UNIVERSITA' DEGLI STUDI DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Psicologia dello Sviluppo e della Socializzazione

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE PSICOLOGICHE

INDIRIZZO: SCIENZE COGNITIVE

CICLO: XX

On the Structure of Semantic Number Knowledge

Direttore della Scuola: Ch.mo Prof. Luciano Stegagno

Supervisore: Ch.mo Prof. Marco Zorzi

Dottorando : Thomas Hope

DATA CONSEGNA TESI

31 luglio 2008

for Mike

Abstract

English: In recent years, there has been a surge of interest – and progress – in the cognitive neuroscience of numerical cognition. Yet despite that progress, we still know comparatively little about the fundamental structure of numerical cognition; the subject is a fertile source of compelling research questions. The answers to those questions also have far-reaching social implications. Recent research suggests that the incidence of extreme disorders of numerical skills (dyscalculia) among students may be as high as 6%; if that proportion is consistent in later life, 40 million people might be affected in Europe alone.

Recent evidence suggests that both animals and pre-linguistic infants are sensitive to number. Coupled with the observation that this sensitivity shares features in common with that displayed by normal adults, these studies establish a role for evolution in the development of a functioning number sense. Chapters 3 through 5 describe a project that explores this connection by "evolving" quantity-sensitive agents in simulated ecosystem; the goal was to discover what kinds of number representations might emerge from a selective pressure to forage effectively. Agents of this sort are notoriously difficult to analyse – so difficult that some researchers have claimed they are simply unsuited to classical interpretation (this is the subject of chapter 3). Chapter 4 is devoted to a novel analytical method that rebuts these claims by "discovering" recognisable representations in a model (of categorical perception tasks) that had previously been thought to eliminate them. Chapter 5 applies the method to quantity-sensitive foraging agents, and reveals a novel format for number representation – a single unit accumulator code – that is nevertheless well-supported by recent neurophysiological data.

Chapter 6 shifts the focus away from number knowledge itself, and toward the decision-

process that might use it. This project describes a novel model-building method that is driven solely by the intuition that neural information processing will be formally optimal (or near optimal). Applied to the problem of making categorical responses to noisy stimuli (most popular theories of number representation are noisy), this method captures both the behaviour that has been observed in humans and other primates, and its apparent, neural implementation.

Chapters 7 through 9 all consider the processing of very large numbers, or long digit strings. The role of digit-level vs. whole-number content in these strings is still quite controversial; chapter 7 reports an experiment that uses a novel prime – the “thousands” string ('000') – to dissociate the two. The prime appears to mediate the subjects' sensitivity to both types of content, so confirms that they are sensitive to both types. The results also appear to dissociate two empirical phenomena – the Size and Distance effects – that are both thought to be associated with interference at the level of number representations. That common cause makes it difficult to explain how the two effects could diverge; just one of the four popular theories of number representation appears to allow it.

Chapter 8 explores multi-digit number comparison from a computational perspective, presenting a model-space (for models trained to solve number comparison problems for numbers in the range 1-999), defined by systematic variation of the representation configurations and hidden layer sizes that define particular models. By performing some standard statistical analyses on the behaviour of the models ion this space, the chapter reveals a specific representation configuration (decomposed representations with the single unit accumulator code) that can reliably drive effective models.

Finally, chapter 9 reports the results of two more experiments, which establish constraints on the perceptual processing of long digit strings. Both experiments use masking to prevent subjects from using saccades to process visually presented digit strings. With this restriction in place, the first experiment establishes that subjects can enumerate at most 6 or 7 digits, and the second suggests that they can identify up to 3 or 4.

Italiano: Negli ultimi anni c'è stato un forte interesse - ed un corrispondente progresso - nelle neuroscienze cognitive della cognizione numerica. Malgrado numerosi e rilevanti progressi in questi ambiti, sappiamo ancora molto poco dei processi cognitivi sottostanti l'elaborazione numerica. Studi recenti suggeriscono come l'incidenza di discalculia fra gli studenti possa raggiungere il 6%; nel caso questa percentuale si riproponesse anche in età adulta, ciò significherebbe che solo in Europa ci potrebbero essere 40 milioni di persone discalculiche. Lo studio dei processi cognitivi sottostanti l'elaborazione numerica non è dunque solo fonte di interessanti quesiti, ma può anche fornire delle risposte che hanno importanti implicazioni sociali.

Lavori recenti suggeriscono come alcuni animali (ad es. scimmie ma anche salamandre) siano in grado di differenziare diverse numerosità. Dal momento che queste capacità condividono alcune caratteristiche con i processi che caratterizzano gli adulti normali, questi studi stabiliscono un ruolo per l'evoluzione nello sviluppo di un "senso del numero". I capitoli dal 3 al 5 descrivono un progetto che esplora questo collegamento tramite esseri "quantità-sensibili" che si sviluppano in un ecosistema simulato; lo scopo era scoprire che tipi di rappresentazioni numeriche potessero emergere in seguito ad una pressione selettiva per procacciarsi del cibo efficacemente. Gli agenti di questo tipo sono notoriamente difficili da analizzare, tanto che alcuni ricercatori hanno affermato che questi sono semplicemente inadeguati per una interpretazione classica (questo argomento sarà trattato nel capitolo 3). Il capitolo 4 descrive un nuovo metodo analitico che rifiuta queste affermazioni, "scoprendo" rappresentazioni riconoscibili anche in un modello (compiti di percezione categorica), che in precedenza si pensava li eliminasse. Il capitolo 5 applica questo metodo agli agenti quantità-sensibili, e suggerisce la presenza di un nuovo tipo di rappresentazione numerica – un "single unit accumulator code" – che è oltretutto ben supportato da dati neurofisiologici recenti.

Il capitolo 6 sposta l'attenzione dalla rappresentazione delle quantità numeriche ai processi decisionali che di queste possono avvalersi. Viene descritto un nuovo metodo per costruire modelli computazionali. Si ipotizza che la semplice l'elaborazione delle informazioni neurali potrebbe essere costituire il miglior modello formale possibile. Quando applicato per risolvere il problema di fornire delle risposte categoriche a degli stimoli "noisy", questo metodo produce dei modelli che può riprodurre sia il comportamento di soggetti primati (RT e gli errori contro la forza di segnale) e la sua implementazione neurale.

I capitoli dal 7 al 9 considerano l'elaborazione di numeri molto grandi, o lunghe stringhe di numeri. Il ruolo del rapporto fra contenuto a livello di cifra e contenuto a livello dell'intera stringa numerica è poco chiaro; il capitolo 7 presenta un esperimento che usa un nuovo tipo di "prime" e cioè la stringa "delle migliaia" ("000") – per dissociare i due contenuti. Il prime sembra mediare la sensibilità dei partecipanti ad entrambi i tipi di contenuto. I risultati appaiono dissociare anche due fenomeni empirici – gli effetti "Grandezza" e "Distanza" – che si pensa siano associati con l'interferenza al livello delle rappresentazioni numeriche. Proprio questa associazione fra "Grandezza" e "Distanza" rende difficile spiegare come i due effetti possano differire; solamente una fra le quattro maggiori teorie sulla rappresentazione numerica contempla infatti questa possibilità.

Il capitolo 8 esplora il confronto di numeri a più cifre da una prospettiva computazionale, presentando un gruppo di modelli (addestrati a risolvere compiti di confronto numerico per numeri nella gamma 1-999), caratterizzato dalla variazione sistematica dei parametri dei modelli. Eseguendo delle analisi statistiche sul comportamento dei modelli nel spazio, il capitolo evidenzia un gruppo piccolo di modelli – ed una specifica configurazione di rappresentazione (rappresentazioni scomposte con il single unit accumulator code).

Infine, il capitolo 9 riporta i risultati di due ulteriori esperimenti, miranti a stabilire caratteristiche e limiti dell'elaborazione percettiva di stringhe numeriche. Entrambi gli esperimenti utilizzano il paradigma di masking per evitare che i partecipanti facciano ricorso a movimenti oculari finalizzati ad un processamento più efficiente delle stringhe numeriche. I risultati differiscono in base al compito proposto. Il primo esperimento mostra come sia possibile enumerare 6 o 7 cifre, mentre il secondo suggerisce invece che il massimo numero di cifre identificabili sia di 3 o 4.

Acknowledgements

I wish to thank my supervisor, Marco Zorzi, both for giving me this opportunity and for helping me to take it. My thanks also to Klaus Willmes von Hinckeldey and Stanislas Dehaene, for their supervision, support and advice during my periods of study abroad. I also wish to thank Brian Butterworth, without whom I might never have known what I was missing.

I wish to thank Ivilin Stoianov, both for the technical advice and for the random conversation, and Peter Kramer, for putting up with raised voices.

Finally, I wish to thank Alice, for keeping me sane – but not too sane.

Contents

III
On the Processing of Large Numbers

IV
Summary, Discussion and Conclusions

Like all men of the Library, I have travelled in my youth. I have journeyed in search of a book, perhaps the catalogue of catalogues... There are official searchers, inquisitors. I have seen them in the performance of their function: they always arrive extremely tired from their journeys... sometimes they pick up the nearest volume and leaf through it, looking for infamous words. Obviously, no one expects to discover anything.

A blasphemous sect suggested that the searches should cease and that all men should juggle letters and symbols until they constructed, by an improbable gift of chance, these canonical books. The authorities were obliged to issue severe orders. The sect disappeared, but in my childhood I have seen old men who, for long periods of time, would hide in the latrines with some metal disks in a forbidden dice cup, and feebly mimic the divine disorder.

Jorge Luis Borges (translated by James Irby), *La biblioteca de Babel*

I

Introduction and Background

Chapter 1

Introduction

Mathematics governs almost every aspect of our lives. From the first ring of the alarm clock (at 0730) to the book that we read before bed (page 74, chapter 3), our days are invaded by numbers. In industrialised societies such as the UK, numeracy – the ability to apply mathematical tools and concepts in practice (e.g. Cockroft, 1986) – is a key predictor of socio-economic status, even when literacy levels are taken into account (Bynner & Parsons, 1997). And throughout the modern world, "those who lack either [the] confidence or [the] skills to employ basic arithmetic, statistics, and geometry lead their economic lives at the mercy of others" (Steen, 1990).

Recent research suggests that between 3% and 6% of children lack precisely those skills (Shalev, Auerbach, Manor, & Gross-Tsur, 2000). If that proportion is consistent in later life, more than 40 million people might be affected in Europe alone[1]. And there are good reasons to suspect that it *is* consistent; despite significant progress, extreme disorders of numerical skills (dyscalculia) still lack the recognition in education that reading disorders (dyslexia) now enjoy. In the UK for example, dyscalculia has only been officially recognised since 2001 (DfES, 2001).

One reason for this is that dyscalculia can be interpreted as an expression of a more general learning disorder that also explains dyslexia. The case in favour stems from the strong co-morbidity of dyscalculic and dyslexic symptoms; these are thought to be roughly as widespread as each other in the general population (Butterworth, 2005), but students with poor numeracy appear to be disproportionately affected by problems with literacy. Since the trend cuts across both linguistic and orthographic boundaries (e.g. Lewis, Hitch, & Walker, 1994; Ostad, 1998; Gross-Tur, Manor, &

1 Based on United Nations population estimates, available at http://esa.un.org/unpp.

Shalev, 1996), it seems natural to explain it by citing a common cause. And if that explanation is right, both kinds of symptoms should be susceptible to similar remediation; by researching dyslexia, we might hope to remediate dyscalculia as well. This same logic can also begin to explain why the neuroscience of numeracy has attracted rather less attention than its counterpart for literacy. A common cause for their respective disorders implies a common cognitive architecture for numerical and linguistic skills – so research on the latter should also be expected to illuminate the former.

Recently, the logic that elides language and number has begun to be rejected. Support for that rejection starts with a series of studies conducted by Henschen during the 1920's, demonstrating that (despite their apparent connection) disorders of language and numerical skills can occur independently. Consistent with that result, more recent studies with neurological patients have identified both disorders of numerical skills with preserved language (e.g. patient CG: Cipolotti, Butterworth, & Denes, 1991) and disorders of language with preserved numerical skills (e.g. patient IH: Cappelletti, Butterworth, & Kopelman, 2001; Cappelletti, Kopelman, & Butterworth, 2002) – a classic double dissociation. At the same time, neuroimaging data have begun to establish a convincing association between numerical processing and the horizontal segment of the bilateral intra-parietal sulcus (e.g. Dehaene, Piazza, Pinel, & Cohen, 2003; Simon, Mangin, Cohen, Bihan, & Dehaene, 2002) – some distance away from the more classical language areas. Indeed, one study (Pesenti, Thioux, Seron, & De Volder, 2000) has found that activity in Broca's area can be *depressed* during numerical tasks, suggesting an apparent opposition between language and number processing.

This evidence leaves a lot of questions unanswered. For example, it is perfectly possible to promote the distinction between numerical and linguistic skills, while still claiming that the latter play a pivotal role in the development of the former (e.g. Mix, Huttenlocher, & Levine, 2002; Carey, 2004). And for all the evidence of difference, linguistic and numerical skills still clearly recruit at least some of the same neural architecture. Nevertheless, the increasing separation of the numerical and linguistic domains presents an opportunity for research that illuminates a fundamental cognitive concept (i.e. of number), with potentially far-reaching social implications. This thesis is inspired by that opportunity.

1.1 Overview of the Thesis

The next chapter (chapter 2) presents a summary of the empirical and computational background to this thesis. The chapter's goal is to depict a context for the novel work that follows, to motivate that work, and to establish a terminology that the subsequent chapters can employ.

Chapters 3 to 5 all address the format of semantic number knowledge, building through a series of stages toward a novel theory of that format. The proposal depends on extensive new methodology, which is introduced in chapter 3. Chapter 4 develops and deploys the method in practice, reporting a model of categorical perception. Chapter 5 extends that work and applies it to numerical cognition, reporting a novel model of number comparison, and proposing a novel format for the representations that drive it.

Chapter 6 shifts the focus away from number knowledge to the decision processes that employ it, introducing a model that interprets noisy representations (of which number knowledge may be an example) with categorical decisions. This work illustrates how that process can be captured with extremely minimal assumptions.

Chapters 7 to 9 all consider problems associated with the processing of larger, multi-digit numbers. Chapter 7 presents an experiment that dissociates subjects' sensitivities to these numbers' integrated values from their sensitivity to the numbers' single-digit components. Chapter 8 presents a model-space for multi-digit number comparison, which yields a preferred representation of semantic number knowledge, and a group of models that capture the relevant empirical phenomena. And Chapter 9 reports an experiment that captures some basic constraints on the perceptual process that subjects use to manipulate very large numbers (i.e. very long digit strings).

Chapter 2

Background

In this chapter, I first describe the empirical phenomena that drive the current work (section 2.1), before supplying an account of the popular theories of number representation (section 2.2), and of the debates that they inspire. Section 2.3 describes the way in which computational modelling has been used in the past to capture the cognitive processes underlying basic numerical skills. Throughout this chapter, and those that follow, empirical phenomena are associated with Capitalised names, and theories of number representation are given *italicised* names.

2.1 Core Empirical Phenomena

Theories of numerical cognition are driven by a relatively small – though still growing – set of empirical phenomena. Rather than attempting a complete review of the field, this section considers only those that are relevant to the chapters that follow.

First identified in 1967 (Moyer & Landauer, 1967), the Number Size and Distance effects (Figure 1) are perhaps the most robust and familiar of this set. The Number Size effect refers to a characteristic trend of increasing errors and reaction times as the numerical magnitudes of compared numbers increase; for example, the number comparison “2 vs. 3” appears easier to solve than the comparison “8 vs. 9”. A distinct but related phenomenon – the Problem Size effect – describes the observation that simple arithmetic tasks (such as additions) seem more difficult to solve when their operands are numerically larger (i.e. the sum ”5+6” takes longer to solve than the sum “4+5”: Ashcraft, 1992; Groen & Parkam, 1972; Zbrodoff & Logan, 2005). The Distance effect refers to a complementary trend of increasing errors and reaction times as the numerical distance between processed numbers increases; the comparison problem, “4 vs. 5” appears a more difficult judgement than “4 vs. 6” (Moyer & Landauer, 1967). Most popular theories of numerical cognition explain these effects as a result of interference at the level of semantic number representations (Zorzi, Stoianov & Umiltà, 2005) – the focus of the next section.

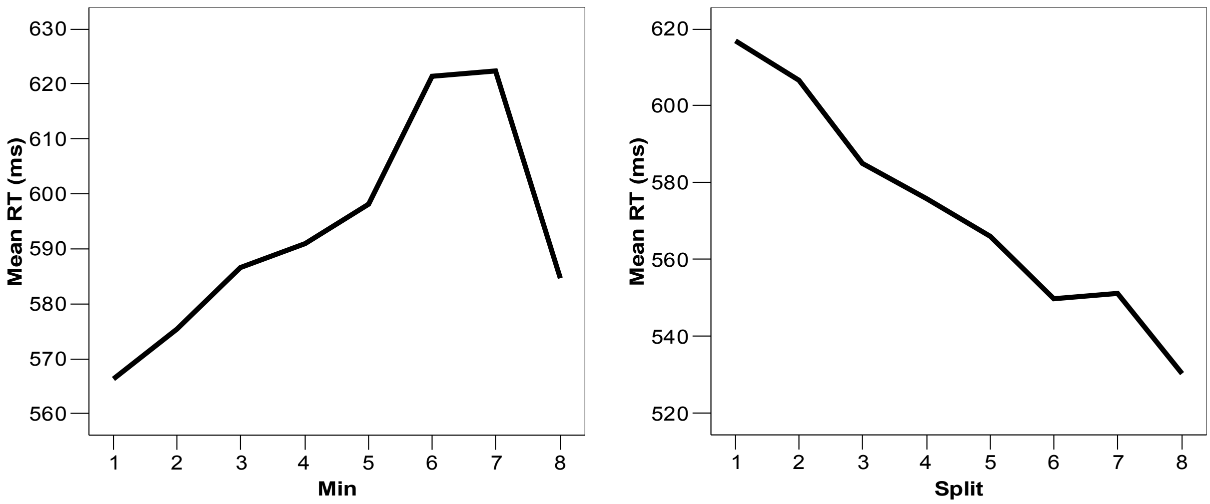


Figure 1: Mean reaction times (ms) of subjects engaged in a single-digit number comparison task. The left-hand graph illustrates the trend toward increasing RT with the minimum of the compared numbers (the Size effect). The sharp decrease for pairs with a minimum of '8' is a characteristic 'edge effect' (e.g. Zorzi, Stoianov, & Umiltà 2005), which arises because pairs that include the maximum number in the range ('9' in this case) are easier for subjects to compare. The right-hand graph illustrates the trend toward decreasing RT with increasing numerical distance between the compared numbers (the Distance effect).

A third phenomenon, which emerges in experiments with multi-digit numbers (considered in chapters 7-9), is best understood as a divergence from the Size and Distance effects. Consider the number pairs “55, 31” (pair A) and “55, 29” (pair B). The minimum of the numbers in pair A ('31') is numerically larger than the minimum of the numbers in pair B ('29'), so the Size effects tells us that the former pair should be more difficult to process (e.g. compare) than the latter pair. The same prediction follows from the Distance effect, since the numbers in pair B are more numerically distant than those in pair A. Early experiments with two-digit numbers appeared to confirm that prediction (e.g. Dahaene, Dupour, & Mehler, 1990), but more recent work – in particular by Nürk and colleagues (e.g. Nürk, Weger, & Willmes, 2001) – has begun to undermine it.

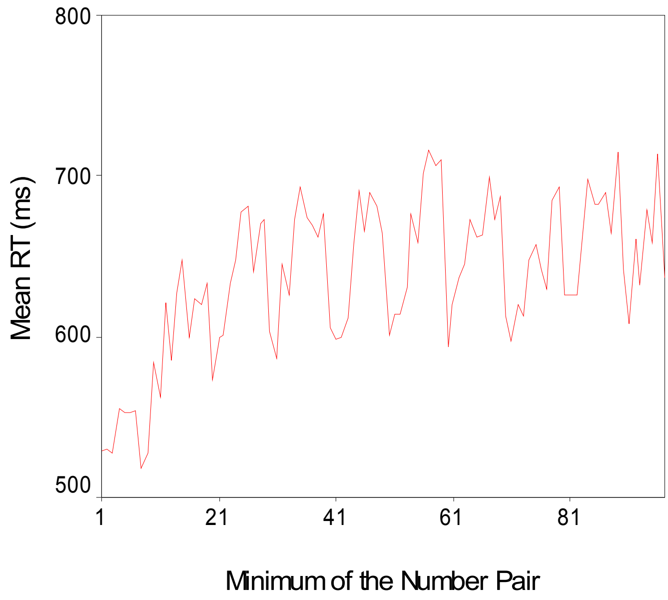


Figure 2: Reaction times in a two-digit number comparison experiment, plotted against the minimum of the two numbers that participants had to compare. There is a visible Size effect, since reaction times are longer toward the right of the graph than they are on the left, but there is also a significant "wobble" that seems to straddle the decade boundaries; this pattern is evidence of a Congruence effect. These data were derived from experiments conducted at the university of Trieste, and supplied by Simone Gazzellini; for a published reference, see Gazzellini & Laudanna, 2005.

The difference stems from a characteristic "wobble" in the trends that the Size and Distance effects describe when subjects must process multi-digit numbers (see Figure 2). Pairs A and B include two tens digits and two units digits each; rather than integrating these digits, it might be perfectly valid to compare them by performing two separate comparisons at the single-digit level. In pair A, both comparisons ('5-3' and '5-1') favour the the first number ('55'), while in pair B, the number with the larger tens digit ('55') also has the *smaller* units digit. If participants employ this “decomposed” strategy, we might expect the incompatibility between these two comparisons to carry a cost (pair A should be compared more quickly than pair B) – precisely the cost that Nürk, Weger and Willmes (2001) report. Costs of this sort have been observed in tasks involving both two-digit numbers (a unit-decade Compatibility effect: Nürk, Weger, & Willmes, 2001) and three-digit numbers (both decade-hundred and unit-hundred effects: Korvorst, & Damian, 2007). We will revisit this phenomenon several times in the material that follows. For convenience, I will use the shorthand – Congruence effects – to refer to all of the variants of this kind of interference, and apply the terms 'congruent' and 'incongruent' to distinguish between number pairs of type A and B respectively.

The Number Size, Distance, and Congruence effects comprise the empirical backbone of this thesis, but three other phenomena will also be relevant in the material that follows. The first is the numerical variant of the well-known semantic priming effect. First demonstrated by Meyer and Schvaneveldt (1971) – who showed that words such as DOCTOR were read more quickly when preceded by semantically related words like NURSE – these effects are now a familiar feature of research on language and memory (see Neely, 1991, for a review). Several studies have confirmed that semantic priming can also be observed with numbers (Den Heyer & Briand, 1986; Koechlin, Naccache, Block, & Dehaene, 1999; Reynvoet & Brysbaert, 1999; Reynvoet, Brysbaert, & Fias, 2002). This work shows that numerical priming effects are inversely proportionate to the numerical distance between the prime and the target, additive to the effect of repetition priming and symmetrical with respect to the priming direction (Figure 3: Reynovet, Brysbaert, & Fias, 2002). That last property – symmetry – has a theoretical significance, which the next section describes.

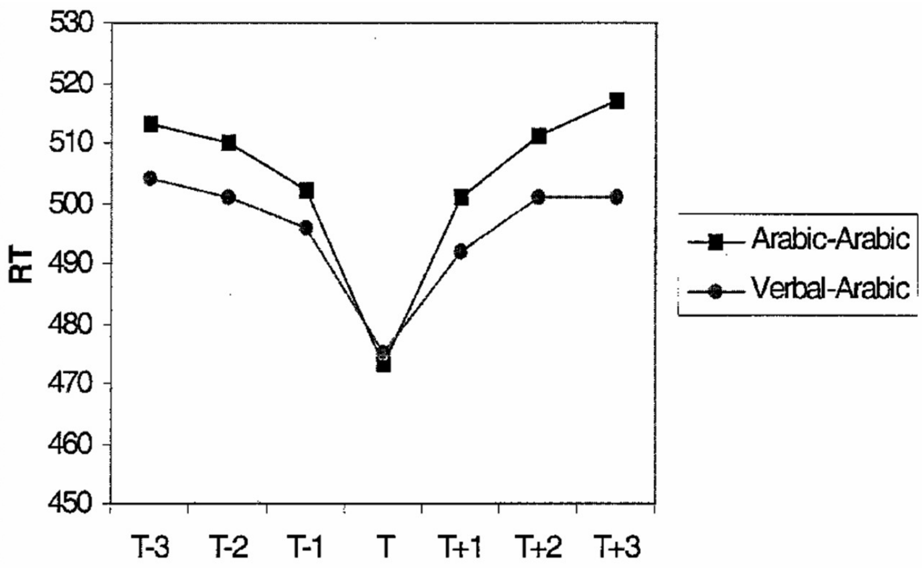


Figure 3: Re-printed from Reynvoet, Brysbaert, & Fias, 2002. Mean reaction times (ms) for number naming tasks in which the target number (T) was preceded by a prime. The primes invoke slower responses when they are numerically close to the target, and the effect is directionally symmetrical.

The second relevant phenomenon was first reported by Dehaene, Dupoux, and Mehler (1990), and confirmed by Dehaene, Bossini, and Giraux (1993) a few years later. In a parity judgement task – requiring odd / even judgements for a series of visually presented Arabic digits – Dehaene and colleagues discovered that participants were significantly faster when responding to smaller numbers (1-4) with the left hand, and to larger numbers (6-9) with the right hand (see Figure 4). The authors called the effect a Spatial-Numeric Association of Response Codes (SNARC), and – remembering that numerical magnitude is irrelevant to parity judgement tasks – interpreted it as evidence for the automatic activation of a mental number line with left-to-right orientation. This interpretation is supported by recent neuropsychological data (Zorzi, Priftis, & Umiltà, 2002), but the effect may not be specific number; SNARC-like effects have also been observed with non-numeric stimuli like letters, days of the week and months of the year (Gevers, Reynvoet, & Fias, 2003; Gevers, Reynvoet, & Fias, 2004). The SNARC effect will be relevant in chapter 7.

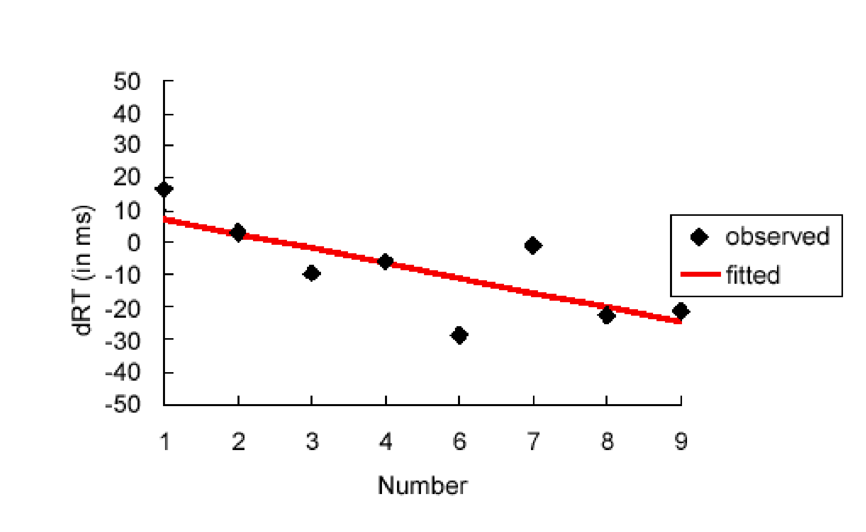


Figure 4: RT for right-sided responses minus RT for left-sided responses in a parity judgement task. Regardless of mathematical training, small numbers elicit faster responses with the left hand, while large numbers elicit faster responses with the right hand (re-printed from Fias & Fischer, 2005).

Reflecting the focus of this thesis, the phenomena discussed so far can all be interpreted as products of interference at the level number representations. The final phenomenon that deserves some mention is pitched at a rather different level; the process that people use to build those representations in the first place. Though the distinction is still contentious (e.g. Balakrishnan & Ashby, 1991, 1992; Piazza, Mechelli, Butterworth, & Price, 2002), a great deal of evidence suggests that the enumeration of small vs. large sets might implicate at least two different cognitive processes (e.g. Atkinson, Campbell, & Francis, 1976; Akin & Chase, 1978; Mandler & Shebo, 1982; Simon & Vaishnavi, 1996; Simon, Peterson, Patel, & Sathian, 1998; Trick & Pylyshyn, 1994). Like the other phenomena in this section, reaction times provide perhaps the best illustration of this distinction; in set enumeration tasks, the relationship between subjects' reaction times and the number of elements presented is significantly different for small sets (1-3 elements) vs. large sets (more than 4 elements; see Figure 5). The process employed for small sets is commonly called *subitizing*, while larger sets implicate either *counting* (a slow, sequential process that reveals the exact number of elements in a display) or *estimation* (a faster, approximate detection of large numerosities). This distinction is relevant to the discussion of computational models in section 2.3, and also to the results described in chapter 9.

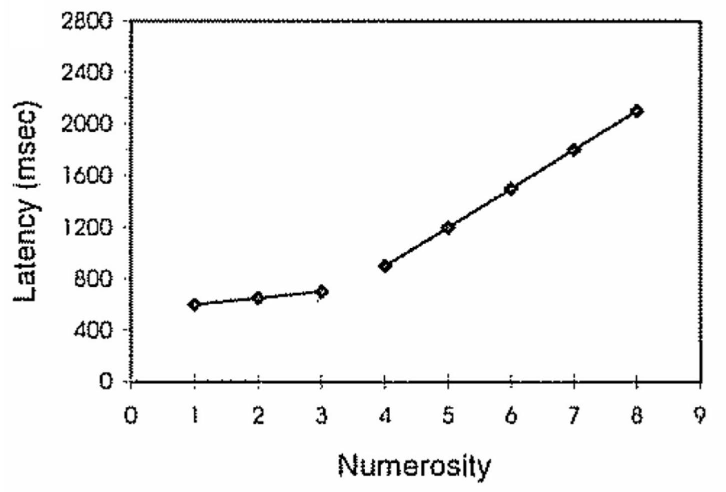


Figure 5: Prototypical reaction times in set enumeration tasks – re-printed from Peterson & Simon, 2000. For sets of 1-3, the relationship between response times and set size is significantly different to that observed in larger sets.

