Attentional Capture by Meaning, a Multi-level Modelling Study

Li Su, Howard Bowman (ls68@kent.ac.uk, hb5@kent.ac.uk)
Centre for Cognitive Neuroscience and Cognitive Systems, University of Kent, UK

Philip Barnard (philip.barnard@mrc-cbu.cam.ac.uk)
MRC Cognition and Brain Sciences Unit, Cambridge, UK

Abstract

We present a computational study of attentional capture by meaning, based on Barnard et al's key-distractor attentional blink task. We highlight a sequence of models, from an abstract black-box to a structurally detailed white-box model. Each of these models reproduces the major findings from the key-distractor blink task. We argue that such multi-level modelling gives greater confidence in the theoretical position encapsulated by these models.

Keywords: Attentional blink; LSA; semantic modulation; multi-level modeling.

Introduction

There are now many different approaches to the computational modelling of cognition, e.g. symbolic models (Newell 1990; Kieras and Meyer 1997), cognitive connectionist models (McLeod, Plunkett et al. 1998) and neurophysiologically prescribed connectionist models (O'Reilly and Munakata 2000). The relative value of different approaches is a hotly debated topic, with each presented as an alternative to the others, suggesting that they are in opposition to one another, e.g. (Fodor and Pylyshyn 1988; Hinton 1990). However, another perspective is that these reflect different levels of abstraction / explanation of the same system that are complementary, rather than fundamentally opposed.

Computer science, which has often been used as a metaphor in the cognitive modelling domain, gives a clear precedent for thinking in terms of multiple views of a single system. An illustration of this is what is now probably the most widely used design method, the Unified Modelling Language (UML) (Booch, Rumbaugh et al. 1999). It is not that this perspective has been completely lost on cognitive scientists; indeed, Marr famously elaborated a version of this position in his three levels of cognitive description (Marr 2000). However, despite Marr's observations, concrete modelling endeavours rarely, if ever, consider multiple abstraction levels in the same context and particularly how to relate those levels.

Multiple Level Cognitive Modelling

In this paper, we can distinguish between the following two levels of explanation of a cognitive phenomenon. Firstly, high-level abstract descriptions of the mathematical characteristics of a pattern of data, e.g. (Stewart, Brown et al. 2005). Secondly, low-level detailed models of the internal structure of a cognitive system, e.g. (Dehaene, Kerszberg et al. 1998). These two levels of explanation really reflect different capacities to observe systems; that is, the extent to which the system is viewed from outside or inside, i.e., as a black or white box. There are clear pros and cons to these forms of modelling, which we discuss now.

Black-box (Extensionalist) Modelling. With this approach, no assumptions are made about the internal structure of the system and there is no decomposition at all of the black-box into its constituent components. Thus, the point of reference for the modeller is the *externally visible behaviour*, e.g. the stimulus-response pattern. That is, such models are *extensionalist* in nature. A critical benefit of black-box modelling is that a minimal set of assumptions are made, especially in respect of the system structure. Consequently, there are less degrees of freedom and fewer hidden assumptions, making data fitting and parameter setting both well founded and, typically, feasible. For example, if the system can be described in closed form, key parameters can be determined by solving a set of equations, if not, computational search methods can be applied.

White-box (Intensionalist) Modelling. In contrast, the *internal* (decompositional) structure of the system is asserted with this approach. That is, such models are *intensionalist* in nature. Although we can bring theories of cognitive architecture and (increasingly) neural structure to bear in proposing white-box models, a spectrum of assumptions (necessarily) needs to be made. Furthermore, typically, many of these assumptions concern the internal structure of the system. While structurally detailed models of cognition are likely to be the most revealing (especially with the current emphasis on neurophysiological correlates), deduction from these models is more slippery and potentially less well founded. Most importantly, many assumptions, such as settings of key parameters, need to be made, many of which may, at best, require complex justification and, at worst, be effectively arbitrary. As a result, parameter setting and data fitting is more difficult and, arguably, less well founded with white-box models.

