learned.
- An efficient mind/brain is capable of constructing new representations, and re-representing its data in terms of them.
- Imagination and creativity, at the lowest functional level, are the result of anticipation of unseen events, made possible by the construction of new representations that afford new meaning.

3 The Global Workspace and In Situ Representations

IDyOT is an instantiation of the Global Workspace Theory (GWT) of Bernard Baars (1988). Very broadly, Baars' Global Workspace (GW) is like an AI blackboard (Corkill, 1991); consciousness is conceptualised as the GW itself, and cognition is implemented by means of "generators" that view the GW and act accordingly. However, Baars' very generic architecture is presented in a von Neumann computational style, with information described as "moving" around the system, "into" the Global Workspace. IDyOT, in contrast, maintains static notions of memory, and, though it is intended for implementation on von Neumann machines, its theoretical operation may be explained in terms of *in situ* representations (van der Velde, 2013), which are substantially more brain-like: essentially, the memory store is also the processor. The Theatre of Consciousness metaphor of Hippolyte Taine (1871), extended by the addition of spotlights, elegantly captures this behaviour (Wiggins, 2012b).

Baars (1988) requires that generators somehow form coalitions to allow their information to enter the GW; this is the analog of attention in the GWT. But he acknowledges that this is paradoxical (the Threshold Paradox), because to form coalitions they need to communicate, but they can only communicate through the GW. IDyOT provides a solution to this problem by using information-theoretic measures to identify which current items occupy the GW (Wiggins, 2012a).

4 Memory formation

To implement this, IDyOT builds a hierarchical, literal memory of its perceived input. There is a system clock, at a frequency sufficient to simulate human time discrimination: around 40Hz. Input is sampled at this rate, and a multidimensional sequence of inputs is maintained. (In future work, forgetting will be included, since this is not a reasonably simulation of echoic memory, and is not a practical proposition even given a very large computer.)

IDyOT's primary operation is chunking of its perceptual input (Gobet et al., 2001). This is implemented as an emergent effect of passage of a chunk into the Global Workspace, as follows. Boundary entropy is well-established as a chunking principle in language (Servan-Schreiber and Anderson, 1990) and music (Pearce et al., 2010); IDyOT affords a hypothesis as to why this should be so: a significant increase in entropy (with respect to a given generator and its competitors) is what triggers entry into the GW (or in in situ terms, focuses the GW on the relevant chunk).

This begs the question, "entropy in what?" IDyOT memory encodes a first-order Markov Model of its perceptual input: a sudden increase in Shannon entropy (Shannon, 1948), as calculated from a first-order Markov model marks the beginning of a new segment. What makes IDyOT different from other Markov-related methods is that each segment corresponds with a symbol in a higher level model, similar to, but not the same as, a hierarchical hidden Markov Model. As IDyOT receives input, it constructs a layered sequence model, one for each simultaneous perceptual input stream (pitch, volume, colour, taste, proprioception, etc.), constructing new layers, each with its own alphabet, as it goes (Wiggins and Forth, 2015).

5 Consolidation and tethered meaning

All this requires some control over the size of alphabets, or they would simply expand indefinitely, and this is afforded by means of measures of similarity afforded by conceptual spaces (Gärdenfors, 2000) that supply a semantics over each alphabet, in terms of mutual similarity: each symbol indexes a region in the space, yielding a model of categorical perception van der Velde et al. (2016), and tethering inferred meaning to perceptual input (Sloman and Chappell, 2005). New chunks that match an existing symbol are labelled to match. The geometry of each conceptual space must be learned from the relevant data, as proposed by Kemp et al. (2004). In general, symbols will relate to sequences of varying length, and therefore spectral representations (in which a point represents, effectively, the Fourier transform of a trajectory through the space below it) are likely to be important (Chella, 2015; Chella et al., 2008).

Evidence for this approach of creating new representations to improve predictive accuracy is given by Pearce (2005) (see also Pearce and Wiggins, 2006, 2012), where multiple representations of data are chosen on the basis of reducing the average information content per symbol in the model.

6 Prediction

In implementation, each layer of IDyOT's memory maintains a first order Markov model of the sequence order, of downward implication, and of upward implication. Thus, predictions, expressed as distributions, can be created anywhere on the leading edge of the memory structure. These predictions are evaluated in terms of their information content, which is taken to simulate attention: attention may be thought of as a spotlight, picking out the nodes where there is the most information. Such nodes may be at any level of the structure, and so may be more or less abstract (like ideas). Given a node as a starting point, reasoning can proceed in any direction up, down, or along the memory structure.