2.2 Mental Number Lines

Contemporary theories of cognitive quantity processing overwhelmingly refer to an intermediate, analogical form of representation: the Mental Number Line (MNL; e.g. Zorzi, Stoianov, & Umiltà 2005). There are four popular theories of MNL format – the *noisy MNL* (Figure 6a; Gallistel & Gelman, 1992, 2000), the *compressed MNL* (Figure 6b; Dehaene & Changeux, 1993; Dehaene, 2003), the *barcode* (Figure 6c; e.g. McCloskey & Lindemann, 1992; Verguts, Fias, & Stevens, 2005), and the *numerosity code* (Figure 6d; Zorzi & Butterworth, 1999; Zorzi, Stoianov, & Umiltà, 2005; Zorzi, Stoianov, Becker, Umiltà, & Butterworth, under revision).

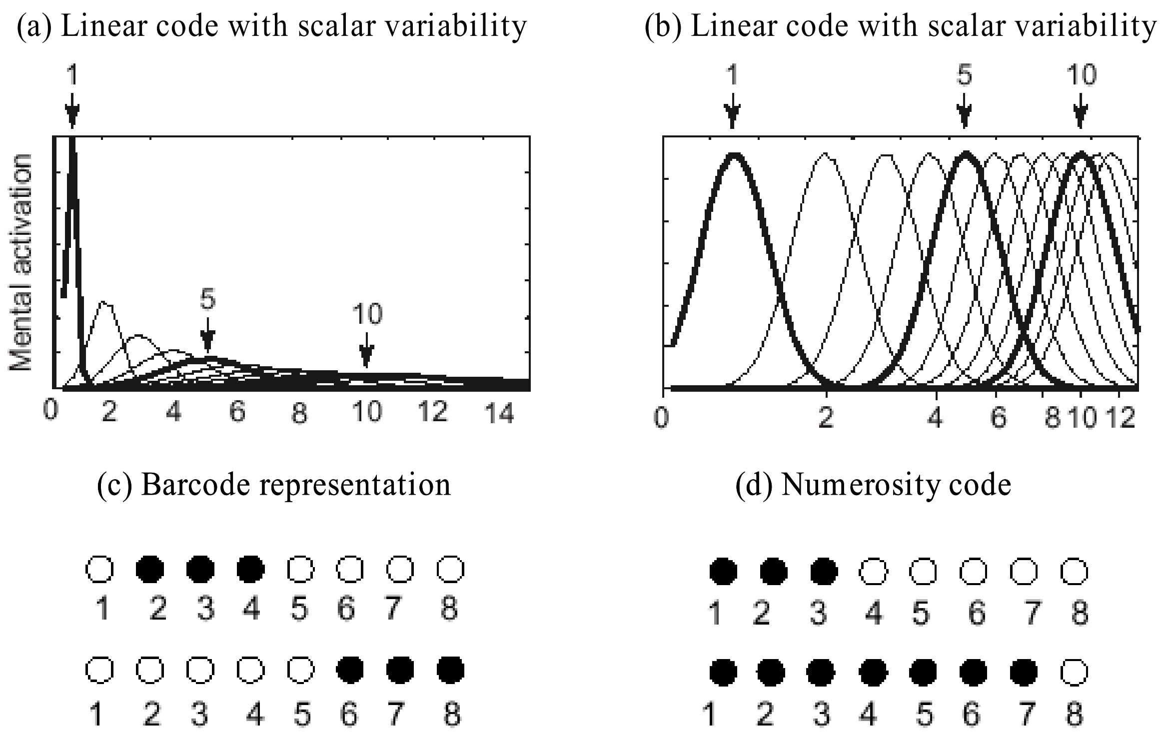


Figure 6: Four popular proposals for the format of semantic number representations. (a) The linear number line with scalar variability (Gallistel & Gelman, 1992, 2000); the *noisy MNL*. (b) The compressed number line with fixed variability (Dehaene & Changeux, 1993); the *compressed MNL*. (c) The “barcode” (or shifting bar) representation (Verguts, Fias, & Stevens, 2005); the *barcode*. (d) The cardinal accumulator representation (Zorzi & Butterworth, 1999, Zorzi et al., under revision); the *numerosity code*.

2.2.1 The Format Debate

Though independent evidence for the existence of the MNL has begun to emerge (e.g. Nieder, Freedman, & Miller 2002; Nieder & Miller, 2003, 2004; Zorzi, Priftis, & Umiltà, 2002), its original inspiration was supplied by the Size and Distance effects, mentioned previously. For example, the *noisy MNL* employs the magnitude-specific tuning of Gaussian-shaped neural receptive fields to represent particular numbers; the field centres are arranged in order of increasing numerosity (tracking the number system's ordinal structure), and their variance also increases for larger numbers. This format defines a scheme of noisy representations, with particular neurons that "prefer" magnitude X, but which are also partially activated by magnitudes X+1 and X-1, and less so by X+2 and X-2, and so on. That noise implies a potential for interference, which carries a cost in terms of accuracy and reaction times. The interference is more significant between the representations of numerically close numbers than those of numerically distant numbers, which can explain the Distance effect. And the pattern of increasing variance makes the interference between the representations of larger numbers more significant than it is for representations of smaller numbers – this can account for Size effects.

The second theory of MNL format – the *compressed MNL* – employs a very similar logic to capture Size and Distance effects, replacing increased field width (or variance) for larger numbers with decreased separation between field centres. After implementing both the *noisy MNL* and the *compressed MNL* to simulate the animal data of Brannon, Wusthoff, Gallistel, and Gibon (2001), Dehaene (2001) concluded that both formats define such similar same metrics of number similarity that most of their behavioural predictions should be identical. The implication is that these two theories can only be separated by evidence of a more directly neurophysiological sort.

One example of this kind of evidence stems from an elegant series of experiments conducted by Nieder and colleagues (e.g. Nieder, Freedman, & Miller 2002; Nieder & Miller 2003, 2004). Working with monkeys, they recorded from single cells in the primate homologue of the parietal cortex during a task requiring judgements of the number of dots in visually presented sets. Their analysis revealed magnitude-dependent responses in apparently number-sensitive neurons – yielding results that are distinctly reminiscent of the *compressed MNL* (see Figure 7).

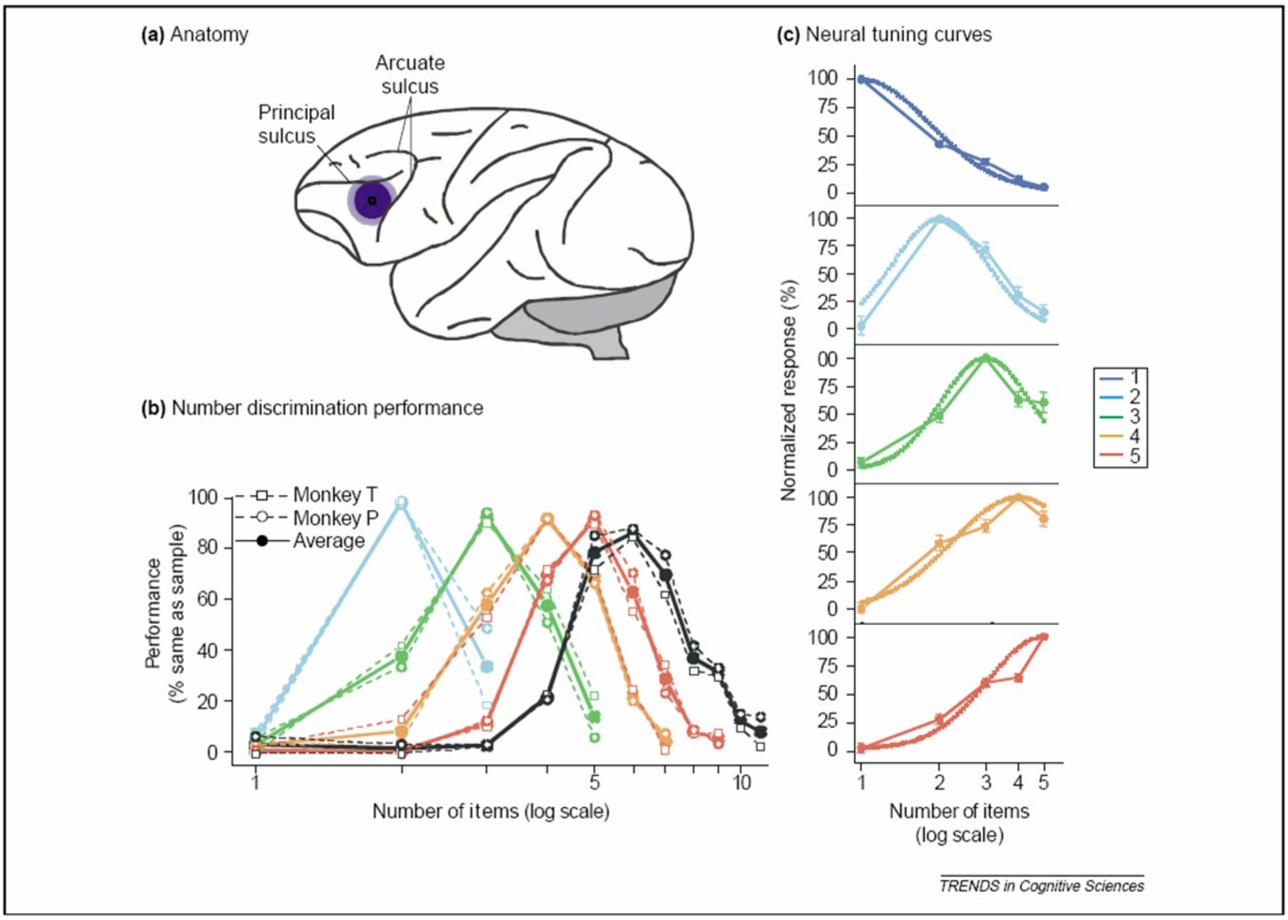


Figure 7: Evidence for logarithmic coding of number in the monkey brain – this figure is re-printed from Dehaene, 2003. (a) The anatomical location in primate prefrontal cortex from where Nieder and Miller recorded number neurons. In their experiments, monkeys were presented with a first set of dots, which they were then asked to discriminate from a second set of dots. (b) The percentage of trials on which they responded 'same' is plotted as a function of the second number (abscissa) for different values of the first number, which ranged from 2–6 during behavioural testing (colour of plot). Performance decreased smoothly with the distance between the two numbers (i.e. the peak occurs when the two numbers are the same). This distance effect assumed a Gaussian shape when plotted on a logarithmic scale. (c) So did the tuning curves of individual number neurons (shown for 1–5).

At about the same time, Dahaene and colleagues began to employ an emerging combination of functional Magnetic Resonance imaging (fMRi) and Repetition Suppression (RS) to make analogous measurements in humans (Naccache & Dehaene, 2001; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). RS is a possible neural correlate of priming phenomena – a characteristic pattern of diminished responses in neurons when the stimuli that activate them are repeated (e.g. Miller & Desimone, 1991; Squire, Ojemann, Miezin, Petersen, Videen, & Raichle, 1992; Schacter, Alpert, Savage, Rauch, & Albert, 1996; Schacter & Buckner, 1998). Piazza and colleagues (2004), exposed passive participants to a series of dot fields of varying numerosity; as expected, the RS phenomenon diminished the response of number-sensitive voxels when the numerosities of successive stimuli were identical. By tracking the recovery of that response when the fields' numerosities changed, these researchers were able to derive approximate, numerosity-specific tuning profiles that appear to confirm the presence (and habituation) of a *compressed MNL* representation (Dehaene & Changeux, 1993).

Though certainly suggestive, these results are also far from conclusive. The mechanisms underlying RS are still largely unknown (Naccache & Dehaene, 2001), so cannot support strong conclusions, and there remain concerns that apparently number-selective neurons (and voxels) may be confounded by the other features of a stimulus (for example, Piazza and colleagues, 2004, report a small mediating effect of item shape – circles vs. triangles – on the number-sensitive voxels that drive their results). The same suspicion of confounds also lingers over the work reported by Nieder and colleagues; in this case, that doubt is amplified by the extensive training required for their primate participants (if the representations are learned, their general significance may be undermined). Further, without neural stimulation and lesion data, the causal roles that these representations play cannot be confirmed. And finally, even in the absence of these doubts, these results cannot prove a negative; in deference to parsimony, we might prefer a theory of number representation that employs exactly one format – but there are good reasons to allow for more.

The possibility that different individuals might recruit different representations in numerical tasks is supported by research reported by Siegler and Opfer (2003), who tested children and adults in number line tasks. Given a series visually presented lines of 25cm length, and told the numerosity ranges that those lines represented (0-100, or 0-1,000), their participants had to mark particular numbers on each line (a number-to-position task). By analysing the pattern of responses, these researchers inferred that different participants were representing the numerosity range differently; some appeared to employ the *noisy MNL*, while others seemed to use the *compressed MNL* – and there was also a general trend toward a more linear representation with increasing age.

That linearity is important because neither of the two MNLs discussed so far can explain it – both employ a directionally asymmetrical pattern of interference (to capture the Size effect), so both imply the over-representation of the lower (left hand) end of the number line. This inconsistency cannot be dismissed as an artifact of strategy because one other phenomenon – directionally symmetrical priming (described in the previous section) – further confirms its significance. Like the line positioning strategies of adults, these priming effects are thought to emerge at the level of semantic representations – and like those strategies, it appears that these phenomena imply access to a directionally symmetrical MNL format.

In response, Verguts, Fias and Stevens (2005) have proposed a return to the rather older *barcode* format (e.g. McCloskey & Lindemann, 1992; see Figure 6c) – a Mental Number Line employing Gaussian receptive fields with constant variance and constant logical separation, projected in a linear space. In this format, the overlap (and interference) between the representations of '7' and '6' is the same as that between '7' and '8', just as the priming effect suggests. But the shift to symmetry also carries a cost – trading an account of the Size effect for an account of symmetrical priming effects.

The authors fill that gap by reverting to the older notion (e.g. Ashcraft, 1992) that Size effects are an artefact of the skewed frequency with which numbers are encountered during learning; subjects are faster when manipulating smaller numbers because these are more familiar. The implied distinction between the mechanisms underlying the Size and Distance effects can also be useful, particularly when we observe data that suggests a dissociation between them. For example, some reports suggest that the Size effect does not emerge in number naming tasks (at least for numbers up to 20: Butterworth, Zorzi, Girelli, & Jonckheere, 2001; Reynvoet, Brysbaert, & Fias, 2002), or in parity judgement tasks (Dehaene, Bossini, & Giraux, 1993; Fias, Brysbaert, Geypens, & d'Ydewalle, 1996; Reynvoet, Caessens, & Brysbaert, 2002) – though both presumably demand access to the MNL. Chapter 7 reports some further evidence along these lines.

In practice though, frequency-based accounts of the Size effect can be difficult to justify because they demand a very extreme skew in favour of smaller numbers – far greater than can be found in, for example, the standard textbooks that are used to teach mathematics (Ashcraft & Christy, 1995). Nor can Size effects be explained solely as a task-specific phenomenon, since they are also found in non-verbal counting tasks (Whalen, Gallistel, & Gelman, 1999). And though the evidence that dissociates the Size and Distance effects cannot be dismissed, it is also predominately negative – depending on the absence of Size effects in circumstances where the Distance effect is observed (e.g. Verguts & De Moor, 2005) – so also somewhat equivocal.

There is one final theory of MNL format that captures the *barcode*'s directional symmetry without sacrificing an account of the Size effect. Unlike the three formats discussed so far, the *numerosity code* (Zorzi & Butterworth 1999; Zorzi et al., under revision) describes an MNL that captures the number system's cardinal structure; the representation of '1' is a subset of the representation of '2', both are subsets of the representation of '3', and so on (see Figure 6d). As the magnitudes of numbers increase, so the proportion of their representations that is not shared with other numbers decreases (logarithmically) – this pattern can explain the Size effect (Zorzi, Stoianov, & Umiltà, 2005; Zorzi et al., under revision). But the format is also symmetrical – the absolute logical difference between the representations of numerically adjacent numbers is presumed to be constant – so should also be able to capture symmetrical priming. Aside from capturing many of the same phenomena as the other MNL formats (Zorzi et al., under revision), this code is useful because increments in the magnitude of representations directly correspond to increments in the numerosity of their referents; that property should be useful in other numerical tasks like enumeration (Zorzi, Stoianov, & Umiltà, 2005).

All of these formats – and perhaps others as well – may play some role in the processes underlying basic numerical skills. Despite that space, the debate that surrounds MNL formats is still largely a fight for supremacy – and unlikely to be resolved in the near future. To date, the only concrete proposals that relate multiple formats together stem from the detailed structures of particular computational models of numerical processes; these are discussed in the next section. These computational contributions are just some among many that might be made to this debate – chapters 5 and 8 report two more.

2.2.2 The Structure Debate

For all their differences, the MNL formats discussed so far all share a common limitation; unlike symbols, a purely analogical representation cannot be *systematic*, in the sense defined by Chomsky (1959). Without that systematicity, a finite pool of resources implies a finite capacity for representations – a limit that sits uneasily with the open-ended nature of the number system itself. This apparent contradiction raises a question; how might larger numbers be represented in the brain?

Congruence effects (mentioned in section 2.1) offer one clue to the answer, suggesting that the single-digit components of multi-digit numbers can play a cognitive role that is independent of the role of their integrated values. The implication is that humans can deploy several MNLs at a time – one for each of the digits in a multi-digit string. *Decomposed*, or *componential* MNLs can be considered a replacement of the traditional, *holistic* view (e.g. Verguts & DeMoor, 2005), or offered as part of a *hybrid* account in which both the components of a multi-digit number and their integrated values are assumed to play some explicit role (e.g. Nürk et al., 2001; Figure 8 illustrates all three options). This latter option implies the activation of three distinct representations when subjects process a two-digit number – one for the number itself and one for each of the number's two digits. *Hybrid* accounts are attractive because they offer an intuitive way to explain both apparently *holistic* phenomena (the Size and Distance effects) and apparently *componential* phenomena (Congruence effects) – yet though both types have been reported, *hybrid* accounts are still rather controversial.

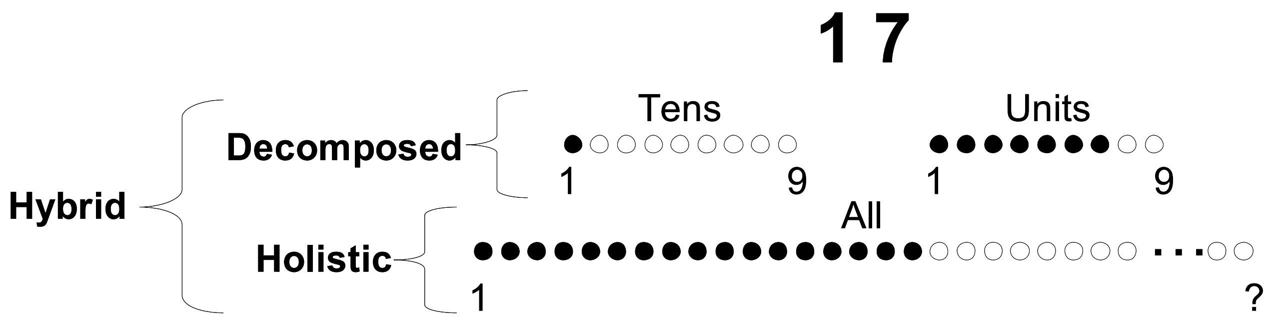


Figure 8: Three proposals for the structure of the Mental Number Line. The Holistic MNL implies a line that tracks numbers' integrated values (the question mark indicates enduring uncertainty concerning the upper bound on holistic number representation), while the Decomposed MNL implies discrete representations for each digit in a multi-digit number. The Hybrid proposals assumes that both holistic and decomposed representations play some role. For the purpose of illustration, the numbers themselves are represented with the *numerosity code*.

One critical source of doubt stems from the equivocal nature of the data in this area. For example, Verguts and DeMoor (2005) compare subjects' performance in a two-digit number comparison task with same-tens pairs vs. pairs in which the tens distance was exactly '1'. From the *absence* of a Distance effect in the latter case, they argue that subjects are only representing the numbers' components. Though plausible, this result appears to contradict that reported by Dehaene, Dupoux and Mehler (1990), who argue for a *holistic* view from the *presence* of a Distance effect that appears to cut seamlessly across decade-boundaries. And even if we assume that inconsistency away – perhaps as an artefact of misinterpretation – there remains a rather bigger problem to solve. Verguts and DeMoor's data imply that, in different-decade pairs, the units digits of each number play no significant role in the participants' reaction times, while the Congruence effects discovered in other two-digit number comparison experiments (Nürk et al., 2001) imply interference between the tens-digit and units-digit comparisons. If trailing digits can be ignored when they are irrelevant to a particular problem, it seems difficult to explain how that interference can emerge.

The second source of doubt is the natural confound between a multi-digit number's components and its integrated value. It is far from clear that results of the form just discussed – regardless of their direction – can tell us anything conclusive; a different-tens Distance effect (e.g. Dehaene, Dupoux, & Mehler, 1990) could simply indicate a residual dependence on unit distance, while the interpretation of its absence (Verguts & DeMoor, 2005) will always be susceptible to criticism. When Nürk and colleagues (2001) reported their Congruence effects they attempted to resolve this issue by regressing the participants' reaction times against a wide variety of predictors, tracking both the numbers' integrated values and the values of their components. The best predictor was the logarithmic numerical distance between those integrated values, suggesting a particular sensitivity to the numbers' integrated content. But the unique contribution of this predictor was small, or only about 1% over and above that captured by the numbers' single digits. The data favour a hybrid interpretation, but cannot support very strong conclusions.

One way to explain this rather equivocal picture is to suggest that multi-digit number processing experiments are tracking a moving target; like MNL formats, MNL structures might be deployed in a context-sensitive manner. Recent work by Ganor-Stern, Tzelgov, and Ellenbogen (2007) supports this view with a series of experiments on the Size Congruence Effect (SiCE). The SiCE is found in physical size judgements tasks with numerical stimuli – participants are slower to make them when the physically smaller number is also the numerically larger number in a pair (or vice versa). In a series of experiments with two-digit number stimuli, these authors reported that the SiCE is mediated by the numerical magnitudes of the numbers' single-digit components (including their unit-decade compatibility), but not by the numbers' integrated values. By contrast, both of these factors appeared to be independently significant in their number comparison task. The implication is that holistic representations of multi-digit numbers are not deployed automatically, but that people can deploy them when they are relevant to a task. Chapters 7 and 8 report work that can contribute to this debate.

2.3 Computational Models of Numerical Cognition

Questions concerning the role of computational methods in the study of cognition are as much a focus of this thesis as numerical cognition itself; they will recur throughout the chapters that follow. This section provides a summary of past answers – a brief account of the way that computational modelling has been employed in the study of numerical cognition in the past. Reflecting the focus of this thesis, the material that follows is primarily concerned with the ways in which computational methods have been used to clarify debate on the structure of semantic number knowledge.

The earliest Connectionist models of numerical processes were all designed to capture simple arithmetic operations like addition and multiplication. Zorzi, Stoianov, and Umiltà (2005) divide these models into two groups: learning models, in which a set of associations is built by an algorithm that requires repeated exposure to a set of training problems, and performance models, in which the parameters that drive the required mapping are defined directly by the models' designers. Two of the earliest models of arithmetic – the network proposed by Viscuso, Anderson, and Spoehr (1989), and MATHNET, designed by McCloskey and Lindemann (1992) – are learning models. Both represent complete arithmetic facts (e.g. 4 * 5 = 20) as activation patterns on a set of network units, and each uses a (different) learning algorithm to to build an association between the components of those facts ('4 * 5' and '= 20'). After learning, both models can complete partial facts (e.g. complete "4 * 5" with "= 20") after undergoing some number of processing cycles; that number is naturally interpretable as a reaction time, which can then be compared to empirical data.

Of the two, MATHNET is probably the stronger model; it uses a more effective learning algorithm (similar to that used by Ackley, Hinton, and Sejnowski, 1985) than its counterpart

(Viscuso and colleagues' model could only learn about 70% of single-digit multiplication facts), can be lesioned to reproduce some of the problems with arithmetic that brain-damaged patients experience (Lories, Aubrun, & Seron, 1994), and displays a strong Problem Size effect (the correlation between fact retrieval time and the sum of a problem's operands was 0.69). Further, MATHNET makes comparatively few assumptions about the representations of problems; both models use a linear *barcode*-like scheme to represent numbers, but Viscuso and colleagues also employed units that were dedicated to the numbers' names. MATHNET also provides a good illustration of the limits of the *barcode* representation, since its account of Problem Size effects depends entirely on a training regime that makes problems with small operands much more frequent than problems with large operands (McCloskey & Lindemann, 1992).

As mentioned previously, the *barcode* is the only popular theory of MNL format that should require this kind of frequency manipulation – but recent results, reported by Zorzi and colleagues (under revision), suggest that the practical picture may be more complex. Working with MATHNET-like learning models of simple arithmetic – but without fact frequency manipulation – these authors report reliable Problem Size effects when arithmetic problems were represented with the *numerosity code,* but *not* when the same problems were represented with the *compressed MNL*. Both of these representations imply greater interference with greater numerical magnitude, but only one appears to be able to put that pattern to work in practice. This result is confusing because, in other models, logarithmic magnitude representations like the *compressed MNL* do seem to be effective. For example, there are two well-known performance models of arithmetic that employ this kind of code (Campbell, 1995; Whalen, 1997) – and both exhibit robust Problem Size effects.

However, these results cannot tell us anything very conclusive about the empirical implications of particular representations, because the connection depends on *both* the representations employed *and* the host of other parameters that make these models work. In performance models, those other parameters (like the weights of connections between input and output layers) are usually selected to provide best fit to the data; that process can make number representations largely incidental to generation of robust Size effects. This point is evident in the work of Verguts and Fias (2004), who present a performance model of multiplication that employs a topographic memory of facts (i.e. knowledge of the fact '7 x 7 = 49' is stored closer to that of '7 x 8 = 56' than to '3 x 4 = 12'). The memory's structure promotes spreading activation among related facts, and the resulting interference can account for the Problem Size effect (as well as some others, like the Tie and Fives effects, that will not concern us here) without making any assumptions at all about the format of number representations. Performance models are attractive because they are flexible – but in the context of the format debate, that flexibility can also be a weakness. Armed with hand-coded (if arguably plausible) structures, the empirical behaviour of these models can be difficult to connect to the properties of the representations that drive them.

The same criticism applies to performance models of single digit number comparison, in which the problem is a number pair (e.g. '6 vs. 7') and the answer a decision between them (e.g. '7 is larger'). One example of this kind of model is Grossberg and Repin's Spatial Number Network (SpaN; 2003), which produces robust Number Size effects with a representation similar to the *noisy MNL*. But the evidence of Dehaene and Changeux's (1993) *learning* model of the same process (this time with the *compressed MNL*) is not so easily dismissed. Like the SpaN, this model produces robust Number Size and Distance effects – but unlike the SpaN, this model employs Hebbian learning (modulated by a reward signal), without fact frequency manipulation, to connect its representations to task responses. That result was also confirmed by Zorzi and colleagues' (under revision) own learning model of number comparison; Size effects in simple arithmetic may be elusive, but Size effects in number comparison seem to be well-captured by the properties of the *compressed MNL*.

To some extent, the distinction might reflect the difference between these two phenomena; both imply increasing response times with increasing numerical magnitudes, but each emerges in a different functional context. But at least in terms of its apparent connection with the properties of other MNL formats, the Number Size effect does seem to "play the same game" as the Problem Size effect. For example, when learning models use linear analogue representations like the *barcode*, Number Size effects (like Problem Size effects) will only emerge when the training set is skewed in favour of small numbers (Zorzi, Stoianov, & Umiltà, 2005). And the *numerosity code* appears to explain Number Size effects in exactly the same way as it explains Problem Size effects – as a product of the decreasing discriminability of number representations as their magnitudes increase (Zorzi et al., under revision). In other words, there is an inconsistency here that has yet to be resolved.

The *numerosity code*'s linear symmetry has also been shown to support symmetrical priming (Zorzi, Stoianov, Priftis, & Umiltà, 2003) in a learning model that performs transcoding – mapping from semantic number representations to symbols (e.g. Arabic digits) and vice versa. After training, the model could activate the appropriate semantic code given only the corresponding symbolic code. To capture numerical priming, the authors tested transcoding for numbers 4-9, with primes ranging from n-3 to n+3 (following the approach of Reynvoet, Brysbaert, & Fias, 2002); particular trials involved pre-activating the model's semantic field with the relevant prime, before clamping the target to its symbolic field. Like most of Zorzi and colleagues' other models, this system's response times were defined as the number of processing cycles required to reach a steady state. And as predicted, those response times are symmetrically mediated by the numerical distances between targets and primes. Like Problem and Number Size effects, these priming effects can also be captured by careful hand-coding, even when the representations that drive a model seem unequal to the task. For example, Grossberg and Repin's SpaN (2003) produces symmetrical priming effects with an asymmetrical representation similar to the *noisy MNL*, but that symmetry owes rather more to the model's detailed structure – specifically, to the progressive deactivation of the number line after numerical stimuli are removed – than it does to the representations employed.

Two of these models – Grossberg and Repin's SpaN (2003) and Dehaene and Changeux's (1993) number comparison model – also include hand-coded accounts of the process that builds number representations from lower-level perceptual information. The details of each account are different – Dehaene and Changeux favour an associative system, while Grossberg and Repin's approach employs a serial accumulation of sequences of items – but each also shares one important feature in common. In both cases, the output of the pre-processing stage (which is mapped directly onto also the authors' favoured number representations) is a linear magnitude representation similar to the *numerosity code* (see Figure 9; these codes are also similar to the older "thermometer" system proposed by Meck and Church, 1986). The same connection is also evident in Verguts and Fias' (2004) model, which employs Hebbian learning to associate both approximate numerosity (represented with the *numerosity code*) and symbols (a localist representation) with numerosity-detectors (structured according to the *barcode*). The central conclusion of this work is that symbolic input representations can "tune" number-sensitive neurons, reducing the variance in their responses.