We can summarise then by saying that black-box modelling describes *what* a cognitive system does and it describes it in a relatively contained and well-founded manner. However, white-box modelling cannot be ignored, since it enables us to describe *how* a cognitive system functions, which is a concern for both traditional information processing and more recent neurophysiological explanations. Thus, when tackling the computational modelling of a particular cognitive phenomenon, one should

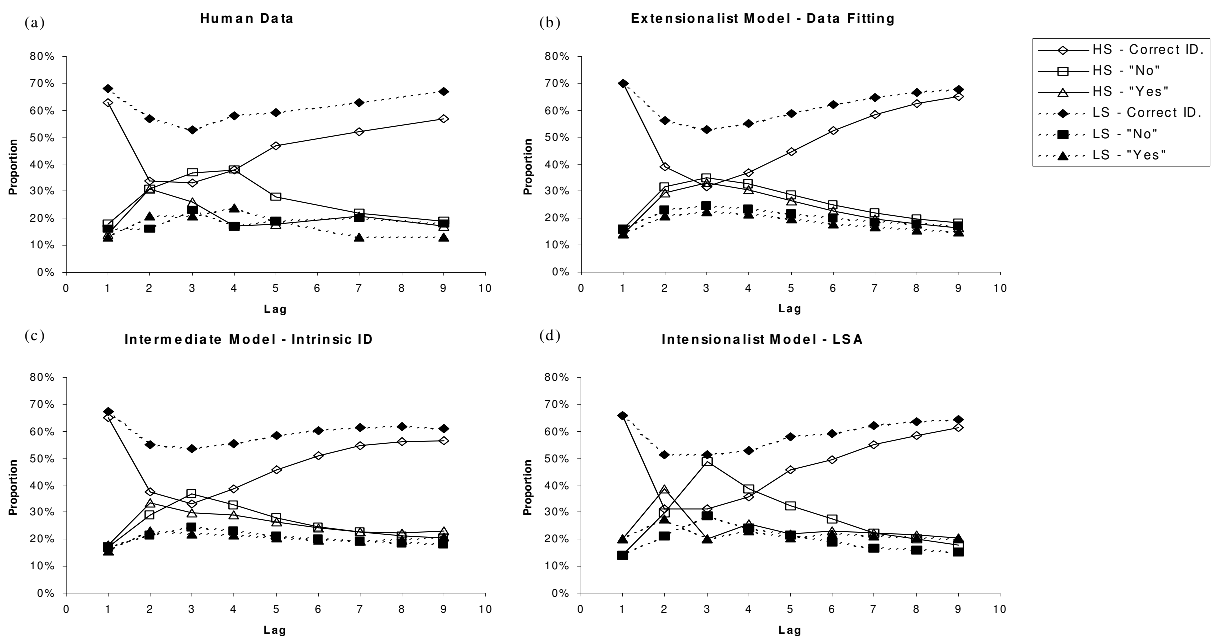

Figure 1 Proportion of different types of responses. HS and LS denote high and low salient condition respectively; Correct ID denotes correct report of target identity. "Yes" denotes response if subject was confidant a job word had been there but could not say exactly what it was. "No" denotes responses if subject did not see a target.

start with an abstract black-box analysis of the observable behaviour arising from the phenomenon. Then, from this solid foundation, one could develop increasingly refined and concrete models, in a progression towards white-box models. Importantly though, this approach enables cross abstraction level validation, showing, for example, that the white-box model is correctly related to the black-box model.

This paper provides an initial step in the direction of multilevel cognitive modelling. In particular, the refinement we present is more from black to dark-gray, then to light-gray! More complete instantiation of our approach awaits further theoretical work on how to relate the sorts of models developed in the cognitive modelling setting.

A key contribution of the article will be the identification of analogous parameter manipulations in all the three models. These cross-model relationships effectively serve as a verification that the theoretical claims we make of our most intensionalist model are well-founded.

Key-distractor Attentional Blink

We illustrate our approach in the context of a study of temporal attention. To do this, we reproduce data on the key-distractor attentional blink task (Barnard, Scott et al. 2004), which considers how attention is drawn to semantically salient items. A particular reason for focusing on this task is that it maps out the profile of attentional capture by meaning over time. This is encapsulated in the serial position curve; see Figure 1.

In order to examine semantic effects, (Barnard, Scott et al. 2004) used a variant of the Attentional Blink (AB) paradigm in which no perceptual features were present to distinguish targets from background items. In this task, words were presented at fixation in Rapid Serial Visual Presentation (RSVP) format, at around 10 items per second. Targets were only distinguishable from background items in terms of their meaning. This variant of the paradigm did not rely on dual target report. Rather, participants were simply asked to report a word if it refers to a job or profession for which people get paid, such as waitress, and these targets were embedded in background words that all belonged to the same category, e.g. nature words. However, streams also contained a key-distractor item, which, although not in the target category, was semantically related to that category. The serial-position that the target appeared after the key-distractor was varied.