7 Imagination and Creativity

If the reasoning is forwards in time, past the end of experienced memory, then IDyOT is being creative: because of the existence of the conceptual spaces, which are continuous representations of meaning, it is possible for IDyOT to reproduce existing symbols in new sequences, or to produce new symbols altogether, motivated by unexplored regions of conceptual space. Thus, IDyOT both produces new symbols and structures and is able to explain them in terms of its own experience.

8 Timing

In order to explain overt and explicit timed behaviour (entrained clapping, in music, for example) and much more subtle and implicit behaviour such as detailed temporal prediction of speech timing between speakers and listeners, IDyOT memory retains explicit durations for each of its symbols. Thus, the collection of durations experienced for a given symbol forms a distribution that predicts the duration of the percept corresponding with each symbol. These distributions can be used to time utterances, to predict the timing of perceived input, and to disambiguate (for example) between words of similar timbre. This aspect of IDyOT is explored in detail by Forth et al. (2016); it affords a new hypothetical account of the phenomenon of *entrainment* in music and language (Fitch, 2013).

9 Other points

Because IDyOT's memory maps directly to experience, it is learned incrementally. This has the following consequences: 1) meanings tethered to symbols depend on the order of events that the model learns; and, 2) it is necessary from time to time to re-optimise the model, after an extended phase of incremental learning. This latter is the Consolidation phase, described above. One consequence is that meanings can change retrospectively as the system learns. Consolidation is performed by attempting to find small changes in the mappings between symbols and their conceptual spaces, in such a way as to reduce the average information content of the overall model (van der Velde et al., 2016).

10 Conclusion

IDyOT constitutes an outline model of the basic processes of cognition. It fails to account for many important aspects of mind: for example, there is no attempt to model emotion, nor to explicitly model reflection. However, it forms a skeleton around which these effects may be added, and it has been designed with the intention of admitting such extensions freely and without preclusion.

References

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

Chella, A. (2015). A cognitive architecture for music perception exploiting conceptual spaces. In *Applications of Conceptual Spaces: The Case for Geometric Knowledge Representation*, number 359 in Synthese Library. Springer.

Chella, A., Frixione, M., and Gaglio, S. (2008). A cognitive architecture for robot self-consciousness. *Artificial Intelligence in Medicine*, 44(2):147–154.

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

Fitch, W. T. (2013). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience*, 7:68.

Forth, J., Agres, K., Purver, M., and Wiggins, G. (2016). Entraining idyot: timing in the information dynamics of thinking. *Frontiers in Psychology*, 7:1575.

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

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., and Pine, J. M. (2001). Chunking mechanisms in human learning. *TRENDS in Cognitive Sciences*, 5(6).

Kemp, C., Perfors, A., and Tenenbaum, J. B. (2004). Learning domain structures. *Proceedings of the 26th annual conference of the Cognitive Science Society*, pages 672–677.

Pearce, M. T. (2005). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. PhD thesis, Department of Computing, City University, London, London,UK.

Pearce, M. T., Müllensiefen, D., and Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception*, 39(10):1367–1391.

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

Pearce, M. T. and Wiggins, G. A. (2012). Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science*, 4(4):625–652.

Servan-Schreiber, E. and Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychologyrimeatal Psychology*, 16(4):592–608.

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

Sloman, A. and Chappell, J. (2005). The altricial-precocial spectrum for robots. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*.

Taine, H. (1871). *On Intelligence*. Savill, Edwards and Co., London, UK.

van der Velde, F. (2013). Consciousness as a process of queries and answers in architectures based on *in situ* representations. *International Journal of Machine Consciousness*, 05(01):27–45.

van der Velde, F., Forth, J., and Wiggins, G. A. (2016). Neural representation and processing of conceptual structures. *Frontiers in Psychology*. In preparation.

Wiggins, G. A. (2012a). Crossing the theshold paradox: Modelling creative cognition in the global workspace. In *Proceedings of the International Conference on Computational Creativity*.

Wiggins, G. A. (2012b). The mind's chorus: Creativity before consciousness. *Cognitive Computation*, 4(3):306–319.

Wiggins, G. A. and Forth, J. C. (2015). IDyOT: A computational theory of creativity as everyday reasoning from learned information. In *Computational Creativity Research: Towards Creative Machines*, Atlantis Thinking Machines. Atlantis/Springer.