From the discussion so far, it should be clear that this kind of representation is, if anything, more powerful than any of the popular alternatives; these models can therefore be accused of inserting an unnecessary stage into the process. Moreover, neither of these models has much to say about the distinction between subitizing and more general numerosity estimation; Dehaene and Changeux's (1993) model is limited to small set sizes, and Grossberg and Repin's (2003) model assumes that all sets are perceived in serial order, regardless of the number of items that they contain. And the only Connectionist model that *does* capture the distinction – proposed by Ahmad, Casey, and Bale (2002) – also employs a thermometer-like representation (as input to a one-dimensional self-organising map; Kohonen, 1995), so further underlines the ubiquity of this kind of "accumulating" code. In other words, the connection between a computational model and the theories it supports can be complex; even apparently non-critical features can be theoretically significant.

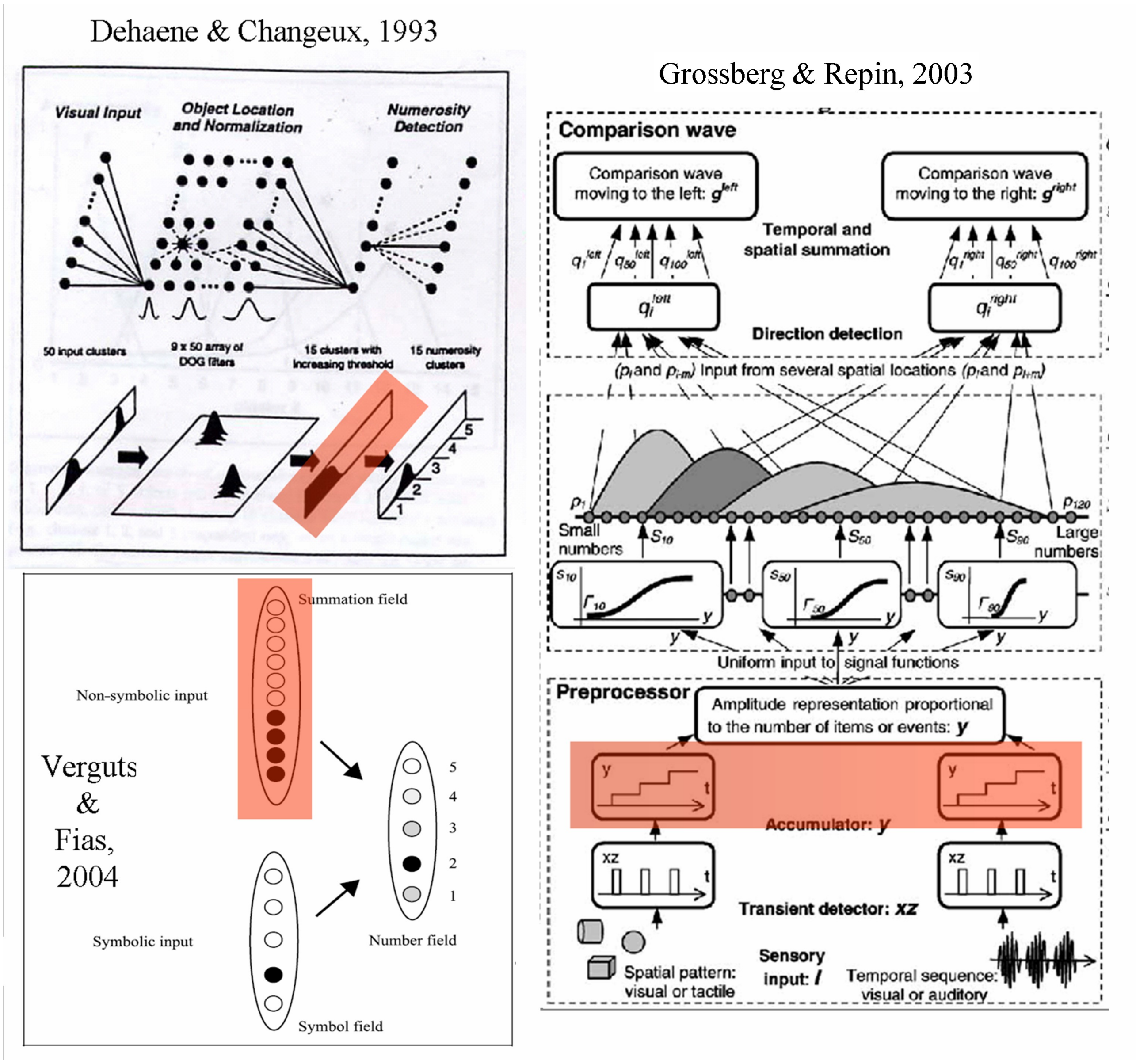


Figure 9: Illustrating the prevalence of accumulator-like codes in computational models of numerical cognition; these models actually encode a preference for either the *compressed MNL* (top left), the *barcode* (bottom left) or the *noisy MNL* (right), but all three also use an accumulator, similar to the *numerosity code*; the relevant features of each model are highlighted in the figure. Each of the these three diagrams are re-printed from the authors' own published papers, which are also referenced above.

The format debate (section 2.2.1) is clearly quite well-captured by these models' results – broadly, they confirm that particular MNL formats really do have the empirical implications that their authors predict. But the structure debate is largely absent from the history that these models compose, which reflects the comparatively minor role that multi-digit numbers have played in the field so far. Chapter 8 begins to redress that balance, with a suite of models of multi-digit number comparison.

These models also demonstrate how the properties naturally associated with particular representations can be reproduced by learning with skewed training sets, or with carefully hand-coded architectures. That flexibility is important because, in some cases at least, pure associative memory seems unequal to the task of modelling particular cognitive skills. For example, simple arithmetic is generally agreed to involve "...some mixture of fact retrieval from memory *and procedures for transforming the problem if the memory search fails.*" (Zorzi et al., under revision, page 9, my emphasis; see also Cambell & Xue, 2001; Groen & Parkman, 1972; LeFevre, Sadesky, & Bisanz, 1996), and none of the learning models discussed so far even attempt to capture that latter, procedural component.

But though the cognitive scope of associative learning may be limited, these models still make an important contribution. Indeed, from the perspective of the format debate, (unstructured) associative learning may be the most effective of the model-building approaches discussed so far – because it cuts through the levels of indirection that can dissociate the logical properties of particular representations from their practical, empirical consequences. Chapters 3, 4 and 5 propose, develop, and apply a new methodology that can reproduce – and perhaps even surpass – that kind of directness.

II

On the Processing of Small Numbers

Chapter 3

Dynamicism and Cognitive Science

This chapter begins with a discussion of the assumptions that constrain conventional modelling methodology (3.1), and suggests that they might be circumvented by a Dynamicist approach, which is introduced in section 3.2. These sections define the context that drives the new models described in chapters 4 and 5.

3.1 The Limits of Convention

In the last chapter, we saw that Connectionist models of numerical cognition tend to employ representation-level interference, structured learning, or hand-coded architectures (or some combination of the three) to capture empirical phenomena. All of these options have a plausible role to play, and each can be well-justified – perhaps even necessary – in the right circumstances. But in the particular context of the format debate, computational models can make a clearer contribution when they are *transparent*; assumptions that obscure a representation's influence can also weaken the evidence that it can supply. That logic motivates a preference for unstructured learning over both structured learning and extensive hand-coding – but Connectionist learning is far from assumption-free.

Like any other Connectionist architecture, associative memories depend on a host of architectural assumptions to make them work. In some cases at least, those assumptions are certainly *not* biologically plausible. Supervised learning systems, like error backpropagation networks (Rumelhart & McClelland, 1986), depend on unit-specific error signals that are hard to justify from what we know of neurophysiology (e.g. Crick, 1989). But often, even *un*supervised learning is only just barely justifiable. For example, the mean field Boltzmann machine (Ackley, Hinton, & Sejnowski, 1985; also used by Zorzi et al., under revision) employs only locally available information (pre- and post-synaptic firing rates), but requires symmetrical network connectivity (the weight of the connections from units A to unit B is the same as the weight of the connection from unit B to unit A), and a carefully structured two-stage learning regime (more on this in chapter 8); neither of these two features makes much biological sense. Connectionist learning models can be biologically plausible – but more often than not, that plausibility is available only because of how much we *don't* know about the relevant neurophysiology.

Indeed, even the training regimes that drive “unstructured” learning assume a great deal of structure – in particular, that number processing problems are experienced in a neat, serial order. Though two of the models described in section 2.3 (Dehaene & Changeux, 1993; Grossberg & Repin, 2003) make some attempt to capture lower level perceptual processes in numerical cognition, none comes close to a properly closed sensor-motor loop. In the context of the format debate, it makes sense for models to avoid too much architectural complexity, so a focus on functional modules can be justified as fitting a model's scope to the problem that it is designed to address. But the modularity assumption is itself open to criticism (Brooks, 1991; Harvey, 1996). Module interfaces can have a profound impact on model behaviour, so model-designers will usually demand great control over them – but it is precisely that control that must be surrendered if the modules of today are ever to be integrated into the embodied, behaving whole of tomorrow (Brooks, 1991). In other words, the modularity assumption may actually amplify the problem of behavioural integration; by assuming away that complexity, modular models may also mislead.

Finally, conventional models of numerical skills tend to ignore – and can actually obscure – questions concerning the origin of those skills. Hand-designed models certainly do not address them, since they can only capture those skills in their “mature” form. Learning models can be interpreted as attempts to capture *both* the skill under study *and* its acquisition – but there are good reasons to suspect that at least some basic numerical skills are not learned at all. In recent years, a wealth of evidence has emerged which suggests that pre-linguistic infants (Feigenson, Carey, & Hauser, 2002) and some animals – from apes (Biro & Matsuzawa, 2001) and monkeys (Nieder, Freedman, & Miller, 2002) to pigeons (Brannon, Wurstoff, & Gallistel, 2001) and salamanders (Uller, Jaeger, Guidry, & Martin, 2003) – are sensitive to quantity. Though learning is clearly involved in the development of an adult human's number sense, this evidence supports the theory that some rudiments of that sense – an "evolutionary start-up kit" (Butterworth, 1999) – are genetically determined. This evidence is also particularly relevant to the MNL format debate, because instances of the phenomena that drive it have also been observed in animals (see Gallistel & Gelman, 1992 for a review, and more recently, Brannon & Terrance, 2003; Cantlon & Brannon, 2006; Jordan & Brannon 2006a, 2006b). Since these phenomena are thought to emerge at the level of the MNL, the implication is that this genetic contribution might include that MNL. And at least as currently employed, computational cognitive modelling methods offer us no principled way to explore that phylogenetic contribution.

Taken together, these three themes – architectural assumptions, behavioural integration, and the role of evolution – motivate an openness to methods that go beyond conventional Connectionist learning. The next section describes a framework that can address all three.

3.2 Dynamicism – The Third Contender

Dynamicism (Van Gelder, 1995, 1998, 1999) is an umbrella term, intended to capture what has been variously been described as Evolutionary Robotics (Harvey, Husbands, & Cliff, 1996), the Embodied Cognition approach (Clark, 1999), the Adaptive Behaviour approach to cognition (Bakker, 2001), and the Dynamical Systems perspective on cognition (Beer, 2000). First described as such by Eliasmith (1996), Dynamicism is the "third contender" only in the sense that

"Structuralism" (e.g. Cooper & Shallice, 2006) is the first, and Connectionism (e.g. Feldman & Ballard, 1982) is the second. Eliasmith himself was not convinced by this distinction – and the discussion that follows will further undermine it – but it does provide a useful way to understand what Dynamicism is and how it might (and might not) be related to its more conventional counterparts.

3.2.1 Structuralism and Connectionism

Structuralism is a relatively new term for a rather older idea – employing the Computer Metaphor (see Crowther-Hyke, 1999, for a good definition) to drive cognitive models and theory. Terms such as "classical cognitive science", and "Good Old-Fashioned Artificial Intelligence" (GOFAI) capture the same essential perspective, casting cognition as a product of the information-processing structure of the brain. Structuralism articulates the intuition that the logic of that structure should somehow correspond to the logic of the behaviour that it drives – for example, that goal-directed behaviour implies the explicit representation of goals by the cognitive system (e.g. Cooper & Shallice, 2000). This is a powerful idea because it implies that there should be an intelligible neural counterpart to our more familiar concept of "knowledge"; cognitive neuroscience can then be construed as the search for a way to "translate" the folk psychology (e.g. Stitch & Ravenscroft, 1994) that we naturally employ into the more verifiable and quantifiable language of neurons and neural behaviour.

Connectionism was first introduced as a potentially very different approach, associated with a suite of practical benefits – principally "brain-like" architecture (networks of simple processing units and weighted connections between them), fault tolerance and "natural" learning (Aleksander, 1989; Feldmand & Ballard, 1982; Honavar & Uhr, 1989; McClelland & Rumelhart, 1986; McCleod, Plunkett, & Rolls, 1998; Seidenberg & McClelland, 1989; Sejnowski & Rosengerg, 1987 – but also see Berkeley, 1997). Conventional, symbolic architectures can also be designed to capture these benefits (Fodor & Pylyshyn, 1988); Connectionism is valuable more for the elegance with which it supplies them, than for their presence in itself. And the benefits were also thought to be balanced by a cost – an apparent resistance to computational interpretation. At the time, that resistance inspired some very extreme interpretations; some researchers claimed that Connectionism had made older, more Structuralist concepts (like representations) irrelevant (Ramsey, Stich, & Garon, 1991), while others dismissed the new approach as "mere implementation" (e.g. Fodor & Pylyshyn, 1988). The emergence of Dynamicism has inspired almost exactly the same debate.

3.2.2 Dynamicism Defined

Dynamicists claim that cognitive agents are best understood as dynamical systems, and that cognition emerges from the close-coupled interaction between an agent and its environment. At a formal level, a dynamical system is simply a mathematical object that describes how the state of a system evolves through time – or more specifically, a triple <T, S, F>, where T is an ordered time series, S is a state space, and F is a function that describes the evolution of subsequent states from current states (Beer, 2000). Thus defined, even desktop computers will support a dynamical description. Dynamicism – like Connectionism (Fodor & Pylyshyn, 1988, Bechtel & Abrahamsen, 1991) – is at least potentially consistent with even the most stringent Structuralist intuitions (e.g. Crutchfield, 1998). So – like Connectionism – the value of the Dynamicist approach stems more from a difference in emphasis than a difference in principle (Beer, 2000); where Structuralism emphasises the information-processing structure of the brain (e.g. Johnson-Laird, 1988), and Connectionism emphasises fault tolerance and natural learning (e.g. Rumerlhart & McClelland, 1986), Dynamicism emphasises the interactive, spatially and temporally situated nature of the whole, behaving agent (e.g. Port & Van Gelder, 1995).

Though perhaps rather abstract in itself, that distinction underlies some very concrete differences in practice. The first difference flows directly from a focus on the whole, behaving agent; unlike their more modular counterparts, Dynamicist cognitive models almost always include agents, with bodies, and the environment that they inhabit (e.g., Beer, 1996; Blumberg, 1995; Harvey, Husbands, Cliff, Thompson, & Jakobi, 1996; Seth, 1998). That trend naturally alters the focus of Dynamicist models – the second difference – away from the information-processing structures that might explain cognitive behaviour and toward the behaviour itself.

A focus on whole agents also raises some rather significant technical challenges. The third (and for our purposes, final) difference that sets Dynamicism apart can be understood as a response to these challenges – a preference for design by behaviour-based selection (using genetic algorithms: see Goldberg, 1989, for a review). Often, neural networks – or systems of equations inspired by the Hodgkin-Huxley description of neurons (Hodgkin & Huxley, 1952) – are the foci of Dynamicist model-building methods, both because they are thought to yield "brain-like" models, and because small changes to particular network parameters (the driving force of behaviour-based search) tend to invoke small changes in global network behaviour. Models of this sort are also naturally comparable to their Connectionist counterparts; I focus exclusively on these in the chapters that follow. But other alternatives – for example the classical equations for simple oscillators like pendulums (which can naturally express the dynamics of ballistic limb movements: e.g., Feldman, 1966) – can work just as well.

3.2.3 Benefits and Costs

Dynamicist models emphasise all three of the themes identified in section 3.1. A focus on the whole, behaving agent may make for tough, model-building challenges, but it also ensures that successful Dynamicist models will seamlessly integrate the cognitive process being studied with their more "natural" behavioural context. And though certainly pragmatic, the preference for behaviour-based selection can also be justified as encouraging an "assumption-light" approach to model-building – free of some of the architectural constraints that more conventional design methods impose (e.g., Harvey, 1996; Beer, 2000). By focusing on an agent's interaction with its environment, this approach can release designers from the need to, for example, specify an internal modular architecture – replacing that designer-driven constrain with the opportunity for more problem-driven emergence. To the extent that these “minimal” (e.g. Nowack, 2004) models are successful (a judgement that remains contentious: e.g., Cooper & Shallice, 2006), there are good reasons to prefer them to their more constrained counterparts. Finally, if given the right task, behaviour-based selection is also a promising model of evolution, so Dynamicist models can be used to explore the role of phylogeny in cognitive development.

However, like Connectionism, Dynamicism appears to undermine some of our most basic, Structuralist intuitions. With minimal architectural constraints, behaviour-based selection permits the recruitment of a very broad range of chaotic dynamics in the generation of cognitive behaviour. And though nothing in the logic of this approach forbids the emergence of recognisable computational structure, that structure simply does not seem to emerge from this chaos (Beer, 2000, 2003). Just as it did for Connectionism, this apparent inconsistency has inspired a range of interpretations, from extreme Eliminativism (e.g. Harvey, 1996), to outright dismissal (e.g. Lewin, 1992, page 164). The only obvious compromise in unsatisfying at best – to assert, without much justification, that representations (and other classical structure) *will* emerge when (or if) Dynamicist models capture sufficiently “cognitive” behaviour. Clearly, another rather more practical compromise is needed.

3.3 Toward a Compatibilist Dynamicism

The apparent connection between the debates that Connectionism and Dynamicism have inspired is important because the former appears to have been resolved. On closer inspection, Connectionist models have turned out to be entirely consistent with our more classical intuitions – just as predicted by Bechtel and Abrahamsen (1991), with a position that they called Compatibilism.

Compatibilism articulates the familiar intuition that apparently antagonistic approaches may, in practice, turn out to be completely compatible. In principle, Dynamicism is certainly susceptible to Compatibilist interpretation (Crutchfield, 1998) – though in practice, the features that drive this interpretation (like recognisable representations) simply do not seem to emerge in Dynamicist cognitive models (Beer 2000; Beer 2003). But appearances can be deceiving. My contention is that classically recognisable structure *can* be discovered in Dynamicist cognitive models, once we know how to look for it.

The key to that contribution is a novel method for analysing Dynamicist models, founded on Marr's (1982) distinction between different levels of description, and on Smolensky's (1987) concept of Approximationism. The former emphasises the notion that a single system can support multiple (apparently distinct) accounts, while the latter articulates the intuition that Structuralist accounts of cognition can be useful, and approximately correct, without necessarily capturing every detail of the underlying causal process. The implication is that we can search for – and discover – classical structure in dynamical systems, while accepting that the implied "computational story" will be at best a good approximation to the underlying "causal story". Critically, I will also propose a metric that quantifies the correspondence between these two levels of description – the extent to which a Dynamicist model justifies a Structuralist interpretation.

The next two chapters chart a practical route toward a Compatibilist Dynamicism. Chapter 4 is a practical illustration of the discussion so far, illustrating the Dynamicist approach, its benefits, and costs – as well as their possible resolution – with a model of categorical perception. Chapter 5 applies the same approach to the problem of number comparison, and makes a novel contribution to the format debate.

Chapter 4

Dynamicism and Categorical Perception

In the last chapter, we saw that Dynamicism offers three important benefits in cognitive models: elegant behavioural integration, minimal model architectures, and a transparent way to capture the impact of phylogeny. On the other hand, Dynamicist models have tended to appear to eliminate features like cognitive representations. This chapter develops and applies a method – the Behaviour Manipulation Method (BMM) – that can discover classically recognisable representations in Dynamicist models.

One early example of the Dynamicist approach was a model designed to address categorical perception – an agent's ability to assign sensed objects to discrete categories (Beer, 1996). This domain is the closest thing that Dynamicism has to a standard environment, and is therefore a good medium for the introduction of new methodology. For pragmatic reasons, the current implementation of this system is not an attempt at precise replication – the goal was simply to generate agents that can perform this familiar task. But the differences between this version and its precursors are of secondary importance to the analyses that follow.

4.1 Model Design

The environment is a 2-dimensional square plane, with sides measuring 100 units. Positions on this plane are denoted with the notation <X, Y>. The agent is a circle of radius 5, which begins each run at the centre of the plane's lower boundary (i.e. with centre <50, 0>). Agents are exposed to series of trials in which shapes (squares or circles) “fall” from the square's upper boundary toward its lower boundary; the trial ends when a shape touches either the agent or the x-axis. The agents' goal is to categorise the shapes that fall toward them, catching (i.e. touching) circles, while avoiding squares.

Shapes fall with a speed of 0.5 units per time step, and occur with a range of possible radii (3-6). Squares are specified relative to the circumcircle defined by their radii, and also occur with random rotation. Together, these two sources of variation (size and rotation) complicate the relationship between apparent shape width and actual shape type – a confound identified by Beer (2003). Each shape starts with a random X position, but their centres will always fall within two 10-unit bands, one on each side of the agent's starting position (i.e. at the start of each trial, shape centres have a Y value of 100, and X values in the ranges 20 to 30, or 70 to 80); this latter restriction was intended to eliminate shapes that fall from directly above the agent, since these special cases have previously been shown to raise particular problems in the past (Beer, 2003).

Agents are rate coded, continuous-time dynamic recurrent neural networks, updated synchronously in time steps. The activity $\boldsymbol{u}$ of unit $\boldsymbol{i}$ at time step $\boldsymbol{t}$ is calculated using equation 1:

$$u_i(t) = u_i(t-1) + (1/\tau_i)\sigma\left(\sum_{j=1}^{N} w_{ji}(u_j(t-1))\right) \qquad (1)$$

where $\boldsymbol{w_{ji}}$ is the weight of the connection from unit $\boldsymbol{j}$ to unit $\boldsymbol{i}$, $\boldsymbol{\sigma()}$ is the sigmoid function (bounded in the range 0-1) and $\boldsymbol{\tau}$ refers to a unit-specific time constant (higher time constants indicate a lower dependence on incoming activity).

The agents' visual systems are analogous to a laser range-finder. Seven rays project upwards from the centre of each agent, spanning a 60° angle – adjacent rays subtend an angle of 10°, and each ray has a length of 110 units. If a ray intersects with a shape, activity is passed to its associated sensor unit; the value passed is inversely proportionate to the distance to that intersection. Agents can move in only one dimension, along the x-axis – at each update, the change in an agent's position is proportionate to the difference between the activity values of two effector units, with a maximum speed of 5 units per update. To improve the similarity between our agents and those analysed in previous work, we also restrict the agents' hidden layers to include exactly 7 hidden units (as in Beer, 2003). With the exception of the sensor units, which are always fixed by properties of the "world", so receive no incoming connections, the agents' neural networks are universally connected; every unit is directly connected to every other, and to itself. Figure 10 presents a schematic of the agents' network architectures (left), and an illustration of an agent in its environment (right).

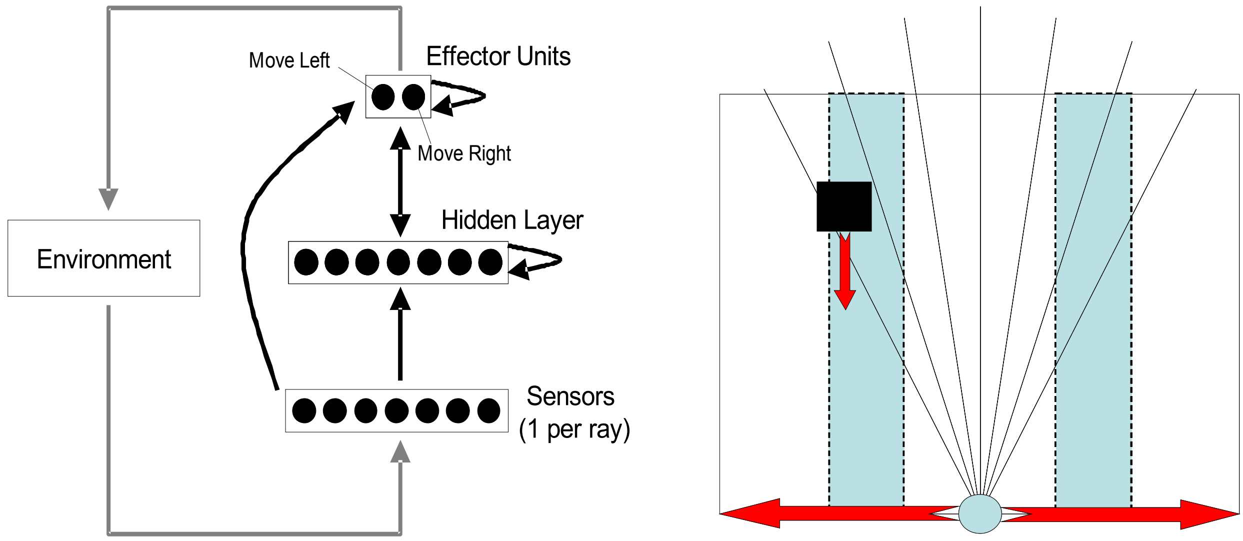


Figure 10: (Left) Schematic network structure for agents designed to solve a visual object classification problem. The agent has 7 sensor units, 7 hidden units, and two effector units. The sensor units' activities are always fixed by the agent's environment, but the hidden and effector units all receive direct input from every other unit, and from themselves. (Right) An agent in its environment. Shapes fall from the environment's upper boundary toward its lower boundary – their centres always fall within one of the two, shaded areas. The agent's task is to touch falling circles, and to avoid falling squares; they can move in only one dimension, along the environment's lower boundary.

The agents in this system, which were initially specified at random, were designed with a Microbial genetic algorithm (Harvey, 2001). During each iteration, two "parents" are randomly selected from the population and compared. The weaker of the two parents is replaced by their "child", which is defined by mixing the parents' weight vectors and time constants (each parent contributes a parameter with 50% probability), and applying a mutation. The mutation operator usually implements a small, random change (± 0.01) to a randomly selected weight, but will sometimes (p = 1%) increment or decrement a unit's time constant instead.

An agent's fitness score is the sum of the absolute distances between that agent and each of a set of 100 shapes at the end of each of 100 trials; each set is composed of 50 pairs of shapes, identical in every respect but for their type. Distances to circles (which should be caught) are counted negatively, and distances to squares (which should be avoided) are counted positively; the fitness **f** of individual **i** is calculated as in equation 2:

$$f_i = \left| \left(\sum_{s=1}^{N} \left| x_i^s - x_s \right| \right) - \left(\sum_{c=1}^{N} \left| x_i^c - x_c \right| \right) \right| \qquad (2)$$

where **s** indicates squares, **c** indicates circles, $\boldsymbol{x}^{\boldsymbol{T}}$ is the x-axis position of a particular shape (of type **T**) at the end of a trial, and $\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{T}}$ is the x-axis position of agent **i** at the end of the same trial. Sets of shapes were generated randomly for each competition, but two prospective parents were always compared against the same set. Evolutionary runs were ended once an agent in the population had achieved 100% accuracy on any shape set.

The best agents in this system achieve good performance after about a 5 million iterations of the microbial algorithm. The material that follows will focus on one agent, which achieved 100% accuracy on one shape set, and retained good performance (>98%) when tested against 100 other randomly generated sets.

4.2 Behavioural Analysis

Figure 11 graphs the absolute lateral distance between the agent's centre and the shape centres in trials of different type; the data illustrate that the agent does in fact catch circles, while successfully avoiding squares. There is also an apparent similarity in the paths during the early stages of both trial types; an active scanning strategy (at least superficially similar to that identified by Beer, 2003) that gives way to genuine divergence only after the ~90th time step. Intuitively, this pattern suggests a perceptual process, which drives a categorisation "choice" that defines subsequent behaviour. The material that follows offers three analyses designed to explain it.

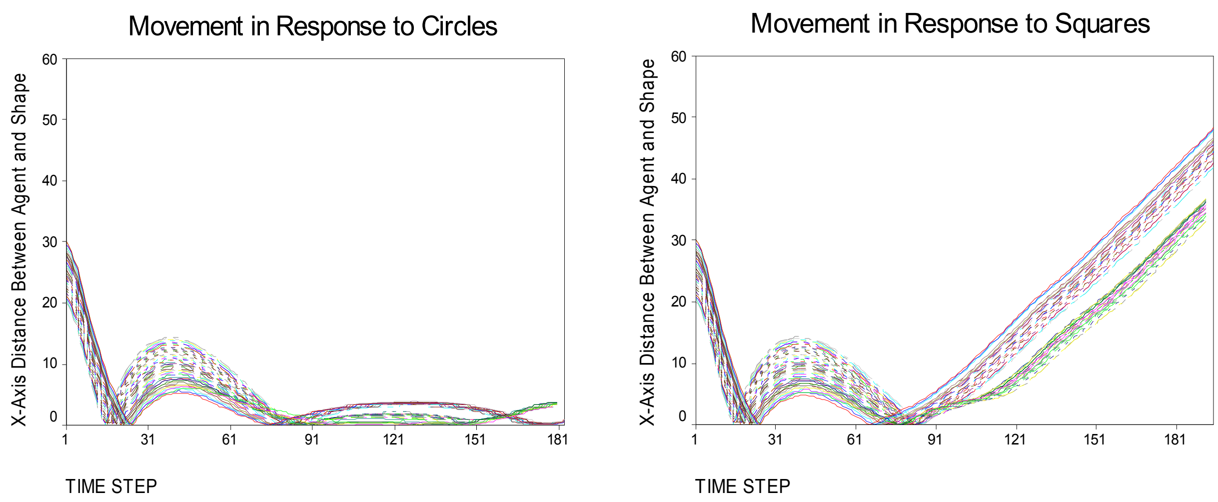


Figure 11: Absolute lateral distances between the agent and shape-centres in a random selection of 100 shapes (squares and circles). Each series refers to a single trial. (Left) Behaviour in response to circles. (Right) Behaviour in response to squares.