Participants could report the target word (accurate report), say "Yes" if they were confident a job word had been there but could not say exactly what it was, or say "No" if they did not see a target, and there were trials on which no target was presented. When key-distractors were household items, a different category from both background and target words, there was little influence on target report. However, key-distractors that referenced a property of a human agent, but not one for which they were paid, like tourist or husband, gave rise to a classic and deep blink as shown in Figure 1a (e.g. the HS - Correct ID curve) & Figure 2b. The horizontal axis denotes lag, which indicates the serial position of the target relative to the key-distractor. The vertical axis denotes the proportion of each types of responses used. Thus, (Barnard, Scott et al. 2004) showed that the level of salience of the key-distractor, i.e. how related it is to the target category, modulates how strongly attention is captured. In this paper, we will concentrate on quantitatively modeling this key effect that semantic similarity modulates the blink

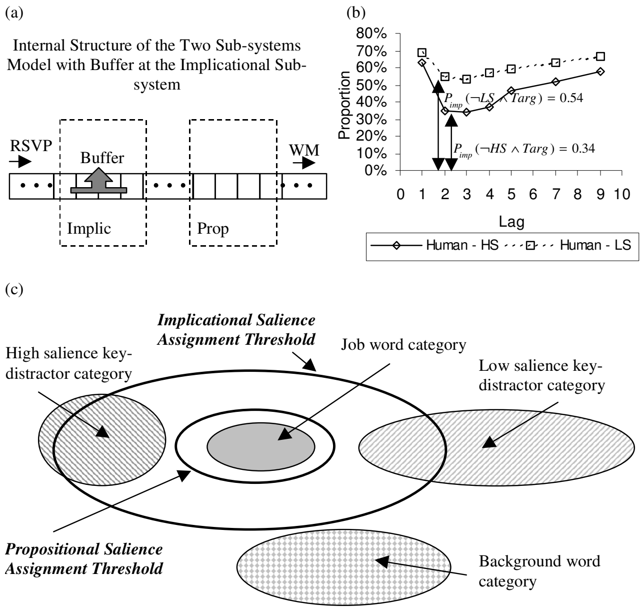

Figure 2 (a) Internal structure. (b) Target report accuracy by lag in humans for high and low salient key-distractors with intrinsic identifications. (c) Salience assignment. Semantics in LSA are expressed in a high dimensional space. This illustration is 2D for ease of depiction.

depth as shown in Figures 1a & 2b, and present both (black-box) extensionalist and (white-box) intensionalist models.

Extensionalist Model – Data Fitting

The most extensionalist approach begins with behavioural data from Barnard's key-distractor task. Accordingly, this model fits the behavioural data using a closed-form equation. This approach has been applied to almost every branch of science in order to characterise the observed behaviour and formulate mathematical models of the underlying mechanisms. This technique has also been widely used in modelling response time distributions (Van Zandt 2000) and, more recently, in modelling serial position curves of AB tasks (Cousineau, Charbonneau et al. 2006).

In our context of exploring the key-distractor AB task, the human data has a sharp blink onset and shallow recovery as shown in Figure 1a (e.g. the HS - Correct ID curve) & Figure 2b. This shape matches an inverted Gamma distribution (GD). (Note, there is a shape parameter in the GD, which determines the skewness of the distribution. Increasing the shape parameter, moves the GD towards a normal distribution; decreasing it, moves the GD towards an exponential distribution.) Hence, we use the following equation to model our AB curves.

$$f(x) = a + b \cdot y(x)$$

where x denotes lag; a is the baseline parameter, which sets baseline performance and, thus, performance following blink recovery; b is the depth parameter, which sets the difference between the deepest point of the blink and the baseline; and y(x) denotes the GD, which also has parameters. However, b is the only parameter that changes significantly when different key-distractors are used in the experiment. The function becomes the baseline if b is 0, i.e. complete absence of the blink and baseline performance at all lags. Hence, we argue that b is related to salience of the key-distractor and thus characterises the attentional capture by salience effect we are interested in.

A simple search of the parameter space has proved sufficient to yield a good fit to the experimental data. We show this fit in Figure 1b. Note, the ratio of the *b* parameter between low and high salient conditions is around $0.4/0.9 \approx 0.44$. Moreover, the GD shape parameter is relatively small for all curves. This suggests that the blink curves are asymmetrical. It will become clear that this relationship is consistent among our different models.

Intermediate Model – Intrinsic Identification

In this section, we model the internal structure of the system as shown in Figure 2a. Three principles underlie our model: sequential processing, 2-stages and serial allocation of attention. We discuss these principles in turn.

Sequential Processing. With any RSVP task, items arrive in sequence and need to be correspondingly processed. Thus, we require a basic method for representing this sequential arrival and processing of items. At one level, we can view our approach as implementing a pipeline. New items enter the front of the pipeline from the visual system; they are then fed through until they reach the back of the pipeline, where they enter working memory (WM). Every cycle, a new item enters the pipeline and all items currently in transit are pushed along one place. The key data structure that implements this pipeline metaphor is a delay-line as shown in Figure 2a. It could also be viewed as a symbolic analogue of a sequence of layers in a neural network; a particularly strong analogue being with synfire chains (Abeles, Bergman et al. 1993). It is a very natural mechanism to use in order to capture the temporal properties of a blink experiment, which is inherently a time constrained order task.

2-Stages. Like (Chun and Potter 1995; Bowman and Wyble 2007), (Barnard, Scott et al. 2004) and (Barnard and Bowman 2004) argued for a two-stage model, but this time recast to focus exclusively on semantic analysis and executive processing. In particular, (Barnard and Bowman 2004) modelled the key-distractor blink task using a two-stage model. In the first stage, a generic level of semantic representation is monitored and initially used to determine if an incoming item is salient in the context of the specified task. If it is found to be so, then, in the second stage, the specific referential meaning of the word is subjected to detailed semantic scrutiny; thus, a word's meaning is actively evaluated in relation to the required referential properties of the target category. If this reveals a match, then the target is encoded for later report. The first of these stages is somewhat akin to first taking a "glance" at generic meaning, with the second akin to taking a closer "look" at the relationship to the meaning of the target category. These two stages are implemented in two distinct subsystems as shown in Figure 2a: the implicational subsystem or *Implic* and the propositional subsystem or *Prop* (Barnard 1999). (We consider how these subsystems fit into a larger cognitive framework, ICS, in the conclusion.)

These two subsystems process qualitatively distinct types of meaning. One, implicational meaning, is holistic, abstract and schematic, and is where affect is represented and experienced (Barnard 1999). The other is classically "rational", being based upon propositional representation, capturing referentially specific semantic properties and relationships. Semantic errors make clear that sometimes we only have (referentially non-specific) semantic gist information available to us, e.g. the Noah illusion illustrates implicational meaning (Erickson and Mattson 1981).

Serial Allocation of Attention. Our third principle is a mechanism of attentional engagement. It is only when attention is engaged at a subsystem that it can assess the salience of items passing through it. Furthermore, attention can only be engaged at one subsystem at a time. Consequently, semantic processes cannot glance at an incoming item, while looking at and scrutinising another. This constraint will play an important role in generating a blink in our models. When attention is engaged at a subsystem, we say that it is buffered (Barnard 1999). (In the context of this paper, the term buffer refers to a moving focus of attention.) Thus, salience assignment can only be performed if the subsystem is buffered and only one subsystem can be buffered at a time as shown in Figure 2a. The buffer mechanism ensures that the central attentional resources are allocated serially, while items pass concurrently, i.e. all items throughout the overall delay-line are moved on one place on each time step.

How the Model Blinks. In this model, words are expressed by their roles in Barnard et al's blink task, i.e. background, target, and key-distractor, which has two subtypes: high salient and low salient. The buffer movement dynamic provides the underlying mechanism for the blink.

Initially, *Implic* is buffered as shown in Figure 2a. When, in response to the key-distractor being found implicationally salient, the buffer moves from *Implic* to *Prop*, salience assessment cannot be performed on a set of words (i.e. a portion of the RSVP stream) entering *Implic* following the key-distractor. So, when these implicationally uninterpreted words are passed to *Prop*, propositional meaning (which builds on implicational meaning) cannot be accessed. Target words falling within this window will not be detected as implicationally salient and thus will not be reported.