4.3 Conventional Network Analyses

At least one – very detailed – analysis of this type of agent already exists (Beer, 2003). Throughout the material that follows, it will be important to remember the restricted scope of this chapter, which is not intended either to repeat, or replace, an analysis of that kind. The current focus rests solely on the issue of representation – on the extent to which classically recognisable representations can be identified in Dynamicist models. From this perspective, the most relevant conclusion of that earlier work is:

> "Whatever "meaning" this interneuronal activity has lies in the way it shapes the agent's ongoing behavior, and in the way it influences the sensitivity of that behavior to subsequent sensory perturbations, *not in coding particular features of the falling objects*". Beer, 2003, p. 238 (my emphasis)

In other words, the analysis uncovers nothing in these agents that appears to "stand for" properties of the agent's environment – nothing that represents in the classically expected manner. This conclusion is all the more interesting because, in the author's own words, categorical perception is a "representation-heavy" task (Beer, 1996): a task that might naturally be expected to require access to cognitive representations. In the sections that follow, we will attempt to discover representations of precisely this sort; internal states that *can* be interpreted as instantiating the agent's knowledge of shape type.

To understand the motivation – and contribution – of the analysis that I will propose, it will be useful to be able to compare its results to those obtained by more conventional means. A complete review of prior art is beyond the scope of this thesis; for the purpose of illustration, I consider just two examples from the field.

4.3.1 Principal Components Analysis (PCA)

One of the most popular tools for neural network analysis, PCA is a technique for expressing high dimensional data sets as lower dimensional data sets, while preserving the data's underlying variance. Neural network state-spaces have at least as many dimensions as they have units – usually far too many to comprehend directly. PCA can reduce that apparent complexity, exposing the fundamental dimensions of a network's state trajectory.

The mathematics underlying PCA are well-described elsewhere (e.g. Gonzalez & Richard, 1992; Oja, 1989; Rao, 1964); I will provide only a brief summary. My version of this method, based on that used by Elman (1991), begins by recording step-by-step hidden unit activities as the agent attempts to categorise falling shapes. These series compose an [N x T] matrix, where N is the number of hidden units (7 in this case), and T is the total number of time steps required to complete the 100 trials[2]. From this "activity matrix", we can calculate a covariance matrix; the dimensions that PCA identifies are eigenvectors of this matrix, and their eigenvalues correspond to the variance that each accounts for.

Three principal components account for 88% of the variance in hidden unit activities; Figure 12 illustrates the way these components change throughout each trial. The features of interest here are *differences* between these components in trials of different type – because a sensitivity to shape type is a precondition (i.e. necessary, though not sufficient) for the representation of shape type.

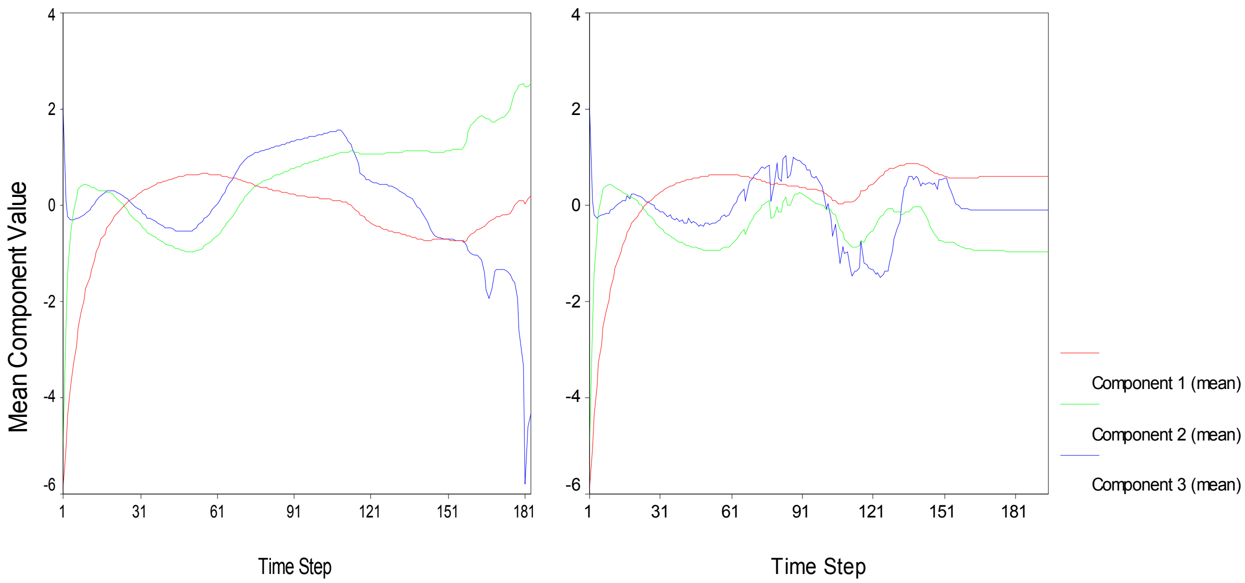


Figure 12: The agent's hidden unit state trajectory, projected onto three Principal Components, during shape categorisation trials involving circles (left) and squares (right).

Moving a bit beyond Elman's method, we can quantify these differences statistically – comparing the values of each component (and of the x-axis distances from Figure 11) at each time

2 This number can vary from trial to trial, because trials can end before a shape touches the x-axis – when they are "caught" by an agent – and because shape size can vary.

step in trials of different type. Remember that the shape set includes 50 squares and 50 circles, and every square is paired with an equivalent circle, identical in every respect but for its type. At each time step, there are therefore 50 values for each component in square-trials, paired with 50 values for each circle-trial. None of the samples deviates significantly from a normal distribution, so we can use t-tests for paired samples to quantify the differences at each time step. Figure 13 displays the T-values (where $p < 0.001$) derived from these tests; these series represent the extent to which each principal component, and also the agent's lateral distance from the shapes, is different depending on the categorisation decision that the agent is required to make.

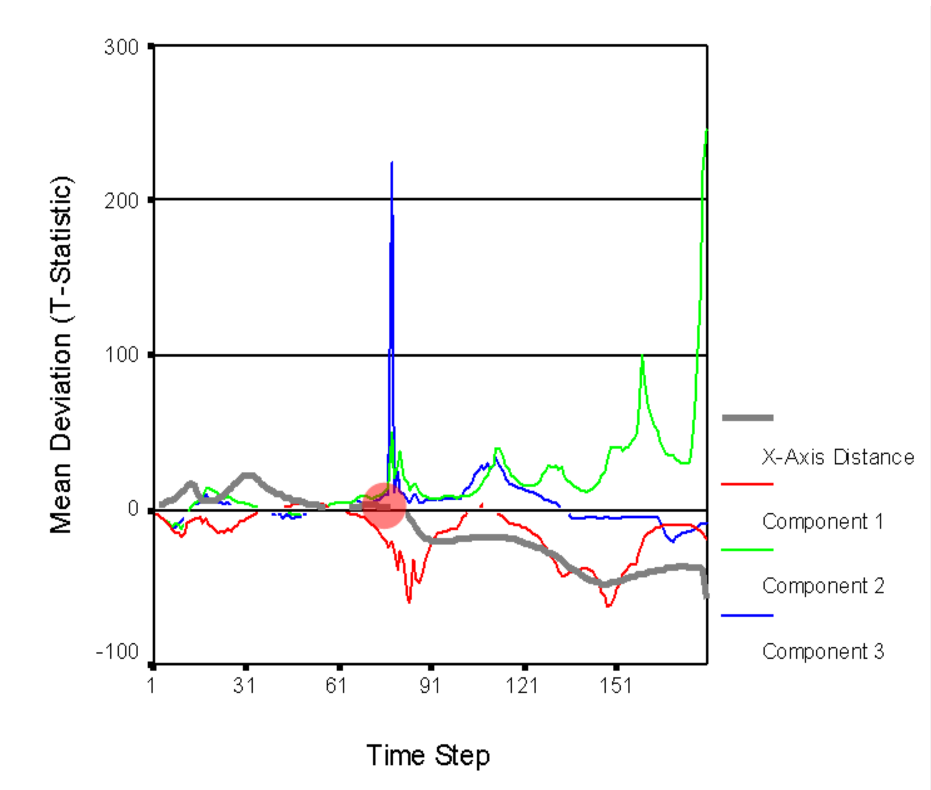


Figure 13: T-statistics ($p < 0.001$) for tests of shape-sensitive deviation. The red circle marks a possible "decision point"; the point beyond which the agent's shape-sensitive behavioural deviation is consistently significant.

The proposed decision point is visible in Figure 13 as the final period of behavioural similarity before *consistent* behavioural deviation is observed. Further, one of the principal components (component 3) displays a very extreme pattern of shape-sensitive deviation during that period; this pattern reflects a stark reduction in the variance of component 3 at that time, which implies significant uniformity across trials involving the same shape type. The temptation is to conclude that this agent does make a decision, and that component 3 conveys its “knowledge” of shape type.

However, that temptation should probably be resisted. Though intuitive, this computational interpretation is also rather circular; we have decided that there is a decision point, then “found” that point in the data and used it to drive our interpretation. That kind of logic can clearly lead observers to see structure that simply is not there – precisely the kind of mistake that we want to avoid. Following conventional logic, we can use linear regression to associate that deviation with deviations in particular components. Employing the x-axis deviations (t-values) as the dependent variable, and the component deviations (t-values) as separate independent variables, we have three regression analyses with series that each contain 50 values; the results emphasise components 1 ($p < 0.001$, $R^2 = 0.33$) and 2 ($p < 0.001$, $R^2 = 0.32$), while marginally dismissing component 3 ($p = 0.054$, $R^2 = 0.02$). However, shape-sensitive behavioural deviation is evident very early in each trial – and certainly before our proposed decision point – so it is far from clear that these associations can justify claims that any of these components convey classically recognisable representations. In other words, this conventional analysis seems perfectly consistent with the claim that representations need play no part at all in the agent's behaviour.

Even ignoring these problems – and accepting our initial intuition despite them – significant obstacles remain. Consider the values in Table 1, which indicate the correlations between hidden unit activities and the extracted components. How can we “reverse” the extraction – for example, of factor 3 – to view the supposed representation in its original form? Six of the seven hidden units are significantly correlated with factor 3; must we inspect all six of these units to observe our representations? If so, the analysis has done little to reduce the agent's apparent complexity. Should we define some minimum correlation below which we can ignore particular units? Though perhaps appropriate in some circumstances (such as the analysis of fMRi data: e.g. Friston, Worsley, Frackowiak, Mazziotta, & Evans, 1994), this approach seems a poor compromise when better options are available.

As we will see, better options *are* available. The next section considers an alternative that addresses a general concern which lurks behind many of the more specific issues raised so far – correlations and covariance offer at best a limited view of their object's underlying causal structure. To begin to garner evidence of this more causal sort, lesion studies are required.

Hidden Unit	Principal Components		
	1	2	3
Unit 1	.699**	.279**	-.041**
Unit 2	.070**	-.313**	-.876**
Unit 3	.984**	-.120**	-.003
Unit 4	.410**	.608**	.401**
Unit 5	.108**	.196**	.966**
Unit 6	.128**	.957**	-.119**
Unit 7	-.015*	-.639**	.195**

**p < 0.001, *p < 0.05

Table 1: Correlations between hidden unit series and the three principal components.

4.3.2 Multi-Perturbation Shapley Value Analysis (MSA)

Just as neurological disorders can illuminate the functional structure of normal brains (e.g. Shallice, 1988), so lesion analyses can clarify the functional architecture of neural networks. There are almost as many specific methods for lesion analysis as there are researchers to use them. In this section, we will focus on one of the method's more systematic variants, called Multi-perturbation Shapley value Analysis (MSA). MSA was originally inspired by the economics of share-dividend calculation (Keinan, Sandbank, Hilgetag, Meilijson, & Ruppin, 2004a), and its results associate each of a network's hidden units with a Contribution Value (CV) or causal significance, relative to some defined measure of behavioural performance. That CV is essentially a Shapley value.

The Shapley value (Shapley, 1953) is a familiar concept in game-theory, and describes an approach for calculating the fair allocation of gains obtained through the cooperation of groups of actors – allowing for the possibility that some actors may make a greater contribution than others. In formal terms, this situation can be described as a *coalitional game*, defined by a pair (**N**, **v**), where **N** = {1, ..., n} is the set of all *players* and **v** is a is a real number associating a worth, or *payoff*, with the game; the goal is to calculate a *payoff profile*, associating each player with a specific proportion of that total payoff. Shapley's approach started by measuring the *marginal importance* of each actor **i** relative to each subgroup of actors (**S**, where $\mathbf{S} \subset \mathbf{N}$) – this is the difference between the payoff for group ($\mathbf{S} \cup \mathbf{i}$) minus the payoff for group **S** alone. Actor **i**'s Shapley value is then simply its average marginal importance for all permutations of **S**. As applied to the analysis of neural networks, this formulation requires access to the performance scores associated with every subgroup of the networks' units. That “full information” approach may be prohibitive for large networks – and the MSA method's authors do offer a reduced, or “predicted” approach to reduce that load (Keinan et al., 2004a, Keinan, Hilgetag, Meilijson, & Ruppin, 2004b) - but the current agent is quite small, so perfectly susceptible to this kind of exhaustive analysis. The agent has 7 hidden units, so there are $2^7 = 128$ subgroups in all (one of those groups includes all of the agent's hidden units), so the analysis requires that we conduct 128 performance tests in all. Each performance test is defined by a “Lesion Configuration”, which specifies the units that will be removed for that test.

Following Keinan and colleagues' own preference (e.g. Keinan et al., 2004a), the current work also employs “Informational Lesions”, rather than the more traditional “Biological Lesions” to implement each Lesion Configuration. Biological lesions are so-called because they mimic the probable impact of neural lesions, effectively removing units either by setting the weights of their outgoing connections to zero (e.g. Joanisse & Seidenberg, 1999), or by adding random noise to their activity values (e.g. Plaut & Shallice, 1993). As the name suggests, informational lesions are merely intended to remove a unit's information, and work by fixing its activity to an average value. There are numerous reasons for this choice (see Aharanov, Segev, Meilijson, & Ruppin, 2003 for a discussion), but the central intuition behind it is that functional analyses can be misleading if their objects – the networks under study – are too far removed from their “natural” state (e.g. Seth, 2008). The different roles of biological and informational lesions can also be illustrated with the simple example of bias units.

Bias units, a common feature of neural networks, have activity values that are always close to '1' regardless of a network's other dynamics. Often explicitly specified, bias units can also (and often do) emerge through learning, or simulated evolution. When applied to bias units, biological lesions can have a profound effect on a network's state trajectory, since the lesion drastically alters a consistent feature of the network's default state. By contrast, informational lesions will have no effect whatsoever when applied to these units. The preference for informational lesions can therefore be interpreted as expressing a position on what constitutes a “good” explanation of network function – bias units are of minimal interest in those explanations.

In practice, the MSA method starts with a baseline performance test (with no lesions), during which the activity values of every hidden unit at every time step are collected; these data define the average values that informational lesions employ, as well as a performance standard (a categorisation accuracy rate) for the analysis that follows. We then repeat the same series of trials (with the same shapes in the same order), while applying the informational lesions defined by each of the Lesion Configurations (one per performance test). Lesion Configurations can be thought of as binary lists with one cell for each of the agent's hidden units; if the value in a unit's cell is '1', a lesion is applied, whereas a '0' indicates that the unit is allowed to vary freely. The results associate performance scores with each Lesion Configuration – and by implication with every subgroup of network units; Figure 14 displays normalised Contribution (Shapley) values for each unit as calculated by Keinan and colleagues' own Matlab implementation of the process.[3]

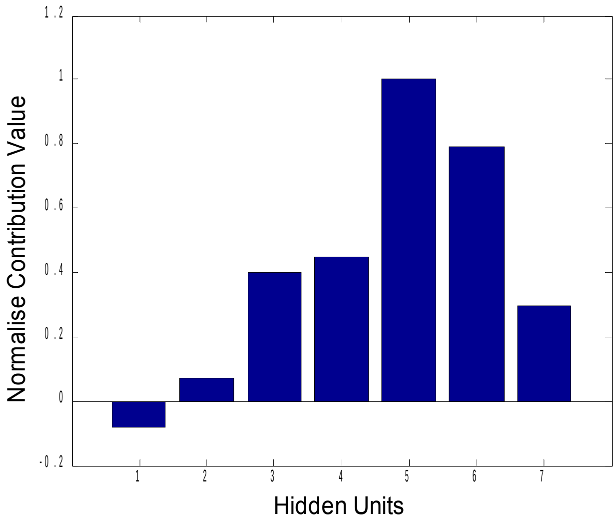


Figure 14: Normalised Contribution Values of each of the agent's 7 hidden units; higher values indicate units that appear to play a more significant role in the agent's classification performance.

Of the seven units, two (units 1 and 2) appear relatively insignificant; we can probably ignore those in our search for representations. Note that, though largely causally insignificant, unit 2 did display a strong correlation with the most intuitive source for representations (principle component 3) that was observed with PCA – a good illustration that correlations and covariance really are an imperfect metric of causal significance. The five other units all do seem to play some role, and two of them (units 5 and 6) may justify particular attention. To interpret these results, we need to inspect the unit activity series themselves; Figure 15 displays the average series for each of those 5 units during trials involving circles and squares respectively.

3 Available on request; see http://www.cns.tau.ac.il/msa/

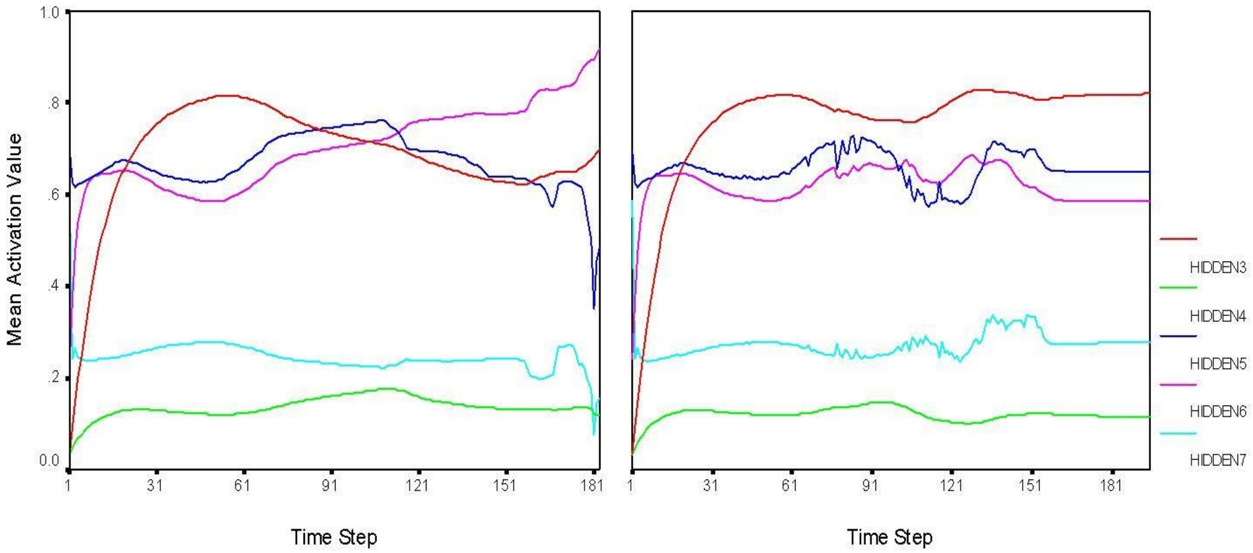


Figure 15: Average activities of 5 units that are causally significant to the agent's categorisation performance. (Left) Trials involving circles. (Right) Trials involving squares.

The visual similarity between Figures 15 and 12 is unsurprising – three of these hidden units are extremely highly correlated with the three factors that we previously extracted. Can we see representations in this Figure 15? There are certainly some sensitivities, some units whose characteristic activity series differ for different shape types. But – as before – it is not clear that we *should* interpret that sensitivity as a representation. Any interpretation that we do make will depend on simple "eye-balling", so will certainly be susceptible to criticism.

This dependence on eye-balling raises a further problem; at least 5 of the agent's 7 units seem to demand some scrutiny. The MSA method's original motivation stemmed from the intuition that, usually, task-specific functional significance will be localised to small subsets of units (Keinan et al., 2004a; Keinan et al., 2004b). If this prediction is satisfied, MSA may be useful – but there is no guarantee that it will be. Larger networks, with more complex behaviour, might yield results that are simply too complex to be useful.

Like PCA, the results of MSA simply do not go far enough to answer the questions with which we are concerned. The shift to causal evidence is an improvement over PCA, but both methods still depend on a subjective eye-balling process, which raises both logical and practical concerns. These problems highlight the need for a method that addresses interpretation directly.

4.4 The Behaviour Manipulation Method (BMM)

Consider a hypothetical agent, designed to solve our categorical perception problem, which *does* represent shape type in the classically accepted manner (i.e. with internal states that stand for either squares or circles). Since the agent is simulated, we have great freedom to constrain its state trajectory; if we know what those representations are, we should be able to force the agent to "perceive" squares as circles, and vice versa. And when subject to that forced perception, the agent should behave as if squares really are circles, and circles really are squares.

The Behaviour Manipulation Method (BMM) is a practical extension of this example's logic – an analysis based on targeted lesions, which shares some features in common with MSA. Like MSA, the BMM employs integer vectors to define the lesions, but the meaning that those vectors convey is rather different. In deference to that difference, these lists are called "Candidates" (rather than Lesion Configurations) in the material that follows. In the language of the hypothetical example, Candidates can be construed as hypotheses concerning the best way to control an agent's perception of its environment; better Candidates permit ever-more effective and predictable mediation of the agent's behaviour.

The current implementation of this method borrows from the concept of the informational lesion, described previously, which involves fixing a unit to its average activity. Extending this concept, we can define a "Partial Informational Lesion", which involves fixing a unit to its average activity *in specific circumstances*. Where the informational lesion is designed to remove a unit's information, the partial informational lesion offers a positive hypothesis concerning the meaning of that unit's activity – that average activity values convey representations of the circumstances that define them. This choice is a useful starting point because it reduces the complexity of unit activity series, and because it accords with the way in which, in practice, researchers manage the apparently random variation in neural spike train data (e.g. Tomko & Crapper, 1974).

4.4.1 The BMM in Practice

Like MSA, the BMM begins with a series of “natural” trials, which provides both a baseline for the agent's categorisation performance (the number of correct categorisations: 100% in this case) and a record of its hidden units' activity values throughout each trial. From these latter data, we can calculate two average activities for each unit – one for trials involving circles and one for trials involving squares (these averages group every time step in each trial type together). As with MSA, we then repeat the same series of categorisation trials (employing the same shapes in the same order) while lesioning the agent's hidden units; each new experiment is defined by a Candidate, which specifies the lesions that should be performed.

Like Lesion Configurations, our Candidates associate each of the agent's hidden units with either a '0' or a '1'. In the former case, the unit is allowed to vary freely, whereas in the latter, a Partial Informational Lesion is applied. To assess the quality of each Candidate, we attempt to use them to reverse the way the agent responds to shape types; in trials involving squares, lesioned units are fixed to their average activities for trials involving circles, whereas in trials involving circles, lesioned units are fixed to their average for trials involving squares. “Good” Candidates will encourage the agent to catch squares and avoid circles. The best Candidates should encourage incorrect categorisations of most, or all, of the shapes. As with MSA, the current version of the BMM implements an exhaustive search of the agent's Candidate-space, testing each of the 2^7 Candidates.

4.4.2 Results

Startlingly, the results of this analysis yield several Candidates that permit very accurate manipulation of the agent's categorisation choices. One Candidate yields a perfect result – a 100% categorisation error rate. At least in this case, the implication is that some of the agent's hidden units really can be interpreted as conveying a classically recognisable "knowledge" of shape type. The results also define a concrete role for particular hidden units, which appear to convey representations through their average activity values. Figure 16 displays the best Candidate that was found: a distributed solution that implicates units 1 and 3 through 7. Each of these units displays deviations in their average activities depending on the shapes involved in particular trials; these deviations can be used to "fool" the agent into confusing squares for circles and vice versa.

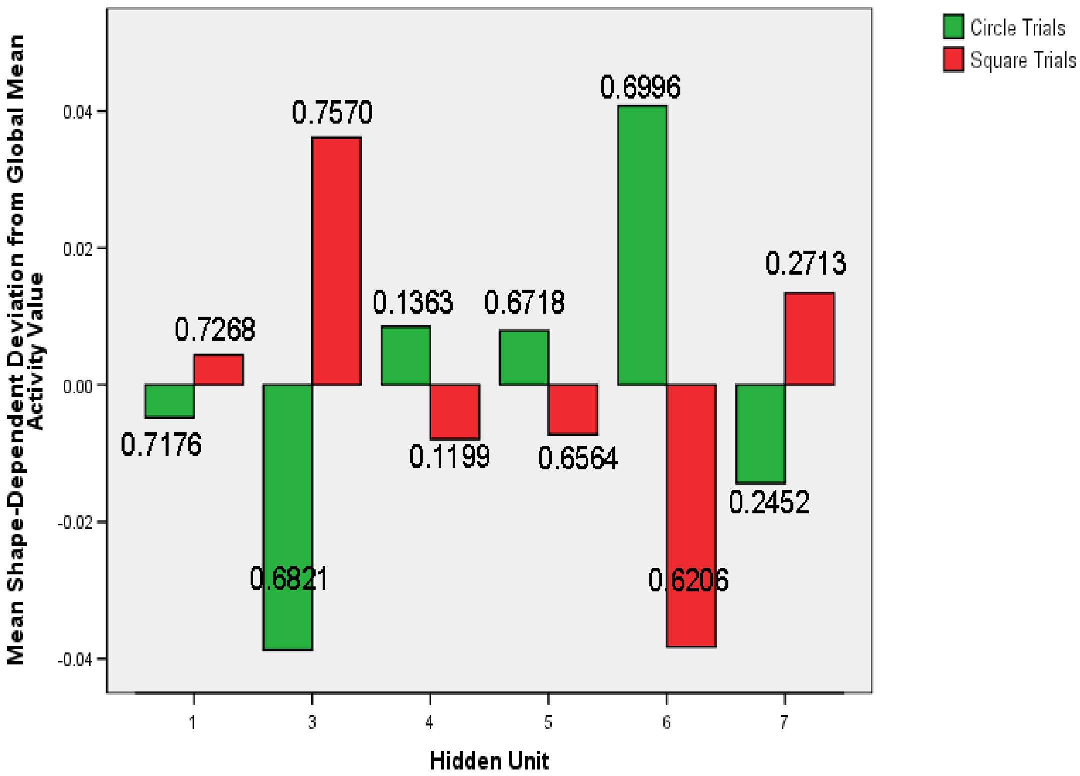


Figure 16: Average shape-type dependent deviation in hidden unit activity values. The bars indicate positive and negative shape-dependent deviations from each unit's average activity across all trials (i.e. including both circles and squares). Unit averages for trials involving particular shape types (circles vs. squares) are appended to each bar. This deviation can be recruited to reverse the way the agent categorises shapes – the implication is that this deviation stands for (or represents) the agent's knowledge of shape type.

4.5 Comparing the BMM to PCA and MSA

Like PCA and MSA, the BMM can be interpreted as a kind of filter, directing attention to those features of a network's dynamics that drive the behaviour of interest. In section 4.3, we saw that there is no guarantee that these "analytical filters" will provide results that are both sufficiently well justified to be useful, and sufficiently simple to be interpretable. Though the BMM also implicates many hidden units, its results are much more interpretable; each unit is associated only with a pair of values (averages) – and we know that these values have played a definite, causal role.

Even on its own, that knowledge is useful. Figures 12, 13 and 15 all graph mean values of the series under study, but the choice is pragmatic (designed to clarify the presentation); nothing in the logic of either PCA or MSA can demonstrate that these mean values are causally significant in themselves. The temptation to make strong claims after eye-balling average series is another good example of the circularity that we are trying to avoid. This conflation is all the more tempting because it is largely accepted in the analysis of, for example, neural spike train data (e.g. Roitman, Brannon, & Platt 2007). Just as minimum correlation thresholds are acceptable in the analysis of fMRi data, so this conflation is acceptable when no better methodological options are available. But as the BMM illustrates, computational models permit far more invasive analyses than their biological referents. Since we *can* verify the causal significance of unit averages directly, it seems reasonable to require that we should.

Another encouraging feature of these results is that, though different in scope, they appear to be largely consistent with those derived from MSA. The best representational theory that the BMM identifies includes units 3-7, precisely those units to which the MSA assigned high Contribution Values. If Partial Informational Lesions are applied only to these units, a 96% categorisation error rate can be achieved; the natural influence of the dynamics of these five units appears to be well captured by their average behaviour in trials of different type. This consistency also clarifies the different contributions that each method makes. MSA allows us to rank hidden units by their causal significance, and that information is absent in the results of the BMM (at least as currently defined). On the other hand, the BMM supplies a justifiable interpretation of the meaning that those units convey – in this case by standing for the agent's knowledge of shape types – while MSA leaves this to the observer. Yet despite this overall consistency, the BMM does seem to disagree with MSA in the way it characterises hidden units 1 and 2.

Unit 1 has a negative Contribution Value, implying that informational lesions actually improved the agent's performance when applied to this unit, but Unit 1 is also part of the best Candidate that we found (displayed in Figure 16). The implication is that the effect of our Partial Informational Lesion is very close to that of the more conventional informational lesion. The shape-dependent averages for all units are numerically quite close, but they are closest for unit 1; in this case, the partial informational lesions' probable primary role is to remove the unit's variance, helping the agent to act on its knowledge of shape type (encoded by units 3-7).

In the case of Unit 2, a positive Contribution Value does not yield a positive role in our best Candidate. The implication here is that unit 2 helps the agent not by encoding its knowledge of shape type, but by helping to guide the shape-following and avoidance behaviour that this knowledge informs. Note that if the agent's control of movement depended mostly on its hidden units, performance would fall apart when a PIL is applied to them – but the Candidate that includes all hidden units displayed fairly accurate behaviour (88% reversed accuracy). The implication is that the control of movement behaviour is largely carried out by the agent's direct sensor-to-effector connectivity. This is not surprising, because given knowledge of the target shape's type, catching / avoiding behaviour can be expressed by a linear mapping from the sensor units. Nevertheless, a freely varying unit 2 is clearly critical to the perfect performance that this agent achieves.