There is normally lag-1 sparing in key-distractor AB experiments, i.e. a target word immediately following the key-distractor is likely to be reported. This arises in our model because buffer movement takes time, hence, the word immediately following the key-distractor may be implicationally interpreted before the buffer moves to *Prop*.

When *Prop* is buffered and detects an implicationally uninterpreted word, the buffer is passed back to *Implic*, which can assign salience to its items again. After this, target words entering the system will be detected as implicationally and propositionally salient and thus will be reported. Hence, the blink recovers.

Generating a Blink Curve. Humans though perceive information imperfectly; as a result, salient items may be missed. In the current model, we assume that the ease of detecting that the key-distractor is implicationally salient determines the depth of the blink curve. We work here with what we call "intrinsic probabilities of identification", i.e. if an item (distractor or target) is presented alone in an RSVP stream, what is the probability that it will be seen. Thus, $P_{imp}(Dist \wedge Targ)$ is not the probability that both the key-distractor and target are seen in an AB setting, but rather the probability that both would be seen in two separate idealised "single target events". The intrinsic probability of judging targets to be implicationally salient, $P_{imp}(Targ) = 0.67$, is set by the baseline performance of human subjects. (Barnard et al stated that humans correctly report the target's identity on average on 67% of target only trials; furthermore, at high lags, the blink curve also recovers to this baseline performance (Barnard, Scott et al. 2004).) We assume that the intrinsic probability of detecting a background word as implicationally salient, $P_{imp}(Back)$, is zero. (This sort of error is so rare as to be effectively zero.) The intrinsic probability of detecting a key-distractor as implicationally salient is $P_{imp}(HS)$ in the high salient condition and $P_{imp}(LS)$ in the low salient condition. According to our model, the likelihood of correct report at the deepest point in the blink curve reflects the joint probability of missing the key-distractor and detecting the target. This is because the way the model is constructed, there is indeed no other way that a target can be detected during the blink. From Figure 2b, $P_{imp}(\neg HS \wedge Targ) = 0.34$ and $P_{imp}(\neg LS \wedge Targ) = 0.54$ can be obtained. We assume detecting targets and the key-distractors are independent, in particular, in both cases we assume the buffer is at *Implic* when the assessment is made. So, $P_{imp}(HS) = 0.49$ and $P_{imp}(LS) = 0.19$.

This calculation quantitatively determines how the model generates a blink curve. As a reflection of the relatively high level of abstraction of this model, randomness is imposed globally and externally using a convolution. This technique does not require specification of either the dynamics or the source of noise inside the model. As a result, assumptions about the internal structure of the system are minimised and also the number of simulation runs is reduced. Thus, we convolve Gaussian-distributed noise (GDN) with the (noise free) simulation results. We also gradually increase the deviation of the GDN by serial position, i.e. the GDN is narrower at earlier lags and broader at later lags. We call this a *convolution with sliding noise*. (Note, we explored simpler convolution strategies, but none of these generated a suitable blink curve, see (Bowman, Su et al. 2006) for details). The intuition behind this approach is that there is less noise in earlier phases of processing than in later phases of processing, which influence blink onset and recovery respectively. Application of such a convolution with sliding noise results in a good fit to the human data as shown in Figure 1c. Note, our extensionalist model achieves this blink curve asymmetry by setting the GD shape parameter, which determines how skewed it is from a normal distribution.

In our simulations, the meaning of a target word can be processed to three different degrees, which, we argue, reflect different types of response. Words that are both implicationally and propositionally fully interpreted can be reported correctly with their identity. Some target words can be implicationally fully un-interpreted, reflecting complete unawareness of the presence of target words, i.e. the "No" responses. Finally, some target words can be partially processed, reflecting the "Yes" response.

The resulting percentages of correct report of target identities, "No" responses and "Yes" responses are shown in Figure 1c. These graphs also illustrate the difference in performance between the high and low salience conditions. The results are consistent with the experimental results from humans (Barnard, Scott et al. 2004) shown in the same graph. Moreover, the ratio between low and high salient key-distractor intrinsic probabilities of identification is $0.19/0.49 \approx 0.39$, which is similar to the ratio of the depth parameters (0.44) in the previous model.

Intensionalist Model – LSA

In previous models, parameters were derived from human performance on the AB task and assumptions about the internal structure were minimized. However, in this model, word meanings are represented using Latent Semantic Analysis (LSA) (Landauer and Dumais 1997), which was developed outside the AB. In this sense, this model's key parameters were constrained by a general theory that will be used to explain the intrinsic probability and the depth parameter in our previous models.

We hypothesize that a word is assigned to be salient if the semantic distance (an LSA cosine) between the word and the target category is smaller than a specified threshold. As shown in Figure 2c, the target words are within the propositional salience threshold. Hence, they are both implicationally and propositionally salient. On the other hand, background words are outside the implicational salience threshold. Hence, they are both implicationally and propositionally unsalient. Key-distractors can be either implicational salient or unsalient. However, they cannot be propositionally salient. Only job words can be reported and only implicationally salient key-distractors can cause blinks.

In this model, the depth of the blink curve depends on the percentage of key-distractors above the implicational threshold. We calculated the LSA cosines in relation to the meanings: generic human, generic occupation, generic payment, generic household and nature categories (Barnard, Scott et al. 2004). Then, we integrated these cosines as a weighted sum of these five LSA values. Effectively, we "skew" the LSA space according to the extraction of implicational meaning. The five weights characterise this skewing, reflecting the relative emphasis that the implicational schema puts on each of the five dimensions.

We constructed a two layer neural network to determine these weights. The input layer contained five neurons, one for each of the five categories. The output layer was a single neuron. We trained the network using all the words we used in the AB experiment. The learning algorithm used was the delta rule (O'Reilly and Munakata 2000). The inputs were LSA cosines and the expected output was 1 for targets and 0 for non-targets. The learning finished when the weights settled, i.e. their changes were smaller than a given value (0.0001). Using the trained network, we calculated the new LSA values for all words. The results were: 52.5% of high salient and 22.2% of low salient key-distractors were implicationally salient. Nature words were mainly implicationally unsalient, except for one word (so, we excluded this word from our simulation). 63.4% of target words were implicationally salient. Interestingly, the ratio between low and high salient key-distractor LSA calculations was $22.2/52.5 \approx 0.42$, which is consistent with the depth parameters (0.44) and intrinsic probabilities (0.39) derived from our previous model.

As a reflection of the fact that this is a more concrete model than the previous ones, convolutions are not used here. Instead, different amounts of variance are added to the buffer movement delay at different stages, i.e. less variance is added to the delay of buffer movement from *Implic* to *Prop* (which regulates blink onset) than the delay of buffer movement in the opposite direction (which regulates blink offset). Our extensionialist and intermediate models justify this, i.e. GD is a skewed distribution and the sliding noise ensures that the variance increases by lag. Partial responses are modelled in a similar way as the intermediate model. The simulation results are shown in Figure 1d. Full details of these models can be found in (Bowman, Su et al. 2006).

Conclusion

Attentional Capture by Meaning. We have provided a concrete account of attentional capture by meaning and the temporal dynamics of that process. A number of key findings have arisen from our modelling. Firstly, we have provided further evidence for the applicability of LSA in the context of attentional capture by meaning. That is, we have shown that a model that measures semantic distance using LSA can reproduce the key-distractor blink and semantic modulations of blink depth. Furthermore, we have shown that these LSA calculations are consistent with more extensionalist approaches in which the difference in observable behaviour is captured by either the GD depth parameter, or intrinsic probabilities of ascribing implicational salience derived directly from the blink curve. Importantly, in all three cases, i.e. GD depth parameter, intrinsic probabilities of implicational salience and LSA measures of implicational salience, the ratio between high and low salience has been almost identical (around 0.42). This is an illustration of how multilevel modelling can provide converging evidence for a theoretical position.

Secondly, we have clarified the characteristics of attentional redeployment when meaning captures attention. In particular, at an extensionalist level, a skewed distribution was used to characterise the asymmetry of the blink curve. At an intermediate level, the need to use a convolution with sliding noise suggests that temporal noise increases systematically by serial position. At an intensionalist level, this sliding noise is realised as variance in the buffer movement delay. This finding suggests that there is less variance in extracting semantic gist (at *Implic*) than extracting referential meaning (at *Prop*), since *Implic* does not have to fully analysis and generate a concrete referent, which is likely to be affected by many variables. This consistency is again an illustration of converging evidence from different levels of modelling.