In its current form, the results that the BMM provides also lack a temporal dimension, which both MSA and PCA include. This is quite deliberate; by associating static values (unit averages) with static referents (shape types), we have traded this temporal information for improved, interpretative clarity. The cost of this trade is a dissociation between the agent's categorisation performance and its actual behaviour (i.e. the pattern of lateral distances between the agent and each shape throughout each trial) – the best discovered Candidate allows a rather better manipulation of the former than the latter. In Structuralist terms, the BMM captures the agent's knowledge of shape type, but not its natural decision process – we will return to this criticism in the next chapter.

In this canonically sceptical domain, the precise form of the discovered representations is rather less significant than the fact that we find “good” results *at all*. The system considered here is rather more convincing as a spur for the “mental gymnastics” (Beer, 1996) required to develop good analyses than as a source of convincing cognitive theory. And it has played that former role successfully. Armed with the BMM we can turn our attention to other, more overtly cognitive domains.

Chapter 5

Dynamicism and Number Comparison

Turning back to the original focus of the thesis, this chapter presents a Dynamicist model of number comparison. The model is founded on the common intuition that this capacity emerges from selective pressure to forage effectively (e.g. Gallistel & Gelman, 2000); effective foragers will tend to "go for more" (Uller et al., 2003) food, implying an ability to judge relative quantity. This chapter implements that logic by "evolving" quantity-sensitive foragers.

5.1 Model Design

The environment is a simplified "berry world"; a 2D toroidal grid, composed of 100x100 square cells, where each cell can contain up to 9 berries. Food is initially randomly distributed throughout the environment, with a uniform probability that a given cell will take any of the possible food values (0-9). As food is "eaten", it can be replaced by random "growth" in other cells. Growth rates are adjusted to maintain the total quantity of available food at no less than 80% of its original value.

The ecosystem includes a fixed population of 200 agents, which traverse their environment by moving between adjacent cells. The agents are recurrent, asymmetrically connected, rate-coded neural networks; the activation value ***u*** of the unit ***i*** at time ***t*** is calculated using equation 3:

$$u_i(t) \;=\; \sigma\left(\sum_{j=1}^{N} w_{ji}(u_j(t-1))(1-m)\right)+u_i(t-1)m \qquad (3)$$

where $\boldsymbol{w_{ji}}$ is the weight of the connection from unit ***j*** to unit ***i***, ***σ()*** is the sigmoid function (bounded in the range 0-1) and *m* is a fixed momentum term with a value of 0.5. This momentum term replaces the unit-specific time constants used in chapter 4, and is equivalent to fixing all of those constants to '2'; higher values of *m* give greater weight to each unit's previous activity value (and less weight to its inputs) in the calculation of its current activity value. In this case, sensors and effectors can only mediate each other through hidden units (Figure 17).

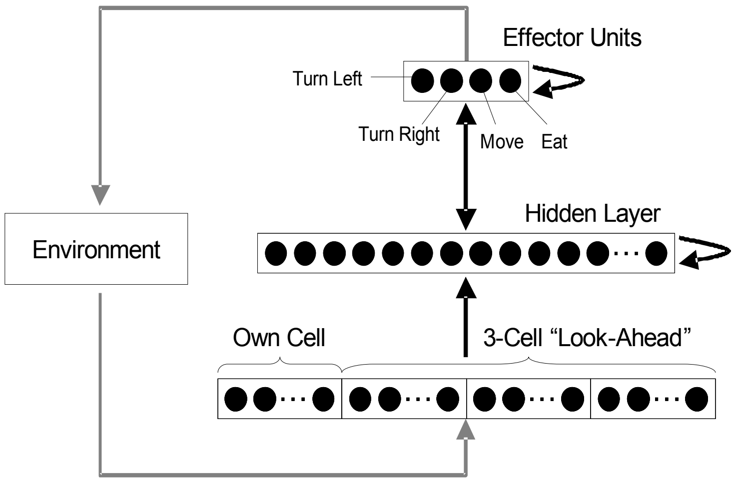


Figure 17: Schematic structure of the quantity comparison agents' neural network architectures. The sensor layer is composed of four cells, each with nine units. The hidden layer is initialised at ten units, but agents in the final population invariably have between 23 and 26 hidden units

The agents’ sensors are always clamped according to the food values of the cells within their “field of view” (see Figure 18) – agents are sensitive to the cell that they currently occupy and to the three cells directly ahead. Each sensor field represents a corresponding food value with a “Random Position Code”; this was used by Verguts and Fias (2004), among others, to capture quantity information without employing any of the popular representational strategies (see Figure 19). By using this code, we are also restricting the problem that agents must solve, assuming away the perceptual cues, such as element size (Miller & Baker, 1968) and density (Durgin, 1995), that mediate numerosity judgements in humans. These simplifications are important, but permissible given the current, methodological focus.

Effector units are interpreted to define the agent's behaviour during each update; agents can turn left or right, move forward, or eat. Each action is associated with a unit, and is executed if its unit's activity is supra-threshold (here, above 0.5). When two inconsistent actions – turning left and turning right, or eating and moving – are attempted at the same time, neither occurs.

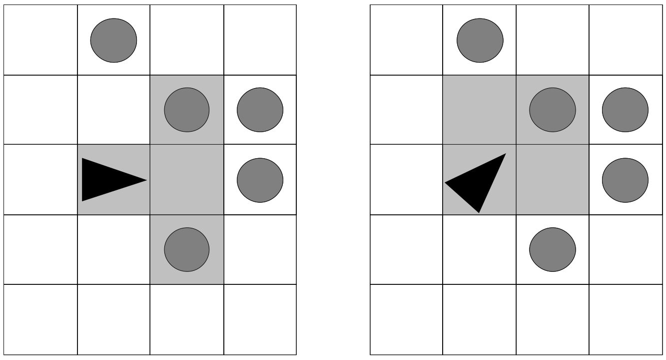


Figure 18: An agent in its environment. (A) The agent – a black triangle – is facing right and can sense food (grey circles) in its right and left-most sensor fields. (B) The same agent, after making a single turn to the left. It can now sense only one cell containing food.

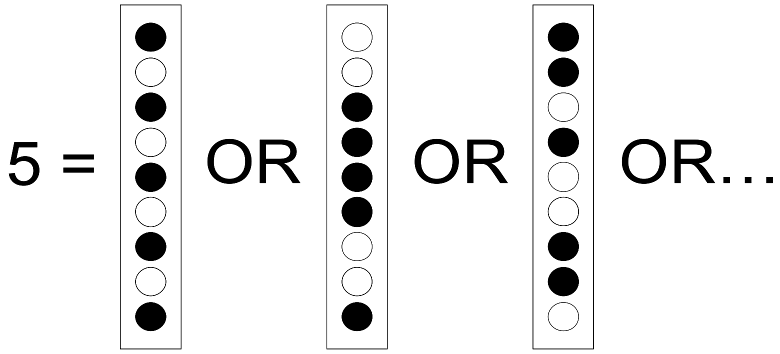


Figure 19: The Random Position Code, similar to that employed by Verguts, Fias, and Stevens (2005). To represent the quantity N, the code requires that exactly N (randomly selected) sensor units be active. The code is illustrated for N = 5.

.

The ecosystem proceeds by iterative update – each update allows every agent the opportunity to sense its environment and act. Agents are updated in a random order, which is re-calculated at the beginning of each time step. As in chapter 4, the "evolutionary" process is implemented with a Microbial genetic algorithm. The crossover and mutation operators are also identical to those used in that chapter, with the exception that the current version includes a dynamic hidden layer that can grow and shrink in size; additions to and subtractions from the hidden layer replace the mutation of time-constants in that model (p = 1%). The fitness of the agents in this system is simply the rate at which they collect food, defined as the amount of food collected since their "birth", calculated by equation 4:

$$Fitness_i = \frac{Food_i}{Age_i} \qquad (4)$$

where the age of individual **i** is the number of time steps since its creation. The goal is to promote the emergence of agents that forage for food in a quantity-sensitive manner – choosing to move into cells that contain the most food by comparing the quantities of food that they can see. The best signal that this behaviour has begun to emerge is high food collection efficiency (food collected per moves made in the environment), which rises toward 5 after about 10 million iterations (~50,000 generations). The evolution was repeated three times, and all three yielded populations that achieved similar distributions of food collection efficiency after a similar number of iterations; the results that this chapter reports are based on the first of those populations.

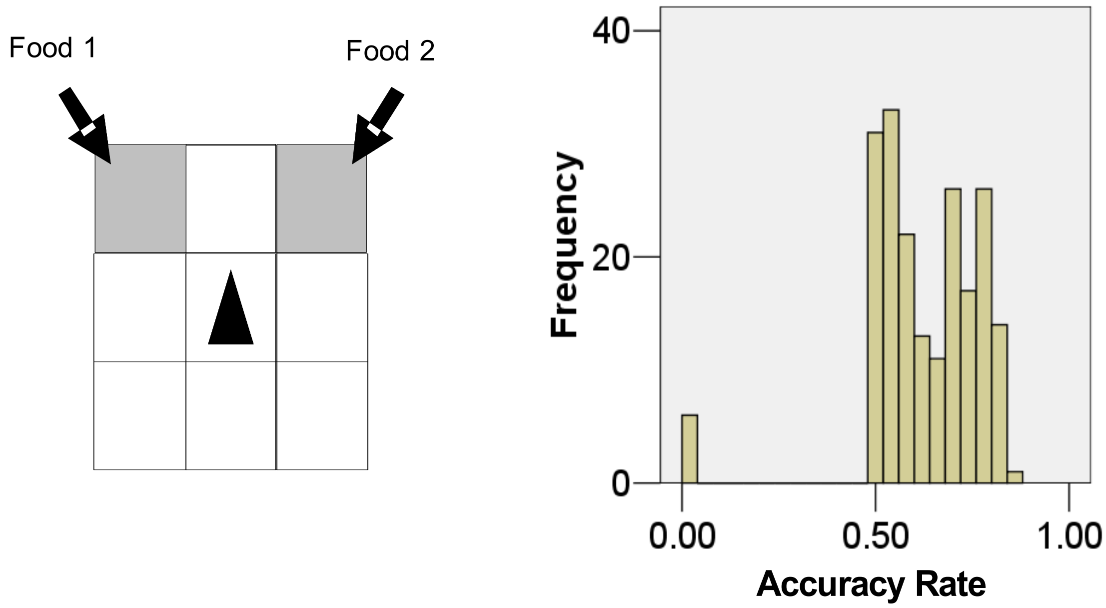
5.2 Behavioural Analysis

To capture the agents' quantity-comparison performance, we can remove them from their "natural" environment and placed them into a 3x3 "mini-world" (Figure 20). Two cells, the top left and top right of the world, contain food of varying quantity. In its initial position at the centre of the world, the agent can "see" both of these food quantities, though it can also turn without constraint once each trial begins. Food selection occurs when the agent moves onto one or other of the filled cells – the only cells onto which it is allowed to move. A correct choice is defined as the selection of the larger of the two food values; this is analogous to method used by Uller and colleagues (2003) to capture quantity comparison performance in salamanders[4]. Every agent in the population was tested

4 One important difference is that Uller et al. (2003) exclude trials in which their salamanders fail to choose one option after a maximum length of time – in our method, these "misses" (failure to choose after 100 iterations) are treated as incorrect choices.

using this methodology, with 50 repetitions of every combination of food quantities (1-9, 72 combinations in all), for a total of 3,600 trials per agent. The results are displayed in Figure 20.

A few of the agents perform extremely badly, indicating that the evolved foraging solutions are brittle in the face of "evolutionary" change. This brittleness may also reflect a more general mutation bias against specialised structures (Watson & Pollack, 2001). The main bulk of the population distribution is also apparently bimodal; agents in the left-most cluster perform at roughly chance levels, whereas agents in the right-most cluster perform significantly above chance – only this latter group appear to discriminate quantity.



.

Figure 20: (Left) The schematic structure of the comparison experiment. The agent (represented by a black triangle) is placed in the centre of the mini-world, facing "up". (Right) A histogram of the population performance in the quantity comparison experiment

The persistence of non-discriminating agents reflects the fact that high rates of food collection can be achieved by sacrificing decision quality in favour of decision speed. A visual inspection of the performance scores for these agents indicates strong asymmetry in their behaviour;

many simply "choose" the right-hand square regardless of the food quantities presented[5]. Using the results displayed in Figure 20, I selected the most accurate agent and recorded its empirical performance in more detail. The results are displayed in Figure 21. As the minimum of the two quantities to be compared increases (Figure 21, left), there is an increase in discrimination error ($p < 0.001$, $R^2 = 0.34$, $\beta = 0.59$); this is an instance of the Size effect. As the numerical distance between the quantities increases (Figure 21, centre), there is a corresponding decrease in discrimination error ($p < 0.001$, $R^2 = 0.55$, $\beta = -0.75$); this is an example of the Distance effect. Strikingly, this agent also displays a Distance effect for *reaction times* ($p < 0.001$, $R^2 = 0.30$, $\beta = -0.56$), just as humans do in analogous tasks. Reaction times are defined as the number of time steps from the start of each comparison trial until the agent chooses one of the two food values (Figure 21, right).

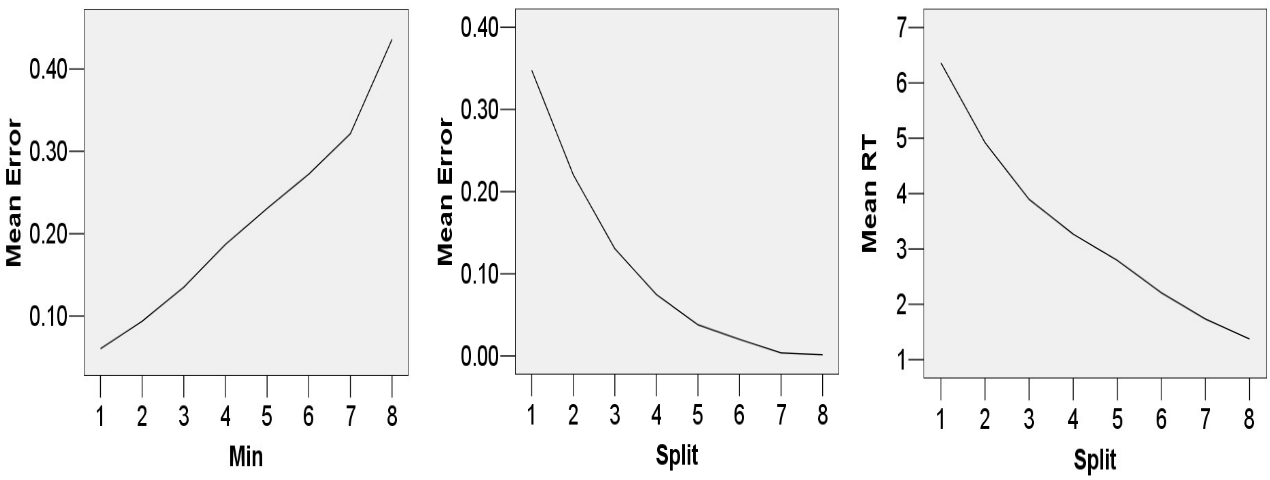


Figure 21: Accuracy scores are rates of correct choices. (a) Mean accuracy vs. minimum quantity of food (Min) in a given trial. (b) Mean accuracy vs. numerical distance (Split) between food quantities. (c) Mean "reaction time" vs. numerical distance between quantities; this latter value is the average number of processing iterations required before the agent makes a defined "choice"

Though non-discriminating foragers can persist by sacrificing decision accuracy for decision speed, this agent is capable of reversing that trade-off, sacrificing decision speed in order to more

5 Though lateral asymmetry is a consistent feature of the behaviour of agents evolved in this system, its direction is not consistent – some runs yield agents that prefer left-sided food.

reliably "go for more". Since Size and Distance effects drive the classical debate on the structure of quantity representation (i.e. the format debate), a representational account of this agent's behaviour – which seems to display those effects – should be able to make a relevant contribution.

5.3 Extending the BMM

Though the logic of the last chapter (the initial application of the BMM) is equally applicable here, the current agent raises some practical issues that demand some extensions. The best quantity-discriminator in our evolved population has 25 hidden units – much more than the 7 considered before. In the previous case, the results were derived from an exhaustive search of the agent's Candidate-space, with 2^7 lesion experiments in all. For much larger spaces of the sort we now face, this approach will be prohibitively time-consuming. The space is further enlarged because the number comparison problem is rather richer – at least in terms of the potential for different representational strategies – than the categorical perception problem. Specifically, there are now multiple "meanings" that we might attribute to each hidden unit, which could represent either of the two quantities independently, or the difference between them. Table 2 displays the list of lesion types – or proposed unit "tuning functions" – that are considered. There are five values in all, so the corresponding Candidate-space contains 5^{25} items.

Lesion Identifier	
0	No lesion
1	Unit average codes for right-hand food value
2	Unit average codes for left-hand food value
3	Unit average codes for relative difference
4	Unit average codes for absolute difference

Table 2: Receptive fields considered in the BMM-driven analysis of the quantity-comparison agent.

To search this space, we can use precisely the same approach as that employed to design the agents themselves – a Microbial genetic algorithm (Harvey, 2001). When designing the agents, the search optimised a population of neural networks, while in this case, we employed the search to optimise a population of Candidates. These Candidates are structurally identical to those employed in the last section, but different in that their cells can contain integers in the range 0-4 (rather than 0-1). The other important difference is that, where the agent was evolved to be an effective forager, the Candidates are evolved to manipulate that agent effectively.

To achieve this goal, we must first record the agent's comparison performance scores for every individual combination of food values (72 in all); the test includes 10 repetitions of each combination, and recorded the number of times that correct choices were made in each case. The result is a vector of performance scores (length = 72), associating each food combination with a score in the range 0-10. A similar list was also generated during the testing of each Candidate; in these tests, agents were always placed in an empty mini-world, and the goal was to discover Candidates that encouraged the agent to behave as if it could "see" particular food value combinations.

Following the logic of chapter 4, we can measure the correspondence between this invoked perception and natural perception by comparing the agent's behaviour in each case; good Candidates should encourage choices that correspond to those made under natural conditions. The fitness of each Candidate is defined as the sum of the absolute item-by-item differences between the baseline scores and lesioned scores; the goal of the search was to find a Lesion List that minimised this "fitness". The calculation of fitness is described below in equation 5:

$$F_i = \left(\sum_{j=1}^{N} \left| (P_j^u - P_j^l) \right| \right) \qquad (5)$$

where P_j^u is the performance score (the number of times a correct choice was made) achieved by the agent for food value combination **j**, P_j^l is the performance score achieved when partial informational lesions are used to simulate the presence of food value combination **j**, but no food is actually present, and **N** is the number of performance scores in each list (72).

After running the lesion-search (~4 million iterations) and identifying the best discovered solution, one further step was required. As mentioned in chapter 4, simulated evolution can yield bias units – units whose activity remains very close to '1' regardless of any environmental input. On closer inspection, five of the agent's units appeared to behave in this way, and two were part of the best Candidate that was discovered. Since bias units do not vary, neither informational nor partial informational lesions should have any impact at all on the agent's behaviour; Candidate cells that correspond to bias units will therefore operate much like “junk” DNA in the genome, since particular values in those cells should have no effect on the solution's overall fitness. That proposal was confirmed by pruning and then re-testing the solution – since no fitness costs were incurred, these units play no part in the results that follow.

Unlike in the previous chapter, the best discovered Candidate does not permit perfect manipulation of that agent's choice behaviour. Nevertheless, the results are encouraging; to assess their quality, we can employ the standard method of linear regression. The mark of a good Candidate is its ability to reproduce “natural” comparison choices when no food is actually present – the dependent variable for the regression is therefore the series of “natural” performance scores that we derived earlier. This series is an integer vector, with 72 cells (one for each food value combination), each containing integers in the range 0-10; a score of '10' indicates that the agent always chooses the larger of the two values when faced with that particular combination. The independent variable is the list of performance scores achieved when our best discovered Candidate is used to lesion the agent, and no food is actually present. In this case, a correct choice is made when the agent moves onto the square that would have been correct if the food that we have tried to simulate were actually present in the mini-world. By measuring the correspondence between these two series, we are measuring the extent to which the best, discovered Candidate has allowed us to manipulate the agent's categorisation choices.

5.4 Results

By linear regression, the relationship between the agent's baseline performance and that obtained using the best Candidate, is very strong: $p < 0.001$, $R^2 = 0.59$. In other words, the best discovered Candidate yields performance scores that are significantly related to the baseline scores, and which account for 59% of the variation in those scores (Figure 22). The theory itself – the best account that we have found of the agent's representational strategy – is graphed in Figure 23.

Is 59% enough? Issues of this sort will always depend on debate. One argument in the result's favour stems from the logic of Spieler and Balota (1997), who argued that a model's item-level predictive power should be judged relative to that of the environmental features that drive the behaviour of interest. In this case, the relevant features are the food values themselves, their mean and numerical distance. When these features are regressed (as independent variables) against the agent's performance scores (the dependent variable), an R^2 value of 0.74 is achieved ($p < 0.001$). That figure of '0.74' is the real target for the Candidates, which are designed solely to capture the agent's representations of those quantities. The best Candidate captures ~80% of the influence of the food values themselves on the agent's choice behaviour, so cannot be lightly dismissed.

The precise form of the agent's proposed representations is also clear; the average activities of almost all of the network's critical units appear to accumulate – positively or negatively, and proportionately – as the magnitudes of their referents increase. Since this linear accumulation is evident at the level of single units in the distributed code, I refer to it as a *single unit accumulator code* in the material that follows. Neurophysiological studies have just begun to tackle the issue of the neuronal correlates of number representations using single cell recording in behaving monkeys. Nieder and colleagues (Nieder, Freedman, & Miller, 2002) have described “number neurons” in the monkey brain with tuning functions that fit the logarithmic coding of Dehaene and Changeux’s (1993) “numerosity detectors”. This finding would seem to be at odds with the type of coding employed by our agent.

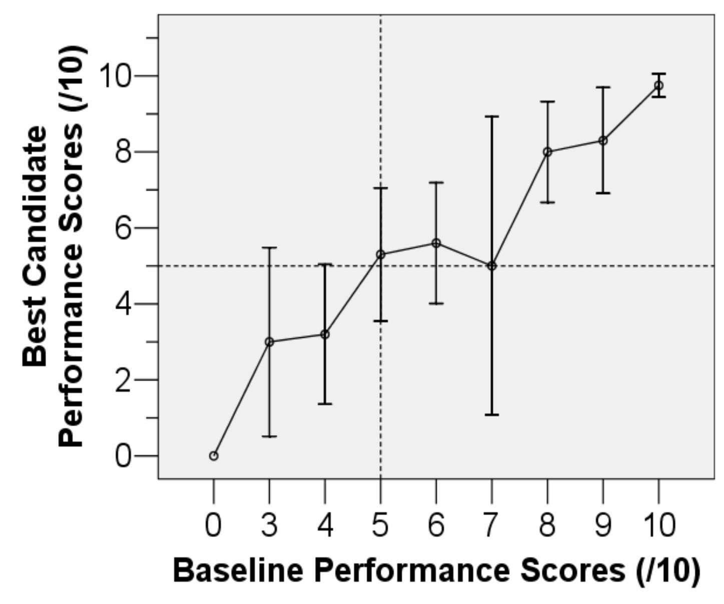


Figure 22: Agreement between performance scores in unlesioned vs. lesioned conditions for the best, identified theory of the agent's representations; error bars are standard errors of the values (mean averages) at each point. Perfect agreement would yield a perfectly straight diagonal line, of the form 'y = x'.

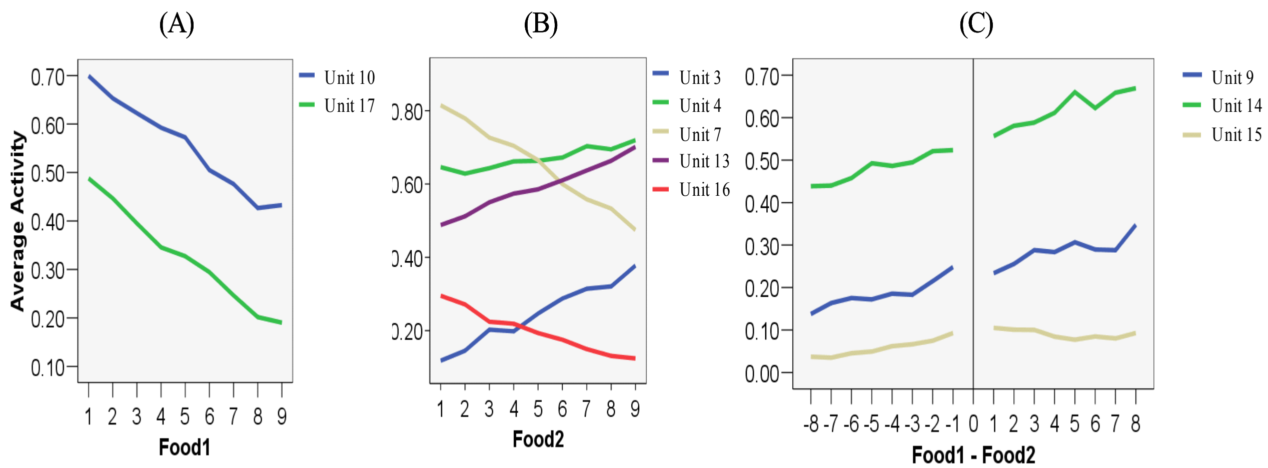


Figure 23: Classically recognisable representations, emerging from in a Dynamicist model of quantity comparison. (A) Representation of food on the agent's right, (B) Representation of food on the agent's left, (C) Representation of the difference between presented food values. Each point in each of these series corresponds to the average activity value of the specified unit in the specified circumstances.

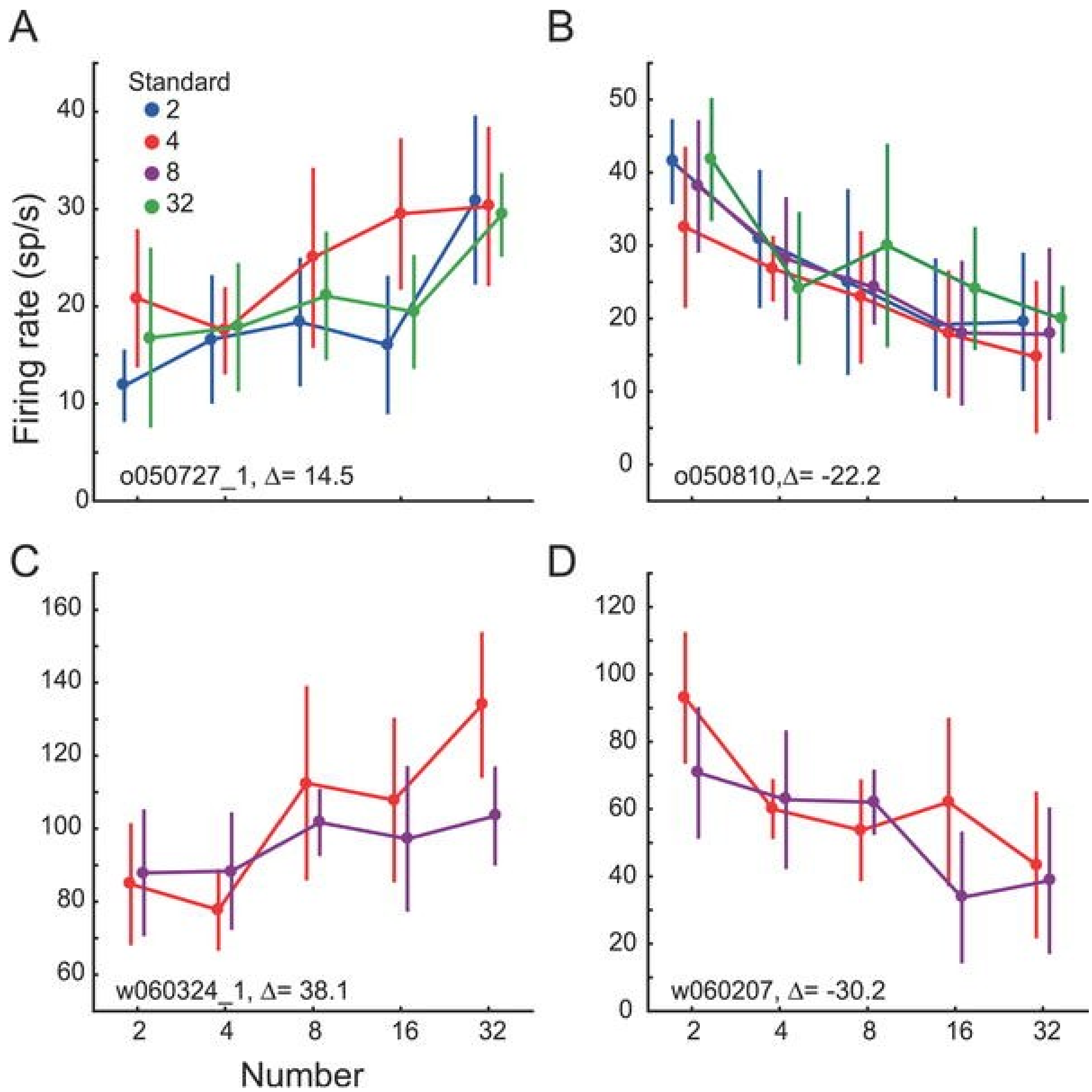


Figure 24: From Roitman, Brannon, & Platt, 2007; 4 examples of neurons' responses recorded in the macaque LIP during a task in which the numerosities of visually presented sets were compared to a fixed standard. In each case, the neurons' spike rates are monotonically related to the number of elements in those sets.

However, a different type of "number neurons" with tuning properties that are startlingly similar to those employed by our agent has been recently discovered by Roitman, Brannon and Platt (2007) in the lateral intraparietal cortex of monkeys engaged in a numerosity comparison task. After averaging the spike rates recorded over a few hundred milliseconds from single, number-sensitive neurons – a process analogous to the current use of circumstance-dependent unit averages – the authors showed that these neurons encode the total number of elements within their receptive fields in a graded fashion (see Figure 24). I was not aware of this work while implementing the model that this chapter reports – nevertheless, these data provide a huge boost to the confidence that we can attach to it. Moreover, the same neural coding strategy (a graded sensitivity to an increase of a particular feature dimension) has been shown to apply to other sensory domains (e.g., the frequency of vibrotactile stimulation; Romo & Salinas, 2003); that result further underlines this representational strategy's biological plausibility, and also raises the possibility that its scope might extend beyond numerical cognition.

This foraging agent is the first example of a quantity-comparison model that encodes numbers with linear single-unit accumulators, though the format is broadly consistent with the accumulator system proposed by Meck and Church (1983), as well as with the coding of Dehaene and Changeux's (1993) "summation clusters" (which precede numerosity detectors) and the *numerosity code* proposed by Zorzi and colleagues (Zorzi & Butterworth, 1999; Zorzi et al., under revision). I refer to this novel format as the *single unit accumulator code* in the material that follows.

5.5 Interim Discussion

We began with some criticisms of Connectionist learning (chapter 3), which motivated the original introduction of Dynamicism. Like Connectionism, Dynamicism carries an apparent, intuitive cost that balances its demonstrable practical benefits; the Behaviour Manipulation Method is inspired by the recognition that dynamical approaches to cognitive science must find a place for classical structure before most cognitive scientists will accept them. Nothing in the logic of Dynamicism forbids Compatibilist interpretation, but in practice, Dynamicist models have not been thought to support it. The first goal of this work – addressed in chapter 4 – was to demonstrate that (and how) Compatibilist interpretations can be made in even the most canonically skeptical circumstances.

Two conventional analytical methods – PCA and MSA – fall short of achieving that goal. Each of these methods can act as a filter, guiding the focus of our attention, but there is no guarantee that either will filter the data into a neatly interpretable form. And even more neatly interpretable cases will still depend on a subjective “eye-balling” of the results, offering little in the way of formal justification for the classically-minded observer. By contrast, the BMM offers a formal, scalable, statistically justifiable route toward the identification of causally significant, classically recognisable representations. It should be noted that our criticisms of PCA also apply to Multi-Dimensional Scaling (MDS) methods, which have gained some currency in recent years (e.g. Botvinick & Plaut, 2004).

The second goal was to illustrate that, armed with the BMM, Dynamicist researchers can begin to “play the same game” as their more conventional counterparts. Using the structure of behaviour-based selection as an analogy for natural evolution, this chapter applied the BMM to “evolved” quantity-sensitive foraging agents, which display characteristic Size and Distance effects when forced to compare two quantities. This environment demanded two important extensions to the BMM – a shift toward the use of search (and away from exhaustive testing), and the definition of a statistical metric for the (probably imperfect) quality of the BMM's results.

The first extension was motivated by the size (25 hidden units) of the agent that we considered, which makes exhaustive search of the Candidate-space impractical. This is a useful extension because it makes the BMM more scalable, but the current form of that extension is largely pragmatic and does carry a cost; we can never be sure that the best identified theory is also the best available theory. Different approaches to searching an agent's representation space may provide better justified results.

Critically though, there is no way to guarantee that the BMM's results will be optimal in a formal sense; their scope will always be restricted by the “representational primitives” that we choose to consider. This point highlights an important opportunity for further extending the BMM;

average unit activities are just one among many primitives that we might reasonably employ. Given the climate of skepticism that Dynamicist models must face, our choice made a justifiable trade of explanatory power for interpretative clarity – but nothing in the BMM's logic precludes the consideration of different primitives, such as time-period dependent means, average rates of change, or even centroid time series. These extensions are attractive because they add a temporal dimension to the results that the BMM might yield – a critical first step on the path to capturing not just an agent's knowledge, but its decision process as a whole.

Yet despite these restrictions, the results are encouraging; the BMM achieved perfect manipulation of an agent's categorical perception performance, and reasonable manipulation of an agent's quantity comparison performance. Even accepting the logic that inspired this approach, the former result is surprising; cognitive theories rarely aim to capture every detail of the performance under study. Like the first extension to the BMM, the second – defining a statistical metric for the "quality" of its results – will therefore probably be key to its scalable application. The form of this extension is useful both for its application-independence, and because it lets us compare the effect of an agent's putative representations to the effect of the referents themselves.

Though novel in detail, the best, discovered theory of the agent's representations is also broadly consistent with some other theories that postulate a linear relationship between (external) numerosity and activation of the (internal) quantity code (e.g., Gallistel & Gelman, 1992, 2005; Meck & Church 1983; Zorzi & Butterworth, 1999), and not with others that represent numerical quantity as a position on a logarithmic analogue continuum (Dehaene, 2003). In other words, the agent does at least appear to "play the same game" as its more conventional counterparts, achieving our second goal. But playing the game is just a first step – Dynamicists should also strive to win it. With this goal in mind, the most valuable source of supporting evidence is the single-cell recording work reported by Roitman, Brannon and Platt (2007), which seems to provide a clear confirmation that the foraging agent's MNL format might also occur in biological agents.

On the other hand, one possible remaining concern is that our foraging agents are really no more "minimal" than their modular counterparts – that the precise form of our results owes more to the details of the artificial ecosystem than it does to a more general connection with the pressure to forage effectively. This kind of connection is probably unavoidable – indeed, its biological relevance is also assumed in the way that researchers employ the statistics of real sensory stimuli to decode the tuning functions of biological neurons (e.g., Atick, 1992; Barlow, 2001; Simoncelli, 2003) – but is presence does suggest a direction for future research. Specifically, these results could usefully be confirmed by reproducing the same selective pressure in a different ecosystem (e.g. with different movement dynamics, sensor representations and / or food types, but with the same essential selective pressure).

More broadly, this Dynamicist approach offers three advantages over its more conventional rivals – the promise of effective cognitive-behavioural integration, the problem-driven emergence of both empirical phenomena (e.g. Size and Distance effects) and representational strategies at the same time, and the opportunity to explore the role of phylogeny in cognitive development. These advantages could be achieved in different ways, but in practice, classically modular modeling approaches do not encourage them.

Alongside those general advantages, Dynamicism also appears to carry an important, general cost; behaviour-based selection is a reasonable model of evolution, but a very poor model of learning. Criticisms of conventional Connectionist learning (such as the reliance on implausible architectures: e.g. O'Reilly, 1998) may be justifiable, but without an alternative account of the role of synaptic plasticity, Dynamicism will struggle to effectively capture a great deal of cognitive behaviour. Biologically implausible design methods can yield biologically plausible structures – but the acquisition of cognitive skills is often at least as important a focus of interest as the structures underlying "mature" skills. Attempts to integrate synaptic plasticity within a Dynamicist framework are beginning to be made (e.g. Phattanasri, Chiel, & Beer, submitted); results of this sort will be critical to the future of the Dynamicist project.

Dynamicist are right to require that empirical evidence should drive the role that Structuralist concepts of representation play in cognitive theory. But in itself, this argument is incomplete; it does not tell us how to *satisfy* that test – to claim with confidence that representations really do emerge. As well as making a novel contribution to the format debate, this section of the thesis has proposed a logic which addresses that problem directly; theories involving classical representations are useful, and approximately correct, if they allow us to manipulate behaviour in predictable ways. The result is a Compatibilist compromise between Structuralist intuition, and Eliminativist doubt. Armed with this method, Dynamicism can move a step further along the path that Connectionism has taken, from peripheral, contentious novelty to accepted, fundamental methodology.

Chapter 6

Evolving Optimal Decision Processes

In the last chapter, I mentioned that the current variant of the BMM captures representations, but not the decision process that employs them. This chapter introduces a distinct but related framework which addresses the latter directly – inspired by the assumption that (biological) neural information processing strategies will be close to optimal for any given task.

6.1 Normative Formal Theories and Cognition

Sometimes, the best way to understand a process is by comparing it to an independent standard. Normative analyses can provide that standard, describing optimal or near-optimal strategies for solving problems of cognitive interest. Studies of vision, in particular, have benefited from this perspective, decoding the relevant neural systems with ideal observer theories (e.g. Geisler, 1989, 2003; Najemnik & Geisler, 2005). More recently, the approach has been fruitfully applied to perceptual decisions, where attention has focussed on three dimensions of variation; the strength of sensory "evidence" that subjects are given (e.g. Kim & Shadlen, 1999; Gold & Shadlen, 2000, 2001, 2003), the probabilistic distribution of correct choices in the past (e.g. Ciaramitaro & Glimcher, 2001; Platt & Glimcher, 1998), and the reward associated with alternative choices (e.g. Platt & Glimcher, 1998). Responses to variation in all three of these dimensions are susceptible to normative analyses.

One task in particular provides an elegant example of the mutual interaction between normative analyses and empirical data. The simplest variant of this task engages subjects in a series of two-alternative forced-choice motion detection problems, driven by fields of moving dots. In each case, a specific proportion of the dots move together, in the same direction (e.g. left or right); the subjects' goal is to identify that direction, and respond accordingly. The most popular, normative analysis of this task employs a Bayesian formalism – a directed, bounded, stochastic random walk – that describes the incremental accumulation of noisy evidence in favour of one or another choice (Shadlen, Hanks, Churchland, Kiani, & Yang, 2006). This model is a special case of the much more general Diffusion-to-bound framework (Ratcliff, 2001; Shadlen et al., 2006), which was first employed to describe the behaviour of gases (i.e. Brownian motion) but is now much more familiar for its applications in finance – in particular to the pricing of options and other derivatives. Diffusion models have also been widely applied in cognitive science (Ratcliff 1978, 2001; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998; Wagenmakers, van der Maas, & Grasman, 2007). The framework's general aim is to describe the behaviour of dynamical systems that are mediated both by a sequence of inputs and by random noise; in concert with the systems' current state, those two factors define their state in the next instant. Given certain other parameters, like the state's possible range of change from one instant to the next (its volatility), the framework can define probability distributions for these systems' behaviour in time.

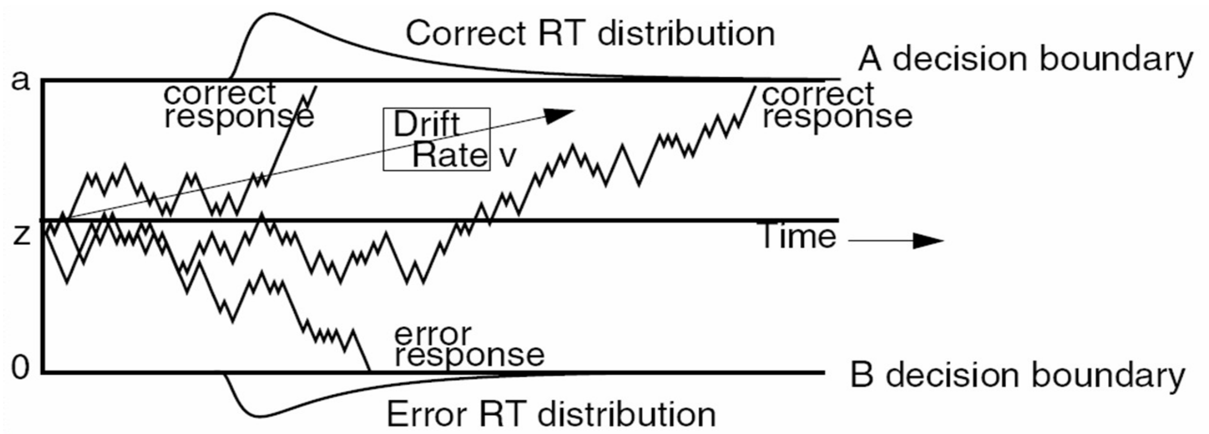


Figure 25: Schematic structure of a Diffusion-to-bound system, with two decision boundaries. Weaker signals yield slower drift away from the system's initial state (Z). Weaker signals also make noise more significant in that drift, so raise the probability that the wrong decision boundary might be reached. This figure is re-printed from Ratcliff & McKoon, 2008.

The simplest variant of the system is illustrated in Figure 25 – a single variable that can change in only one dimension. Imagine that this variable is subject to random noise, and receives a simple, binary signal (a sequence of 1's and 0's); when the system receives a '1', its state is incremented, while a '0' implies a decrement of the same magnitude. Given a completely balanced sequence (e.g. alternating 1's and 0's), this system's behaviour will be largely determined by its noise, but sequences that contain mostly 1's (or mostly 0's) should push the system into definite positive (or negative) pattern of incremental accumulation. As the sequence's bias in favour of one or another value increases, so the rate of that accumulation should increase. If we place boundaries on the accumulation, at equal distances above and below the system's initial state, those different rates of accumulation will translate to different “reaction times”; the system will reach one (or other) decision boundary more quickly when the signal that it receives is stronger. And when the signal is weak, noise-driven drift may allow the system to reach the “wrong” boundary, or to make an incorrect response. Given the right parameters, this framework can therefore capture both error rates and reaction times, relating both to the coherence of a sequential input signal..

With very few free parameters, this model can provide a very good fit to subjects' behavioural data (reaction times and error rates) in the motion-detection task (e.g. Shadlen et al., 2006). However, perhaps its most interesting predictions are pitched at the level of the neurons that drive that behaviour. To collect those neural responses, Shadlen and colleagues recorded data from single neurons in the monkey brain, while these monkeys performed the motion discrimination task. Their particular variant of the task employed eye movements as responses, and the authors recorded data from the primate homologue of the pre-frontal and lateral intra-parietal cortices (PFC and LIP respectively), previously associated with the preparation of eye movements toward particular parts of the visual field (e.g. PFC: Wilson, Scalaidhe, & Goldman-Rakic, 1993; Funahashi, Bruce, & Goldman-Rakic 1993; LIP: Gnadt & Mays, 1995; Colby, Duhamenl, & Goldberg, 1996). What they found was striking – neurons that seem to implement the normative model directly, with spike rate as the accumulating variable (see Figure 26). And like the random walk model, the rate of that accumulation reflects the strength of the available sensory evidence.

This mechanism offers a potentially very general insight into the way that neural systems categorise incremental evidence from noisy sources. Three of the four most popular accounts of number representation – all except Zorzi and colleagues' *numerosity code* – include some concept of noise. Remembering that the BMM, described in chapters 3-5, tracks average unit activity values, noise is also implied in the *single-unit accumulator code* proposed previously. Indeed, as Figure 26 shows, this kind of accumulation is consistent with a monotonic relationship between these neurons' average spike rates and signal strength – precisely the relationship that the *single unit accumulator code* defines. In other words, there are good reasons to suspect that this same accumulation process might be employed in at least some numerical processing tasks (Dehaene, 2007).

Several computational models of this motion discrimination task have already been proposed. One comparatively early example, by Gold and Shadlen (2000), demonstrated that populations of spiking neurons could be effectively pooled to drive the proper accumulation. Another, more simplified model (Usher & McClelland, 2001) explored the possibility of extending the mechanism to capture N-alternative (rather than 2-alternative) forced-choice tasks. More recently, Wong and Wang (2006) analysed an extremely minimal version of the mechanism in detail, and proposed a mechanism that might mediate the accumulation's decision boundary. In the terminology of chapter 2, all of these examples are performance models – hand-coded to capture the the neural responses that have been observed. This chapter takes a rather different approach, which can begin to predict – rather than just reflect – those responses.

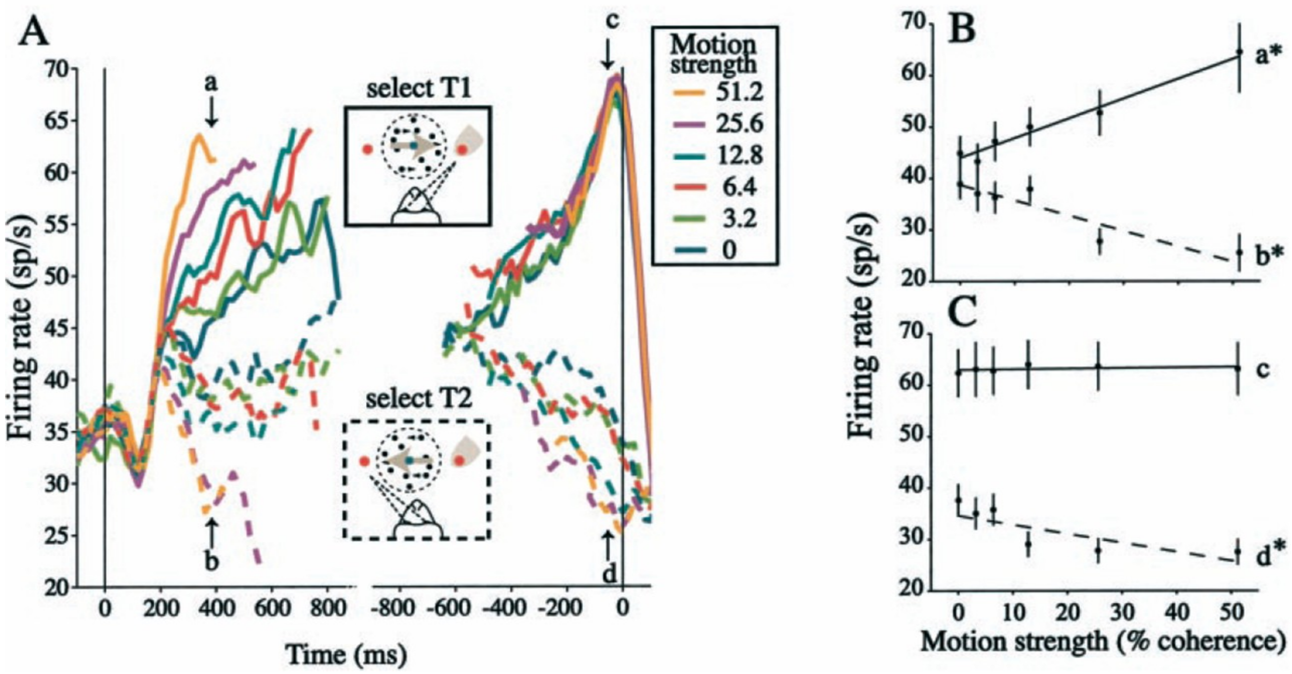


Figure 26: Lateral Intra-Parietal responses to motion stimuli, by signal strength; re-printed from Roitman & Shadlen, 2002. (A) Average response from 54 LIP neurons, grouped by motion strength and choice as indicated by colour and line type. On the left, responses are aligned to the onset of stimulus motion. Response averages in this portion of the graph are drawn to the median RT for each motion strength and exclude any activity within 100 ms of eye movement initiation. On the right, responses are aligned to initiation of the eye movement response. Response averages in this portion of the graph show the build-up and decline in activity at the end of the decision process, excluding any activity within 200 ms of motion onset. The average firing rate was also smoothed using a 60 ms running mean. Arrows indicate the 40ms epochs used to compare spike rate as a function of motion strength in the next panels. (B) Effect of motion strength on the firing rates of the same 54 neurons in the epochs corresponding to arrows a and b above. When motion was toward the RF (solid line; epoch a), the spike rate increased linearly as a function of motion strength. When motion was away from the RF (dashed line; epoch b), the spike rate decreased as a function of motion strength. (C) Effect of motion strength on firing rate at the end of the decision process. Response averages were obtained from 54 neurons in the 40 ms epochs corresponding to arrows c and d. The large response preceding eye movements to the RF (solid line, filled circles; arrow c) did not depend on the strength of motion. Responses preceding eye movements away from the RF were more attenuated with stronger motion stimuli (dashed line; arrow d).

The key to that reversal is a "minimal" model-building method, in the sense described by

Nowak (2004) – a method that, as far as possible, can minimise the architectural assumptions that model-designers must usually make. Rather than using a model to capture fixed intuitions about the neural implementation of this task, this chapter asks – and attempts to answer – a simple question; what neural architecture might be needed to express its optimal (or near optimal) implementation? To answer that question, I propose a variant of the Dynamicist approach described previously: a method that searches the problem's strategy-space to discover effective model architectures.

6.2 Method

The method starts with the definition of the task. Following the logic of prior work (Gold & Shadlen, 2000; Usher & McClelland, 2001; Wong & Wang, 2006), the visual stimuli (fields of moving dots) are expressed by the responses they are thought to invoke in populations of MT movement-sensitive neurons. These responses are encoded as two series of values drawn from two Poisson distributions; coherent motion in a particular direction implies an elevated mean value for the corresponding distribution. When no stimuli are present, the mean value that defines these distributions is '15'. When stimuli are presented, the mean values for the distribution that corresponds to the actual direction of motion are drawn from the range '80-100', while the mean value for the other distribution is always set to '80'. All values drawn from these representations are then divided by '100' before being passed to the network. Taking these series as inputs, our models must “decide” which series has the higher mean value – categorising the implied direction of coherent movement. Note that, as currently defined, these sensor representations are loosely analogous to the *noisy MNL*; like the latter, the former imply that the variance associated with larger numbers is larger than that associated with smaller numbers. Cast in that light, the motion discrimination problem is itself analogous to number comparison against a fixed numerical standard.

The logic of the approach should be applicable to a large range of architectures, but the current work employs rate coded, universally and asymmetrically connected neural networks, updated synchronously in time steps. The activity value $\boldsymbol{u}$ of unit $\boldsymbol{i}$ at time step $\boldsymbol{t}$ is calculated using equation 6 (below) – this is the same approach as used in chapter 5:

$$u_i(t) = \sigma\left(\sum_{j=1}^{N} w_{ji}(u_i(t-1))(1-m)+u_i(t-1)m\right) \qquad (6)$$

where $\boldsymbol{w_{ji}}$ is the weight of the connection from unit $\boldsymbol{j}$ to unit $\boldsymbol{i}$, $\boldsymbol{\sigma()}$ is the sigmoid function (bounded in the range 0-1) and $\boldsymbol{m}$ is a fixed momentum term with a value of 0.5. The network's categorisation choices are represented on two effector units, and each network may (or may not) include a variable number of hidden units (see Figure 27 for a schematic).

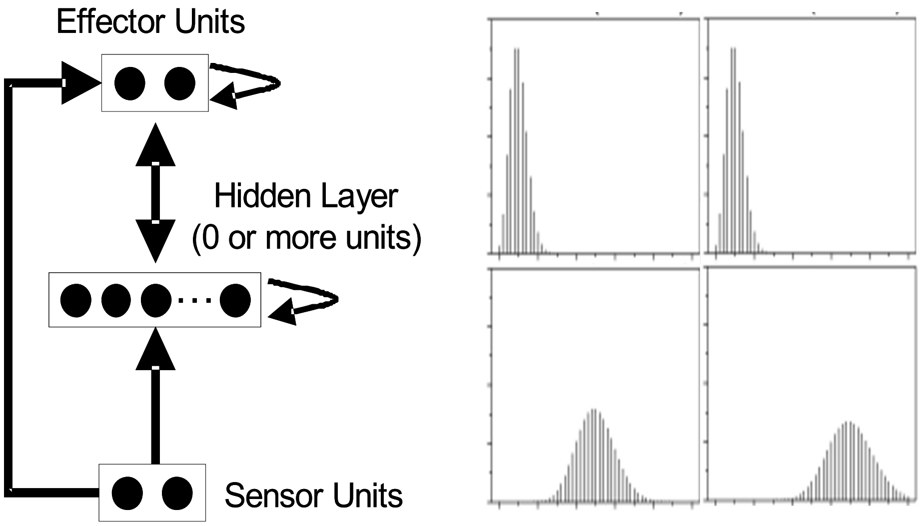


Figure 27: (Left) Schematic structure of the networks designed to solve a motion discrimination problem. With the exception of the sensor units, whose activity is fixed by the input signal, all of the networks' units are directly connected to every other, and to themselves. The size of the hidden layer can also change. (Right) An illustration of the Poisson distributions from which sensor unit values are drawn. During the pre-stimulus phase (top), both of the units' distributions have the same mean (15). During the stimulus phase (bottom), the units' distributions have different mean values; the network's job is to "select" the unit with the higher mean value.

The model-building method is familiar from chapters 3-5: a microbial genetic algorithm (Harvey, 2001). Seeking to avoid any unnecessary assumptions, the current version of this method has a slightly greater scope than that used previously. As before, the process starts with a population of (200) randomly specified neural networks – but in this case, that randomness includes *both* the networks' weights *and* their effector functions. For each of the two possible choices, the target range of effector unit activity values is defined by three real numbers in the range 0-1 – the first two values specify a centre point and the second a radius. Taken together, these values define a circular area in the effectors' (2-dimensional) state space; a choice is considered to have been made when the effectors' state enters one or other of these areas. The weights are initialised as in chapters 4 and 5 – random real numbers in the range 0-1. And in deference to a preference for simpler network architectures over more complex solutions, all of the models are initialised with no hidden units.

The algorithm proceeds by iterations. During each iteration, two networks are selected at random and used to create a “child” individual, defined by a combination of two operators – crossover and mutation. The crossover operator is a simple mixing of the parents (weight matrices and effector functions) that define the two “parents”; each parent supplies a given parameter value with a probability of 50%. The mutation operator implements a small, random change to the child's structure – usually an increment or decrement (with equal probability) of '0.01' to either a randomly selected weight or to one of the values that define the network's effector function. Less frequently (p = 0.01), the mutation operator can also add or remove hidden units, changing the network's total size. This process is also biased in favour of smaller networks, with removals being twice as likely as additions. Once created, the child network replaces the “weaker” of its two parents; after many repetitions, the effect is to propagate features of “fit” networks throughout the population, at the expense of features of “unfit” networks.

Fitness tests are conducted by exposing particular networks to a series of motion categorisation problems. Each trial starts with a pre-stimulus phase, which has a fixed length of 20 iterations, during which no stimulus is present; for each iteration in this phase, sensor unit values are drawn from a Poisson distribution with a mean value of '15' (scaled to 0.15). The networks' goal during this phase is to return their effector units' activity values to a "resting" state (both units have the activity value 0.15 ± <0.1). If they fail, the current trial ends – is counted as a "miss" – and the next trial begins. The stimulus phase lasts for 100 iterations. During each iteration of this phase, sensor unit activity values are drawn from two different Poisson distributions; the mean value for the *standard* sensor unit (representing the direction that does *not* have coherent motion) is '80' (scaled to 0.8), while the mean value for the *coherent motion* unit (representing coherent movement in a given direction) varies, from trial to trial, in the range 80-100 (0.8-1.0). Stimuli continue to be presented throughout the stimulus phase, regardless of whether or not the network makes a response; this approach makes it possible to define a powerful metric for the definition of fitness, discussed below.

In their mathematical treatment of this process, Shadlen and colleagues (2006) suggested that neural accumulation emerges in the pursuit of ever-greater reward rates; better solutions forge an effective compromise between response latency and response accuracy. It is possible to use that metric directly in the current system, defining fitness as the ratio of correct responses to the average response time – but this metric is rather too discrete to be useful. As the search proceeds, the architectures of the models in the population will tend to converge; that convergence naturally emphasises the mutation operator as the population's major source of variation, and each mutation implements a *small* random change. The result is that, often, pairs of randomly selected networks will be very similar indeed – sometimes so similar that they will make the same series of choices with the same response latency. To manage this possibility, we need a metric that identify when one network is *closer* to better behaviour, than another.

The approach used here is to replace reward rate with a metric pitched at the level of effector units, whose activity values are recorded throughout each trial. During the pre-stimulus phase, the networks' goal is to return their effector units to the resting state – we can check that by identifying the minimum distance d^0 between the effector state and the resting state's centre (0.15, 0.15).

During the stimulus phase, the networks' goal is to approach the target state as quickly as possible; that behaviour can be measured by summing the distances D^1 between the effectors' state and the target state throughout the stimulus phase. By this definition, fitter networks will minimise both d^0 and D^1 – but one further feature is required. Since the networks' effector functions can be changed, it is possible to achieve very small distances by making both target states identical. To prevent this from happening, we have to reward networks that use very different target states (i.e. with a larger distance d^C between their states' centres), and penalise them when those states overlap. This latter quantity is calculated as the length d^b of the vector between the points on states' boundaries that intersect with the line that connects their centres. If the direction of that vector is the same as that of the vector between the two centres, the states do not overlap, and d^b is set to '0'. Figures 28 and 29 illustrate how these variables are extracted, and Equation 7 specifies how they are combined in the definition of fitness.

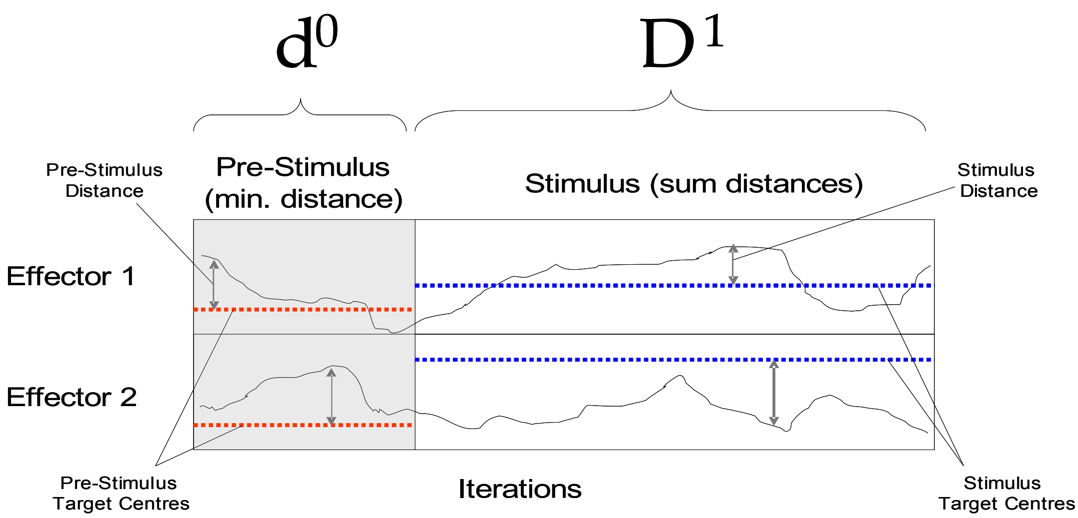


Figure 28: Illustrating the calculation of two distances – d^0 and D^1 – in a motion discrimination trial. d^0 is the minimum distance between each effector's activity values and the pre-stimulus target centres during the pre-stimulus phase. D^1 is the sum of the average distances between each effector's activity (at every iteration) during the stimulus phase. The distances are Euclidean – calculated after projecting the effector activity series (and target areas) onto a 2-dimensional space.