Cognitive Architectures. The general applicability of our models is enhanced since the approach can be placed within the context of a broad cognitive theory: the Interacting Cognitive Subsystems (ICS) architecture (Barnard 1999). Distributed control is inherent in ICS: subsystems are independent components, which interact through exchange of data representations over communication channels (Barnard 1999; Bowman and Faconti 1999; Barnard and Bowman 2004). ICS asserts that cognition emerges as the product of the interaction between a set of autonomous subsystems. Both the delay-line and buffering concepts that we use have their roots in ICS. However, most significantly, the implicational - propositional distinction reflects ICS' dual-subsystem central engine (Teasdale and Barnard 1993).

Multi-level Cognitive Modelling. We have provided a case study for how multilevel modelling can be applied in the cognition setting. Viewing systems from different perspectives and levels of abstraction is just a useful exploratory method for understanding systems, and it is one that the cognitive modelling domain should not miss.

References

Abeles, M., H. Bergman, et al. (1993). Spatiotemporal Firing Patterns in the Frontal Cortex of Behaving Monkeys. Journal of Neurophysiology **70**: 1629-1638.

Barnard, P. J. (1999). Interacting Cognitive Subsystems: modelling working memory phenomena within a multi-processor architecture. Models of Working Memory: Mechanisms of active maintenance and executive control**:** 298-339.

Barnard, P. J. and H. Bowman (2004). Rendering information processing models of cognition and affect computationally explicit: Distributed executive control and the deployment of attention. Cognitive Science Quarterly **3**(3): 297-328.

Barnard, P. J., S. Scott, et al. (2004). Paying attention to meaning. Psychol Sci **15**(3): 179-86.

Booch, G., J. Rumbaugh, et al. (1999). The Unified Modelling Language User Guide, Addison-Wesley.

Bowman, H. and G. Faconti (1999). Analysing Cognitive Behaviour using LOTOS and Mexitl. Formal Aspects of Computing **11**: 132-159.

Bowman, H., L. Su, et al. (2006). Semantic Modulation of Temporal Attention: Distributed Control and Levels of Abstraction in Computational Modelling, Technical Report 9-06, Computing Lab, University of Kent.

Bowman, H. and B. Wyble (2007). The Simultaneous Type, Serial Token Model of Temporal Attention and Working Memory. Psychological Review 114(1):38-70.

Chun, M. M. and M. C. Potter (1995). A Two-Stage Model for Multiple Target Detection in Rapid Serial Visual Presentation. Journal of Experimental Psychology: Human Perception and Performance **21**(1): 109-127.

Cousineau, D., D. Charbonneau, et al. (2006). Parametering the attential blink effect. Canadian Journal of Experimental Psychology **60**(3): 175-189.

Dehaene, S., M. Kerszberg, et al. (1998). A neuronal model of a global workspace in effortful cognitive tasks. Proc Natl Acad Sci U S A **95**(24): 14529-34.

Erickson, T. D. and M. E. Mattson (1981). From words to meaning: a semantic illusion. Journal of Verbal Learning and Verbal Behavior **20**: 540-551.

Fodor, J. A. and Z. W. Pylyshyn (1988). Connectionism and cognitive architecture: a critical analysis. Cognition **28**: 3-71.

Hinton, G. E. (1990). Special Issue of Journal Artificial Intelligence on Connectionist Symbol Processing (edited by Hinton, G.E.). Artificial Intelligence **46**(1-4).

Kieras, D. E. and D. E. Meyer (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. Human-Computer Interaction **12**: 391-438.

Landauer, T. K. and S. T. Dumais (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of the Acquisition, Induction and Representation of Knowledge. Psychological Review **104**: 211-240.

Marr, D. (2000). Vision. Minds, Brains and Computers, The Foundation of Cognitive Science (An Anthology). R. Cummins and D. D. Cummins, Blackwell**:** 69-83.

McLeod, P., K. Plunkett, et al. (1998). Introduction to Connectionist Modelling of Cognitive Processes, OUP.

Newell, A. (1990). Unified Theories of Cognition. Cambridge, Massachusetts, Harvard University Press.

O'Reilly, R. C. and Y. Munakata (2000). Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain. MIT Press.

Stewart, N., G. D. A. Brown, et al. (2005). Absolute Identification by Relative Judgment. Psychological Review **112**(4): 881-911.

Teasdale, J. D. and P. J. Barnard (1993). Affect, Cognition and Change: re-modelling depressive thought. Hove, Lawrence Erlbaum Associates.

Van Zandt, T. (2000). How to fit a response time distribution. Psychol Bulletin & Review **7**(3): 424-465.