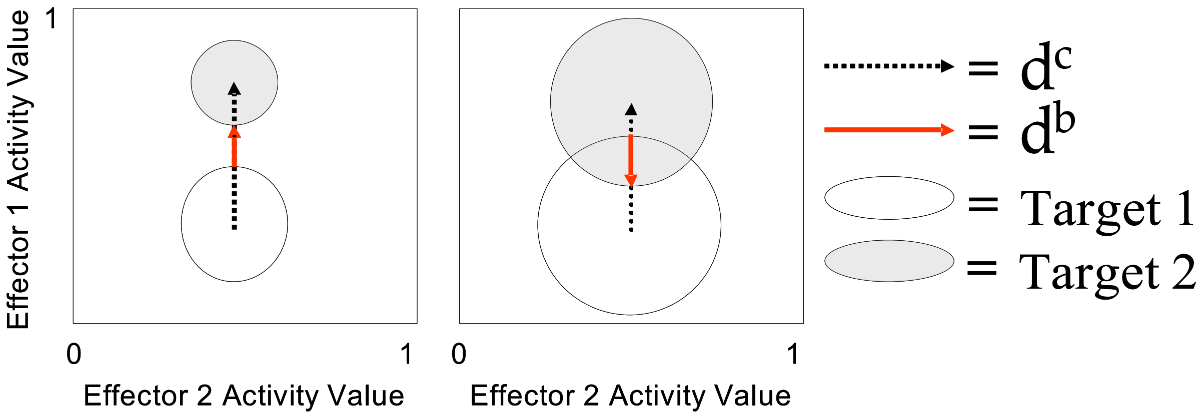


Figure 29: Illustrating the calculation of the effector functions' fitness variables. d^c is the distance between the centres of each target, while d^b is the distance along the same line, between the targets' closest boundaries. When the two vectors run in opposite directions, the targets overlap (right) – otherwise (left), the targets do not overlap and $d^b = 0$.

$$f = \left\{ \sum_{i=1}^{N} \left(\frac{d^c}{d_i^0 d_i^1 (1 + d^b)} \right) \right\} \qquad (7)$$

where d_i^0 is the minimum distance between the network's effector units and the "fixation target" during the pre-stimulus phase of trial i, d_i^1 is the *average* distance between those units and the centre of the correct response region during the stimulus phase of the same trial (i) – dividing the sum D_i^1 by the number of iterations in the stimulus phase (100), and N is the number of trials that compose a fitness test. The resultant measure – f – captures not just the network's rate of reward, but also how *close* its responses were to an optimal balance between speed and accuracy.

6.3 Results

The system requires about 400,000 iterations of the microbial algorithm to produce populations of networks in which the best individuals have low error rates (< 5%). After that time, the most accurate network in these populations reliably exhibits a good fit to the empirical data reported by Shadlen and colleagues (1999, 2001) – both at the level of RTs and error rates (Figure 30), and in terms of the correspondence between model dynamics and neurophysiological data (Figure 31). The learning was repeated five times, to ensure that its results are robust. All five attempts yielded models with equivalent behaviour; the data presented below were all drawn from the best network in the first of these populations, which has two hidden units.

This network is extremely robust to variations in the sensor stimuli that it receives. Though the standard distribution (indicating zero coherent movement in the corresponding direction) during learning was defined by a mean value of '80', the network is perfectly able to compare stimuli to different baselines, with mean values in the range 30-90.

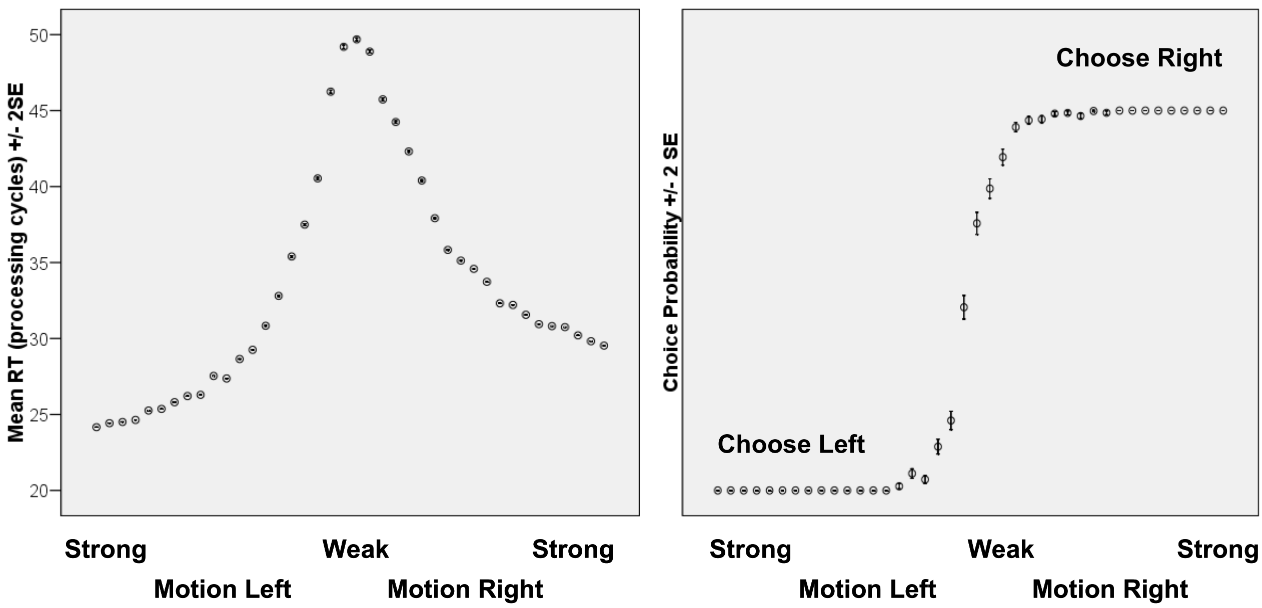


Figure 30: (Left) Choice behaviour as a function of signal strength. The network's responses are reliably accurate when the signal is strong (in either direction), but are more probabilistically symmetrical for weaker signals – like monkeys and humans in analogous tasks. (Right) Network reaction times by signal strength. Easier categorisation problems – with stronger signals (in both directions) are categorised more quickly than harder categorisation problems (with weaker signals). The pattern is similar to that observed in visual motion discrimination with both primates and humans.

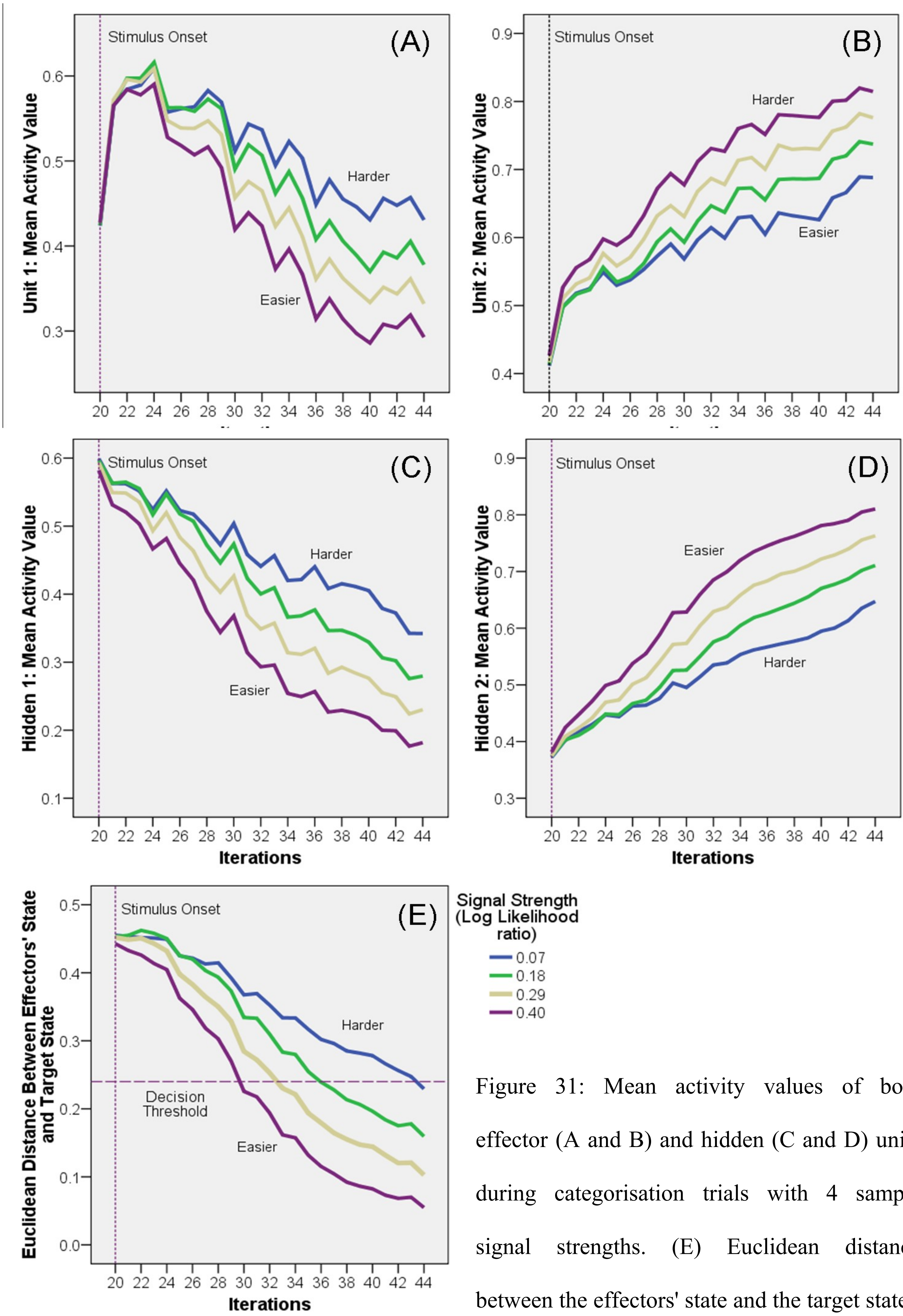


Figure 31: Mean activity values of both effector (A and B) and hidden (C and D) units during categorisation trials with 4 sample signal strengths. (E) Euclidean distance between the effectors' state and the target state.

Further, this model also captures the empirical distinction that separates reaction times for correct responses from those for incorrect responses (given the same stimulus strengths). Restricting the analysis to only those signal strengths that included both correct and incorrect responses, we can capture this by performing a t-test for paired samples (correct vs. incorrect RTs by signal strength) on the remaining values. The results confirm that the incorrect responses are slightly, but significantly, faster than correct responses (mean RTs (iterations): 40.2 (incorrect), 47.45 (correct); $t(16) = 2.252$, p (2-tailed) $= 0.039$).

Finally, it should be noted that the model's continuous operation makes its initial state at the beginning of a trial a function of its history – so potentially different at different times. This is important because it allows the network to respond differently, at different times, to exactly the same stimulus (i.e. exactly the same series of sensor values, presented in the same order). That flexibility is a consistent feature of the relevant empirical data (Shadlen & Gold, 2001), and has also been emphasised by previous models of this process (Shadlen & Gold 2001).

6.4 Interim Discussion

Driven solely by the intuition that neural systems implement near-optimal strategies, this model-building methodology can predict both the empirical behaviour of biological agents, and the dynamics of the neurons that drive it. But the method does not *guarantee* optimality – and for some, that weakness can be critical. For example, Norris and McQueen (in press), are prepared to reject Connectionism *completely* in order to ensure that their model (Shortlist B) of continuous speech recognition will be optimal; that switch is thought to be necessary because “it might be possible to build an interactive-activation network that computed the same Bayesian functions as Shortlist B, but it is also possible to build networks that compute other functions” (page 11, Norris & McQueen, in press). If even *possible* sub-optimality is grounds for rejecting an approach, then the method described in this chapter will be difficult to defend.

However, the pursuit of optimality is not necessarily the same as the pursuit of cognitive relevance. The optimality assumption is useful as much for the way it highlights sub-optimal performance as for its more positive, predictive success. Indeed, the primate neural strategies that the current model captures are almost certainly not optimal, because the formal theory that best describes them – the Diffusion-to-bound framework – is appropriate only when each evidential "event" is conditionally independent (Shadlen et al., 2006); the stimuli that drive the motion-detection problem will almost certainly violate that assumption. In other words, though the model described in this chapter does depart from an optimal strategy, but it does so in much the same way as the neural system that it is designed to capture.

A second example of this kind of connection stems from the model's RTs, which are significantly faster for incorrect responses than for correct responses. That distinction is also visible in the relevant empirical data (e.g. Palmer, Huk, & Shadlen, 2005; Roitman & Shadlen, 2002), but at least as conventionally defined, the Diffusion-to-Bound framework predicts that these latencies should be the same (Shadlen et al., 2006). Given the right parameters, Ratcliff, Van Zandt, and McKoon (1999) have demonstrated that diffusion models *can* capture this distinction; the current work demonstrates that these conditions can emerge without any explicit effort on the part of the designer. In some respects at least, the current method seems a powerful way to capture not just the optimality of the neural decision process, but also its apparent deviations from that optimal model.

One other possibly sub-optimal feature of the model is its size. Despite a strong selective preference for small networks, the best, discovered models reliably contain at least two hidden units. However, Wong and Wang's (2006) minimal treatment of this problem suggests that no hidden units should be necessary to capture Diffusion-to-Bound dynamics. Like the current model, primates appear to employ rather more neurons than they should need to implement the decision process;Kim & Shadlen's early data (1999) emphasised this difference by showing that single neurons could actually be *better* at discriminating sensory evidence than neural populations. In this context, the model has another theoretical significance.

When a population of neurons all have similar response properties, it seems natural to interpret them as components of a population code (e.g. Kim & Shadlen, 1999), which averages those responses to define behaviour. The current model illustrates that this assumption can be misleading; though both hidden units appear to accumulate evidence in much the same way (see Figure 31), they do not appear to play the same, causal role in the model's behaviour. While the model's ability to respond is almost completely undermined when informational lesions (as described in chapter 4) are applied to the hidden unit 1, the removal of the hidden unit 2 preserves the model's basic strategy (i.e. accumulation of incremental evidence), but reduces its power; after the lesion has been applied, the same rates of accumulation are only observed with much stronger signals, and weaker signals cannot be classified at all. The two units "look" very similar, but actually "do" quite different things – and the same might be true of neurons in the brain.

All of these deviations from optimality are probably best explained as local maxima in the models' fitness space; once encountered, these maxima capture the search, and prevent the discovery of more globally optimal structures. The same logic might also begin to explain why naturally evolved neural systems might often be nearly – but not perfectly – optimal. By accepting a model-building method that does not guarantee optimality, we may be able to begin to capture, and explain, that difference.

Perhaps more valuable still, this kind of method might eventually be employed to *discover* the normative formal theories that can drive further research – because the search considers both the strategy-space and its implementation-space at the same time. Since these two levels are not independent, constraints on one level of the search will also infect the other. Recently, Lengyel, Kwag, Paulsen, and Dayan (2006) have proposed a formal theory of memory storage and retrieval, which is pitched at the level of spike co-ordination in attractor neural networks; the current implementation of this method could never discover a solution of that sort because, by using a rate-coded neural network architecture, we have constrained the search to rate-based solutions. Nevertheless, the same logic – searching a model-space, and directing that search with a high-level metric that captures behavioural optimality – could be equally applied to more biologically detailed architectures like spiking neural networks. That extra detail would give the system some freedom to employ either spike coordination, or spike rate, or both, in the pursuit of optimal (or near optimal) behaviour.

III

On the Processing of Multi-Digit Numbers

Chapter 7

Priming the Holistic Content of Multi-Digit Numbers

In chapter 2, I mentioned that attempts to resolve the structure debate are hampered by the natural confound that exists between a multi-digit number's integrated value and its single-digit components. This section presents an attempt to dissociate these two. The intuition behind it is that subjects can be cued to give preferential weight toward either the single-digit components, or the integrated values, of the numbers that they must process. In particular, it seems likely that very large numbers – or very long digit strings – will encourage a much greater focus on the numbers' single-digit components at the expense of their integrated values. The approach centres on the "thousands" string ('000') – perhaps the most common signal that viewed numbers are very large – employed as a prime in a three-digit number comparison task. Appended to both of the numbers in particular trials, this string was intended to encourage the desired shift in the subjects' processing strategies. Any evidence of this kind of shift should support the hybrid theory of multi-digit number representation.

Prior work using identical number stimuli (but without the "thousands" string: Korvorst & Damian, 2007) has confirmed the emergence of three empirical phenomena in a similar number comparison task: the Number Size effect, the Distance effect and the (tens-hundreds) Congruence effect. In the material that follows, I use these effects as metrics for the influence that the numbers' holistic and componential content exerts. To the extent that the thousands string encourages an emphasis on componential over holistic processing, we should expect the addition of the thousands string to diminish the Number Size and Distance effects, while enhancing the strength of the Congruence effect.

7.1 Method

7.1.1 Subjects

Seventeen naïve subjects participated; 3 men and 14 women. The youngest subject was 21 years old, and the oldest subject was 34 years old. The mean age was 24 years. All but one of the subjects were right-handed, and all reported normal or corrected to normal vision.

7.1.2 Design

As mentioned previously, this experiment employed stimuli borrowed from Korvorst and Damian (2007). In their original form, these stimuli provide for a largely balanced manipulation of the Size and Distance associated with both the operands themselves and their single digit components. This experiment focussed on the further manipulation of the thousands string, appended (or not) to both of the numbers in particular number comparison trials. Following the logic of Lorch and Myers (1990), I use linear regression to capture the mediation that three predictors – two for the holistic content and one for the componential content – exerted on each subject's reaction times. The prediction was that the thousands string would cause a characteristic deviation toward zero (diminished strength) in the slopes associated with the holistic predictors (the minimum of and numerical distance between the logarithms of the compared numbers), but increased deviation (i.e. more negative values) in the slopes associated with the componential predictor. This latter predictor was calculated by subtracting the tens digit of the smaller operand from the tens digit of the larger operand; negative values indicate incongruent pairs, while positive values indicate congruent pairs.

7.1.3 Procedure

Each subject was exposed to two blocks of trials, with each trial containing a pair of numbers that they had to compare. In each case, the first block included only the original three-digit stimuli, while the second block included only those pairs in which the thousands string was appended to both numbers. This arrangement was designed to underline the irrelevance of the thousands string to the task in the second block. There were 320 number pairs in all; each was presented once both with and without the thousands string, for a total of 640 trials. Allowing for breaks between blocks, and for an initial series of 12 practice trials, the experiment lasted approximately 30 mins.

The beginning of each trial was signalled by the appearance of a point (a "full stop" symbol) in the centre of an otherwise blank image. After 500ms, this was replaced by a number pair, which remained on the screen for a maximum of 2,000m. Trials could be terminated early if subjects pressed one of two response keys – the "up" and "down" arrows on a standard keyboard – to indicate their comparison decision (the larger of the two numbers). Subjects were instructed to be accurate rather than fast, but we recorded both types of data throughout the experiment.

7.1.4 Visual Presentation Conditions

The experiment took place in a quiet, brightly lit room, with a Pentium III PC, a standard keyboard and a 15 inch colour screen. Subjects sat with their eyes approximately 50cm from the screen, but no chin rest was used. With the exception of the error feedback – a black exclamation mark on a red background – all stimuli were presented as white characters (rgb = 63,63,63) on a black background (rgb = 0,0,0). The numbers that composed each operand pair were presented in horizontal orientation, in Times New Roman font, with a font size of 30. Operand pairs were presented centrally, with vertical orientation (displaced 10 pixels above and below the centre-line) – the "tens" digit of each number was always defined as its centre, regardless of the presence of any trailing digit strings (i.e. the thousands string was always in the right visual field: see Figure 32).

Figure 32: Two examples of 3-digit number comparison problems. (Left) An example from block 1, using the three-digit operands in their original form. (Right) An example from block 2 that uses the same 3-digit operands, but appends a trailing string of zeros to each number.

7.1.5 Data Preparation

The results are derived after excluding first all trials in which reaction times were less than 200ms (thought to indicate anticipation), then practice trials (3.1%), error trials (2.4%), trials in which the hundreds-digits of the two operands were equal (15.7%: in these cases, the tens-digit comparison is relevant to the task, which complicates the interpretation of Congruence effects), and – following the recursive method of Van Selst and Jolicoeur (1994) – trials in which each subjects' reaction times were more than three standard deviations from the subjects' means (8 iterations required, excluding a further 4.8%).

7.2 Results

Linear regressions were computed on the data of individual subjects to obtain regression coefficients for each predictor. Using one-tailed, one-sample t-tests (df = 15 in all cases), we confirmed that, for each predictor, the means of the subjects' standardised regression coefficients (β-values) were significantly different from zero in both blocks (positive for the Size predictor, and negative for the Distance and Congruence predictors). Consistent with prior work (Korvorst &

Damian, 2007) the data display significant Size effects, Distance effects, and Congruence effects in both blocks of the experiment (see Table 3). Using 3-digit vs. 6-digit groups as a binary predictor, I also confirmed that there was no main effect associated with the "thousands" manipulation ($t(15) = -0.72$, p (2-tailed) = 0.481).

	Block	Test Value = 0			Mean Beta
		t	df	Sig. (1-tailed)	
Ln (Size)	1	7.636	15	<.001	.22200
	2	5.825	15	<.001	.14088
Ln (Distance)	1	-17.813	15	<.001	-.33575
	2	-17.481	15	<.001	-.34869
Congruence	1	-2.850	15	.006	-.05200
	2	-4.858	15	<.001	-.08844

Table 3: Main effects in both blocks. The t-tests are driven by samples of standardised regression coefficients, computed for each subject by regressing reaction times against each of the specified predictors.

To assess the interaction between this latter manipulation and the main effects, we can conducted a series of t-tests for paired samples (df = 15 in all cases). Each subject supplies a pair of values (standardised regression coefficients) for each predictor – the first for trials with the original three-digit stimuli, and the second for trials that included the "thousands" string. As predicted, the Size effect was significantly diminished (mean R^2 block 1: 0.062; mean β block 1: 0.222; mean R^2 block 1: 0.029; mean β block 2: 0.141; $t = 2.168$, p (1-tailed) = 0.024), while the Congruence effect was actually enhanced (mean R^2 block 1: 0.008; mean β block 1: -0.052; mean R^2 block 1: 0.013; mean β block 2: -0.088; $t = 1.907$, p (1-tailed) = 0.038). In some respects at least, the "thousands" string seemed to encourage the predicted dissociation between the subjects' sensitivity to holistic vs. componential content (see Figure 33).

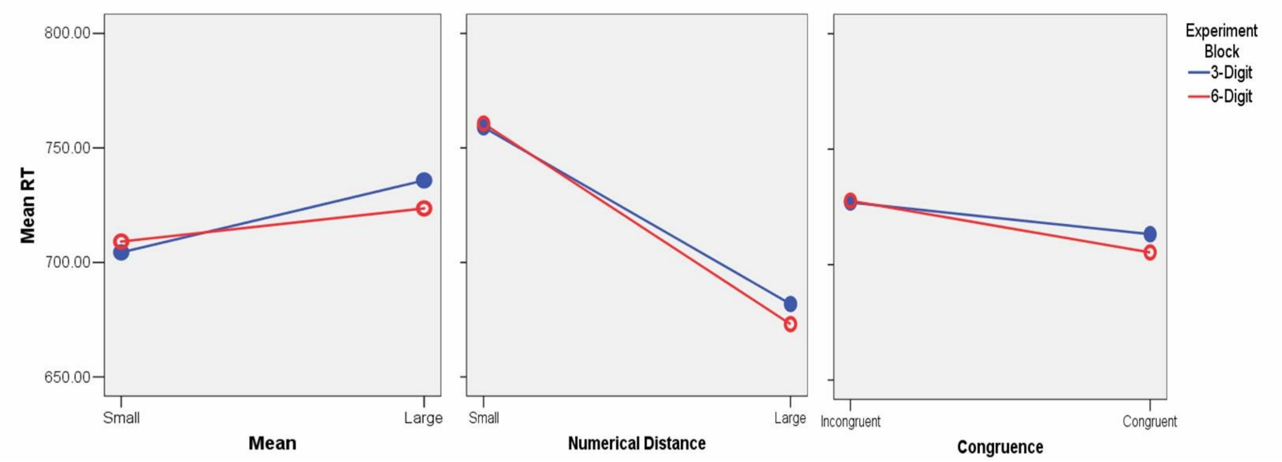


Figure 33: Mean reaction times as a function of Number Size (left panel), Numerical Distance (middle panel) and Congruence (right panel). The ranges of each predictor are divided into two groups (small vs. large) for the purpose of illustration, but were considered in their entirety when employed as predictors in linear regression analyses.

However, there was no significant deviation in the Distance effect (mean R^2 block 1: 0.103; mean β block 1: -0.336; mean R^2 block 2: 0.110; mean β block 2: -0.349; $t = 0.596$, p(1-tailed) = 0.721) – indeed, if anything, this effect displays a (non-significant) trend toward *enhancement*. The key to understanding why stems from a recognition that Congruence effects, which the thousands manipulation appears to amplify, are essentially a kind of digit-level Distance effect (calculated here as the tens unit of the larger operand minus the tens unit of the smaller operand). That connection raises the possibility that other digit-level enhancements might confound the observed mediation of Distance effects. To test the suspicion, I ran multiple, hierarchical regression analyses (by subject), inserting logarithmic Distance only after *linear* Distance had already been considered. As discussed in chapter 2, this latter predictor is much more susceptible to digit-level confounds than the former since, by definition, the numerical distance between two numbers is a linear function of the distances between the numbers' single digits (100 * hundreds-distance + 10 * tens-distance + units-distance). If digit-level enhancement is significant, we might expect the subjects' sensitivities to this linear predictor to be enhanced – and since this linear distance is strongly correlated with our logarithmic predictor (Pearson $r = 0.894$, p (2-tailed) < 0.001), that enhanced sensitivity could certainly mask any diminished effects of the sort we are trying to discover.

The results of this hierarchical analysis associate each subject with a β-value for linear distance, and a unique contribution (R^2) that logarithmic distance makes when linear distance has already been accounted for. Consistent with the intuition that digit-level enhancement can confound linear Distance, the subjects' sensitivities to the linear Distance predictor are enhanced in block 2 (mean R^2 block 1: 0.089; mean R^2 block 2: 0.121; mean β-value for block 1 = -0.293, mean β-value for block 2 = -0.341, $t(15) = 3.254$, $p(1\text{-tailed}) = 0.003$). And by observing the R^2 change associated with logarithmic Distance in both blocks, we can confirm that the unique contribution of logarithmic distance is significantly *reduced* in block 2 (mean R^2 change block 1: 0.037; mean R^2 change block 2: 0.014; $t(15) = 2.039$, $p(1\text{-tailed}) = 0.030$).

7.3 Interim Discussion

Consistent with prior work (Korvorst & Damian 2007), these results suggest that the subjects' reaction times were significantly mediated by both the Size and the Distance associated with each number pair, as well as by their Congruence. The variance explained by the Size and (corrected) Distance predictors was significantly reduced by the addition of a thousands string, which also appeared to increase the strength of Congruence effects. In other words, the prime appears to have been both successful and selective. Since the emphasis of one kind of content over another implies a sensitivity to the each independently of the other, this result supports a hybrid theory of multi-digit number processing.

Though still comparatively new, examples of this kind of top-down modulation has been observed before in number processing tasks. For example, Bächthold, Baumüller, and Brugger (1998) found that the subjects' spatial response biases (SNARC effects, as mentioned in chapter 2) for the numbers 1-11 could be mediated depending on the way they were told to conceive them. When the numbers were construed as measurements on a ruler, a conventional (small-left, large-right) bias was observed, but when the numbers were construed as hours on an analogue clock, the bias was reversed. These spatial biases have also been shown to depend on culturally acquired habits like reading or scanning direction (Zebian, 2005), as well as by learned finger-counting methods (Di Luca, Granà, Semenza, Seron, & Pesenti, 2006). Similarly, it has been shown that subjects' sensitivities to the holistic content of multi-digit numbers may be task-dependent; Ganor-Stern, Tzelgov, and Ellenbogen (2007) report that subjects' RTs are significantly mediated by that holistic magnitude in a number comparison task, but not in a physical size judgement task with the same numerical stimuli. Finally, recent data reported by Bonato, Fabbri, Umiltà, and Zorzi (2007) suggest that, in a number comparison task with fractions, the strategies that subjects use can be mediated by the type of fraction that they are given as a reference (i.e. a fixed standard against which to compare other fractions). Like the current work, that result implies a role for top-down effects that can operate within a single number processing task.

At the same time, the results appear to establish a dissociation between the Size and Distance effects; only the former appear to be significantly mediated by the thousands string. I have suggested that this may be because Distance effects are confounded by enhancement at the level of single digits, but since there is no evidence of corresponding confounds on Size effects, the dissociation between Size and Distance still needs to be explained. There are prior examples of this kind of dissociation. Verguts and Van Opstal (2005) report that subjects RTs were mediated by both effects in a number comparison task, but only by Distance effects in a same / different judgement task with the same numerical stimuli. And Verguts and De Moor (2005) report a dissociation even within a single number comparison task, suggesting that both Distance and Size effects are significant when subjects compare two-digit numbers with the same tens digit, but that only Size effects are significant in pairs with different tens digits. Contrary to the latter study, the current results appear to confirm that there is a Distance effect for holistic magnitude that is independent of digit-level confounds; the change in the F statistic associated with the addition of logarithmic distance after linear distance has been taken into account is significant for all but three subjects in block 1. Nevertheless, the results do support those authors' preferred theory of number representation.

All but one of the popular accounts of number representation (the *compressed MNL*, the *noisy MNL*, and the *numerosity code*) associate both Size and Distance effects with interference at the level of number representations – so none can explain the dissociation that these data define. The fourth theory – the *barcode* – construes Size effects as a processing cost (specific to number comparison tasks) that emerges because small numbers are experienced more frequently than large numbers during learning. Since the effect emerges from the process that maps number representations to task responses, it makes sense that it should be dominated by the integrated values of the numbers that must be compared. And since this theory construes Size and Distance effects as emerging from different causal sources, it can also allow for dissociations between them.

Finally, it should be noted that the cueing effect which drives these results may not be specific to the thousands string; from the current work, we cannot rule out the possibility that the same deviations might be observed with strings of random symbols. Given the results of prior experiments in which number-specific effects were observed even when the magnitudes of number stimuli were irrelevant to the task (e.g. the SNARC effect in parity judgement tasks: Dehaene, Bossini, & Giraux, 1993; Distance effects in same / different judgement tasks: Verguts & Opstal, 2005; magnitude priming effects in number naming: Reynvoet, Brybaert, & Fias, 2002; and numerical flanker tasks: Nürk, Bauer, Krummenacher, Heller, & Willmes, 2005), a number-processing interpretation of the string's effect seems both natural and credible. Nevertheless, these effects' specificity could usefully be confirmed by repeating the experiment with a non-numerical prime. Critically though, the results' value is largely independent of the outcome of that work; the confirmation that subjects can preferentially process different parts of a multi-digit number's semantics is important regardless of the nature of the cues that drive it.

Chapter 8

A Model-Space for Three-Digit Number Comparison

Chapter 8 provides a clear demonstration of the way in which empirical data can interact with both the format debate and the structure debate at the same time. This chapter explores that interaction from a more computational perspective, reporting a structured comparison of models driven by different representation configurations. The method extends an approach previously employed by Zorzi and colleagues (under revision); in that previous work, the goal was to compare the different theories of MNL format that were mentioned previously (see Figure 6), and the authors restricted their analysis to operations involving small numbers. The scope of the current work is expanded in both respects. This chapter considers both MNL formats and MNL structures, and – since MNL structures are only an issue when the numbers involved are large – reports results derived from models trained to compare operands in the range 0-999. To my knowledge, these are the first examples of models that can compare such a large range of numbers.

8.1 Method

8.1.1 Model Architecture

The initial assumption – commonly accepted in the literature (Zorzi, Stoianov, & Umiltà, 2005) – is that number comparison can be effectively modelled by structures that learn associations between pairs of numbers and comparison decisions (e.g. which number is larger). Learning in these structures proceeds by associating successive operand pairs (e.g. 4 vs. 7) with desired responses (e.g. “7 is larger”); if successful, a model will extract and encode appropriate rules from its training set. In the current work, number comparison is modelled using the Mean Field Boltzmann Machine (mfBM) architecture (e.g. Ackley, Hinton, & Sejnowski, 1985), a recurrent, rate coded, symmetrically connected neural network in which every unit is linked by weighted connections to every other unit. This is a pragmatic choice, because mfBM's learn by associating input patterns with point attractors (steady states); given an input pattern (number comparison problem), the number of processing iterations required for the network to reach a steady state is naturally interpretable as an analogy to reaction times (Zorzi & Butterworth, 1999). But theoretical justification is also available because the mfBM architecture uses a local, biologically plausible learning rule (Ackley, Hinton, & Sejnowski, 1985), and has been the basis of successful cognitive models in the past (e.g. Zorzi, Stoianov, & Umiltà, 2005; Zorzi et al., under revision).

The architecture of an mfBM is illustrated in Figure 34. A subset of the network's units are defined as input units, whose activity is fixed to represent those numbers that must be compared. Choices concerning the way in which those inputs are fixed correspond to theories of the way in which compared numbers are represented. The network's "decision" is implemented on two decision units (one for each operand). A correct response is made when, once the mfBM has achieved a steady state, the unit corresponding to the larger operand has supra-threshold activity, and its counterpart has sub-threshold activity - a threshold of '0.5' is used for all models. Finally, the mfBMs can also include a small group of "hidden" units; the current work considers models with 0, 20, or 40 hidden units. The activity **u** of unit **i** at time **t** is calculated exactly as in chapter 5 (equation 8):

$$u_i(t) = \sigma\left(\sum_{j=1}^{N} w_{ji}(u_j(t-1))(1-m)\right) + u_i(t-1)m \qquad (8)$$

where $\boldsymbol{w_{ji}}$ is the weight of the connection from unit $\boldsymbol{j}$ to unit $\boldsymbol{i}$, $\boldsymbol{\sigma()}$ is the sigmoid function (bounded in the range 0-1) and m is a fixed momentum term with a value of 0.5.

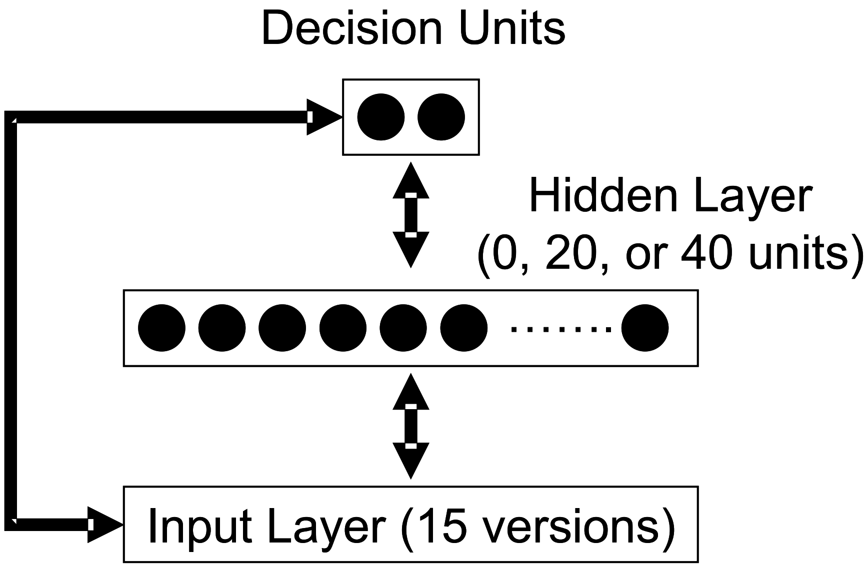


Figure 34: Schematic structure of the Boltzmann machines that were trained to solve number comparison problems with numbers in the range 0-999. With the exception of the input units, whose activity is always fixed according to a given representation, all of the units are directly and symmetrically connected to every other unit.

8.1.2 Learning

The conventional mfBM learning rule (Ackley, Hinton, & Sejnowski, 1985) is a two-stage process. In the first stage, both input and decision units are clamped to values defined by example input and desired decision patterns. Activity from these units is then allowed to propagate freely throughout the network's hidden units, until these also reach a steady state (i.e. maximum change for any unit between two iterations < 0.001). The network's input and decision units are then unclamped and the process repeated. At the end of both the clamped and unclamped phases, the correlations associated with each pair of network units are collected and recorded. The learning rule works by minimising the difference between these two sets of correlations, usually computed over the whole training set. Specifically, the update for the weight ***w*** between units ***i*** and ***j*** is calculated using equation 9.

$$\Delta w_{ij} = \lambda\left((x_i^+ x_j^+) - (x_i^- x_j^-)\right) \qquad (9)$$

where Δw_{ij} is the update for the weight of the connection projecting from unit ***i*** to ***j***, x_i^+ is the activity of unit ***i*** during the unclamped phase, x_i^- is the activity of unit ***i*** during the clamped phase, and λ is a constant that defines the learning rate (smaller values yield slower, more accurate, learning).

The models in this chapter are trained to solve number comparison problems for all numbers in the range 0-999. Pairs of identical numbers (e.g. “7 vs. 7”) are not considered, leaving a total set of 999,000 valid pairs. In deference to the colossal size of this set, these models use an augmented version of the conventional learning rule. Where the original method collected statistics from a batch of training problems, the current variant calculates and applies weight updates after every single problem that is presented during training. Further, the current method employs a semi-supervised version of the conventional BM learning – allowing the decision units to vary freely, but keeping the input units fixed during the unclamped phase. In a few previous applications of the mfBM, it was desirable for the model to learn a *bidirectional* association – allowing either inputs or outputs to drive pattern completion. By enforcing this restriction, we have also restricted the model to learning a *unidirectional* association (from inputs to outputs, but not vice versa) – sacrificing some (in this case, irrelevant) power for faster results.

Thirdly, since single-digit comparison problems represent less than 0.01% of this set (and two-digit problems account for less than 1%), we have to manipulate the frequency with which particular problems are encountered so that single-digit, two-digit and three-digit operands are equally frequent. This is necessary because, without that manipulation, these problems may be extremely rare in the models' training – so subject to implausibly high error rates in their trained performance. Finally, a system is implemented that skews the selection of number pairs toward those that have been error-prone in the past. Each operand pair is associated with an integer (a “tracker”), which is incremented whenever the network offers an incorrect response to that pair, and decremented after correct responses. As the value of a number's tracker increases, so too does the probability that it will feature in subsequent trials; the first increment raises selection probability to

100 times its basic value, the second to 200 times times that value, and so on. Taken together, these four extensions make it possible to train every model configuration to a reasonable standard in a reasonable time. In the current work, that standard was taken to imply accuracy rates of 90% or more when tested against the number pairs in our data sets; for each model, that accuracy was tested after every 3,000 learning problems, and the learning process was stopped once the required accuracy had been achieved.

Though effective, this learning regime also allows for a great deal of stochastic variation into the model-building process. Every number pair for numbers in the range 0-999 *can* be selected for learning, but in practice, the learning for particular models tends to be dominated by a comparatively small subset (the first few hundred pairs on which the particular network made errors). The training sets associated with different networks will share some pairs in common – as mentioned above, single-digit and two-digit pairs are always over-represented – but there will also be significant differences between those sets. To try to ensure that our conclusions are robust to this variation, we consider 9 versions of model driven by each representation configuration; 3 repetitions for models with each of 0, 20, and 40 hidden units.

8.1.3 Representation Configurations

This chapter considers each of the four MNL formats discussed previously (chapter 2) – the *numerosity code,* the *compressed MNL*, the *noisy MNL*, and the *barcode* – plus a fifth – the *single-unit accumulator code* – that was identified in chapter 5. Rather than trying to match this code's original form too closely, the current version is an idealised variant of that form – illustrated in Figure 35 – in which all units accumulate activity at the same rate, with increasing number. The same figure also illustrates a new, continuous variant of the *numerosity code*, which replaces the code's original structure because the latter implies the addition of an extra input unit for every represented number (on a holistic MNL). Models with 999 input units per operand were attempted –

and can be made to work well – but take an extremely long time to train; this change permits the use of a recognisable *numerosity code* in far fewer input units, so ensures that all of the codes can be represented on the same number of units.

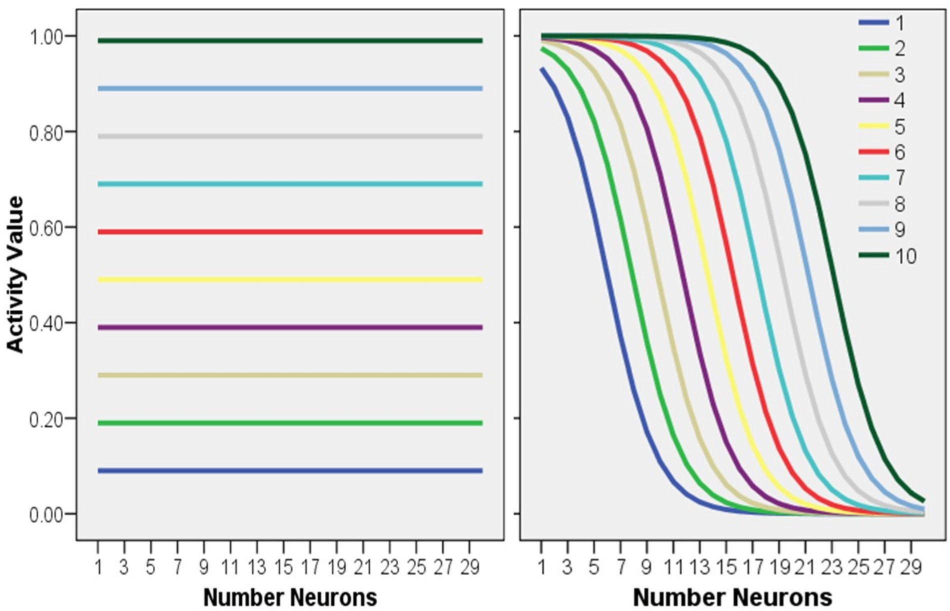


Figure 35: Two MNL formats, expressed as activity values on 30 number-sensitive units. (Left) The *single unit accumulator code*, in which the activity of all units accumulates linearly with increasing number. (Right) A continuous variant of the *numerosity code*, in which activity accumulates linearly *across* the units with increasing number. Both formats are illustrated for numbers 1-10, but both also allow much larger ranges to be expressed on the same number of units, by reducing the logical distance between the representations of numerically adjacent numbers.

For each MNL format, three global structures are also considered – again, as discussed in chapter 2. Holistic structures imply that a single MNL is used to represent the entire sequence of possible operands – in this case, 0-999. Componential structures imply a representation that includes one MNL for every operand digit; the number '999' requires three MNLs, for a total of six to represent a comparison problem. Hybrid structures use both holistic and componential number lines – in other words, a hybrid representation of the problem '999 vs. 998' implies a total of eight MNLs. Five MNL formats and three MNL structures correspond to 15 representation configurations.

8.1.4 Behavioural Data

Alongside the models themselves, the current work also employs three data sets, from three similar behavioural experiments in which human subjects performed number comparison tasks. The first data set, derived from experiments conducted by Butterworth and colleagues (2001), refers to comparison with operands 1-9 (data set 1). The second set was derived from Gazzelini and Laudanna's (2005) study, and refers to operands 0-99 (data set 2). The third data set, collected by Korvorst and Damian (2007), refers exclusively to three-digit operands (data set 3); this set was also used to define the experiment in chapter 7. Data set 1 contains 72 items, data set 2 contains 380 items, and data set 3 contains 320 items, for total of 772 number comparison problems, each associated with a mean empirical reaction time.

In the current context, the most important features of these data are the empirical phenomena that they display. In particular, this chapter's analyses will focus on the Number Size and Distance effects, as well as the Congruence effects that appear in data sets 2 (a unit-decade effect) and 3 (a decade-hundred effect). Each of these effects can be captured by relating reaction times (both empirical and model-driven) to particular predictors with linear regression analyses. Size effects can be captured by regressing RTs against the minimum of the two compared numbers, and Distance with a regression against the absolute numerical difference between them. Congruence predictors are calculated by subtracting the relevant distractor digit of the smaller number from the corresponding digit of the larger number; in the case of units-tens congruence, the relevant digits are the units, while for tens-hundreds congruence, the relevant digits are the tens. In both cases, a negative result indicates an incongruent pair – a pair in which the smaller number contains larger distractor digits. As applied to the subjects' reaction times, the results of the relevant regression analyses confirm that all three effects are present in the data (see Table 4).

The exception to this rule is data set 2, which does not yield a significant Distance effect. That absence stems from the details of the number pairs in this set, which were selected to emphasise Congruence effects (Gazzelini & Laudanna, 2005). The numerical distances between the numbers in each of this sets pairs are always either '2' or '6'; mean RTs are longer for the former distance (638.5 ms, SE = 3.52) than they are for the latter (633.2 ms, SE = 4.81), but the distributions are too variable to allow for a significant dissociation.

Predictors	Data Set 1			
	R^2	p	b	β
Size	0.160	<0.001	7.897	0.415
Distance	0.419	<0.001	-12.446	-0.653
Predictors	Data Set 2			
	R^2	p	b	β
Size	0.217	<0.001	0.979	0.466
Distance	0.002	0.435	-1.157	-0.040
Cong. (U-T)	0.109	<0.001	-4.298	-0.330
Predictors	Data Set 3			
	R^2	p	b	β
Size	0.340	<0.001	0.195	0.583
Distance	0.545	<0.001	-0.216	-0.739
Cong. (T-H)	0.097	<0.001	-4.349	-0.312

Table 4: Linear regression analyses relating subjects' reaction times to predictors for Size, Distance and Congruence effects.

8.1.5 Procedure

The models were all created using Microsoft Visual C++, on an IBM-compatible desktop PC. The data that this chapter reports are drawn from a sample of models, covering the 15 representation configurations, each constructed with 0, 20, or 40 hidden units. Each specific model configuration was reproduced 3 times, for a total of 135 models in all. After training, these models are associated with 135 tables, containing the models' error rates and reaction times for the number pairs in the three data sets. All of the trained models exhibit error rates of less than 10% when tested against the problems in our data sets, indicating good learning performance. I also attempted to produce very large models, with up to 100 hidden units, but those attempts were abandoned in the face of extremely slow training times.

This work is more concerned with the model-space itself than with particular points in that space. In the material that follows, I repeat the analyses employed to generate Table 1 for each of the 135 models, generating regression results that indicate the extent to which each model captures the Size, Distance and Congruence effects described previously. The results yield a series of standardised regression coefficients (β values) associating each model with each predictor for each data set. Those values are normally distributed, so can themselves be analysed with standard statistics (e.g. Lorch & Myers, 1990); that meta-analysis is the focus of the current work.

Following the logic of Spieler and Balota (1995), I also consider the β-values derived from regression analyses that relate model RTs directly to subjects' RTs in each of the three data sets. Each of the three data sets was derived from similar experiments, but no deliberate attempt was made to match them – to guard against artefacts of variation between these sets, β-values were derived for each set separately. The result is that each model yields several β-values; three for Size effects (because Size effects are present in all three data sets), two for Distance effects (which are present in data sets 1 and 3), two for Congruence effects (one for each of the sets that include multi-digit numbers) and three for empirical RTs – for a total of 10 β-values per model.

8.2 Results

Effective models of multi-digit number comparison must display Size effects, Distance effects, and

Congruence effects; these are a natural initial focus for the analysis. In practice though, two of these three – the Size and Distance effects – seem to be rather too common to be useful. Grouping the the models by MNL format and MNL structure, we can use single sample, one-tailed t-tests to assess the significance of both effects in each group. Distance effects are significant in every one of these (15) groups (see Table 5), while Size effects are significant in all but two of the groups (see Table 5). Both of the exceptions are associated with the *barcode* format, reflecting the absence of magnitude-dependent compression in this code (mentioned in chapter 2). As mentioned previously, the models' training regime did include skewed frequency in favour of smaller numbers, but in most cases, that skew does not seem sufficient to produce the desired effect. Since that skew operates *between* rather than *within* orders of magnitude, we should expect Size effects to disappear even in the most apparently effective *barcode* configuration (i.e. with 20 hidden units, and a decomposed structure) when the analysis is restricted to data set 1 (single-digit numbers). The current data include just three models that are relevant to this prediction, just one of which appears to produce a significant Size effect in this case – but though intriguing, this exception also illustrates the variability that our stochastic learning method can inject into the data, so supports the intuition that robust conclusions can only be attached to *groups* of models in this case.

Since both effects are associated with almost every representation configuration, neither can be used to justify a preference for one configuration over another. However, the Congruence predictor is much more selective. Just one of the 15 groups is associated with significant effects – the combination of *decomposed* MNLs with the *single unit accumulator code* (see Table 5).

MNL Format/Structure		Distance				Size				Congruence			
		t	df	p	Mean	t	df	p	Mean	t	df	p	Mean
Numerosity Code	Holistic	-7.557	17	.000	-.45794	3.176	26	.002	.21593	1.486	17	.922	.02089
	Comp	-6.209	17	.000	-.31589	2.316	26	.015	.09670	.198	17	.576	.00767
	Hybrid	-5.004	17	.000	-.23778	2.232	26	.017	.10496	-.817	17	.213	-.02422
Compressed MNL	Holistic	-1.892	17	.018	-.20728	4.865	26	.000	.33196	.706	17	.755	.00828
	Comp	-3.979	17	.001	-.25322	4.189	26	.000	.17544	1.959	17	.967	.09561
	Hybrid	-5.659	17	.000	-.34144	2.752	26	.006	.16681	-1.302	17	.105	-.02178
Noisy MNL	Holistic	-5.337	17	.000	-.32567	3.724	26	.000	.19030	1.878	17	.961	.02717
	Comp	-6.117	17	.000	-.31444	3.338	26	.002	.15785	.286	17	.611	.00733
	Hybrid	-5.502	17	.000	-.28672	3.256	26	.002	.15685	-.786	17	.222	-.01672
Barcode	Holistic	-2.641	17	.005	-.22822	.719	26	.240	.04685	-.300	17	.384	-.00483
	Comp	-4.387	17	.000	-.31472	5.513	26	.000	.18941	1.524	17	.927	.06650
	Hybrid	-4.803	17	.000	-.24700	1.522	26	.070	.07567	2.415	17	.987	.04494
Single Unit Accumulator Code	Holistic	-9.319	17	.000	-.52750	4.113	26	.000	.20111	.938	17	.812	.01261
	Comp	-7.238	17	.000	-.29450	2.184	26	.019	.10767	-2.851	17	.006	-.08972
	Hybrid	-5.395	17	.000	-.26828	3.815	26	.000	.13933	.314	17	.622	.01078

Table 5: One-sample, one-tailed t-tests (vs. zero) for β-values derived by regressing model RTs against each of the three predictors; highlighted cells indicate significant effects ($p < 0.05$). All representation configurations are associated with models that produce significant Distance effects. All but two appear to yield significant Size effects. But only one representation configuration appears to encourage significant Congruence effects. "Comp" refers to the decomposed MNL structure.

The same conclusion can also be reach in a different way, with a univariate analysis of variance (ANOVA) that tracks the mediating effects of MNL format (3 levels), MNL structure (5 levels) and hidden layer complexity (3 levels) on the β-values associated with the Congruence predictor. This analysis reveals exactly one significant effect – the two-way interaction between MNL format and MNL structure ($F(8,360) = 2.867$, $p = 0.003$). And by visual inspection, we can see that this interaction favours exactly the same representation configuration (see Figure 36). This conclusion is robust to analyses that consider β-values averaged over the different data sets (i.e. one value per model rather than two values – one for data set 2 and one for data set 3), and different versions of the same model (i.e. one value per model configuration, rather than three – one for each version of the same model).

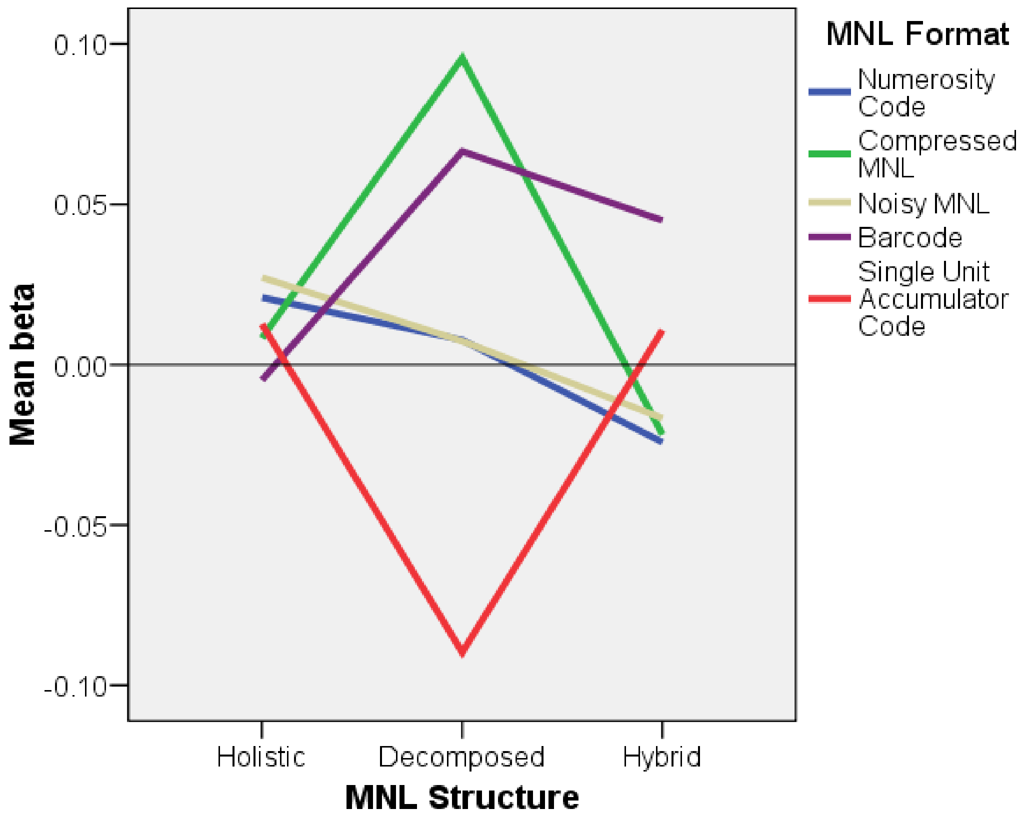


Figure 36: A significant interaction between MNL format and MNL structure in the model-space; regression coefficients are captured by regressing model RTs against the Congruence predictor, and more negative values indicate stronger Congruence effects. In this respect, the most effective representation configuration is the combination of *single unit accumulators* and decomposed MNLs.

From Table 5, we can see that the same group of models also reliably yield both Size effects ($t(26) = 2.184$, $p = 0.019$, mean β-value = 0.108) and Distance effects ($t(17) = -7.238$, $p < 0.001$, mean β-value = -0.296). And perhaps unsurprisingly, the RTs of models in this group are also significantly related to the relevant empirical RTs; in this case, we can use the β-values derived by regressing RTs from each of the models in this group directly against empirical RTs ($t(26) = 4.760$, p = <0.001, mean β-value = 0.157, mean R^2 value = 0.054). Taken together, these results make the case that combination of the *single unit accumulator code* and *decomposed* MNLs can drive effective models of multi-digit number comparison.

8.3 Interim Discussion

Computational cognitive models are usually reported in one of two ways. When dependent on a complex set of parameters, the description tends to focus on the best parameter values, masking the process that was used to select them (e.g. Grossberg & Repin, 2003). Parameter-space exploration is sometimes reported, but tends to demand extremely minimal models, with very few free parameters (e.g. Wong & Wang, 2006). This chapter has attempted to chart a compromise between these two extremes, reporting a space of reasonably complex models, along with statistics that capture some of the properties of that space. With the identification of a significant interaction between MNL format and structure, the analysis supports the intuition expressed in chapter 7, that the format and structure debates might not be quite as distinct as previously thought. It also directs our attention to a particular representation configuration – a combination of the *single unit accumulator code* and *decomposed* MNL structures, that seems to capture the relevant empirical phenomena. This result is particularly interesting because the code itself is novel – discovered in the "evolved" foraging agents described in chapter 5. At the time of writing, I have no convincing explanation for why this *single unit accumulator code* should be so much more effective than its counterparts; that question remains to be solved. But the result is intriguing because it seems quite robust to variations in the detailed structure of the models that it defines.

Alongside the result itself, the importance of a comparatively *weak* phenomenon in this analysis has a theoretical significance. Following the logic proposed by Spieler and Balota (1995), we might be tempted to emphasise the item-by-item correspondence between model and empirical RTs at the expense of other features in our analyses. But when, as now, comparatively weak phenomena are important, the current data confirm that this kind of analysis can be misleading. Since the variance of empirical RTs is dominated by Size and Distance effects (see Table 4), strong relationships between model and empirical RTs can mask non-significant Congruence effects; this is confirmed by an analysis of the group of models that use holistic MNLs with the *single-unit* *accumulator code*; though this is the "best" representation configuration as measured by average R^2 (0.133), these models display no significant Congruence effects. (see table 5; the mean R^2 for this group vs. the Congruence predictor is 0.003).

Nevertheless, if possible, we would like to find model configurations that yield both the desired empirical effects and strong item-level correspondence to empirical data; only the former can be claimed in the current case. Some models do satisfy the latter claim in some data sets – for example, one model (employing the s*ingle unit accumulator code*, holistic MNLs, and 20 hidden units) accounts for 60.0% and 49.2% of the empirical variance in data sets 1 and 3 respectively – but none can satisfy both. By visual inspection, the problem appears to be that the Size and Distance effects (particularly the latter) capture rather more of the variance in empirical RTs than they do in the models' RTs; to attempt to learn why, it may be necessary to investigate each model's detailed architectures directly, or to consider other dimensions of variation in an expanded model-space. However minimum standards for variance explained are difficult to define in this task because that level of analysis is completely absent from the only prior report of comparable work (Grossberg & Repin, 2003).

Consistent with the discussion in chapter 2, the analysis also confirmed that all combinations of MNL structures and format can support Distance effects, and that most can support Size effects. In the context of the structure debate, this result is useful because it tells us that *decomposed* MNLs can drive effects that are usually associated with *holistic* representation – confirming the intuition that *hybrid* representations might be unnecessary. However, of these two holistic effects, only the latter (Distance effects) are convincingly driven by the representations themselves – because even without magnitude-dependent compression, some models that employed the *barcode* do still produce Size effects. That result reflects the problem frequency manipulation that the training regime encodes; problems in which the maximum number has 1, 2, or 3 digits respectively were all equally frequent. As mentioned previously, attempts to train these models without *any* frequency manipulation were largely unsuccessful (because they resulted in models with disproportionately high error rates on single-digit and two-digit comparison problems). If we accept that association is at all reasonable as a strategy for modelling this task, the experience of this work suggests that, contrary to the doubts expressed in chapter 2, the problem frequency manipulation required by the *barcode* must be plausible as well.

For all the complexity of the process that was used to discover it, the models that use our best, discovered representation configuration are also actually quite simple – and certainly simpler than Grossberg and Repin's (2003) Extended ESpaN. The Extended SpaN employs both temporally structured inputs and a topographically organised store of number knowledge, while current model assumes only that particular digits are distinguished at the level of semantic number representations, and leaves the rest to the process of learning.

Further, these results reflect a very significant level of generalisation; though the precise sequence of problems encountered during learning is different for each model, none views more than 10% of the 999,000 number comparison problems that they can, in fact, solve. Without that level of generalisation, it would be difficult to claim any great cognitive plausibility for these models. In normal life, people simply do not see very many multi-digit numbers, and those experienced at school are necessarily a tiny proportion of the range of possible number comparison problems – but normal adults can reliably compare them all. The current models are limited in a way that their referents are not, because their performance will not generalise to numbers with more than three digits – but the generalisation to very large numbers almost certainly implies a significant, perceptual component that is beyond the current scope. The next chapter considers this latter problem directly.

This work has raised some questions that remain to be answered. Though the model-space approach yields rather richer results than any single model, it also poses some significant, technical challenges. In deference to the enormous number of models in this space, this chapter has been rather more *descriptive* than *explanatory*. Why, for example, is the combination of decomposed MNLs and the *single unit accumulator code* so effective? What differentiates it from the much more familiar *numerosity code* (both codes imply a similar metric of number similarity, so both might be expected to yield similar results)? Another question hangs over the roles of the hidden units in this space; since multi-digit number comparison can clearly be performed without any of these units, it may be unsurprising that they play no significant role in the mediation of Congruence effects. Nevertheless, there are some hints that these units *can* play a role – with 20 hidden units, one *barcode*-based model did appear to show significant Size effects in data set 1. Given a fixed preference for a particular representation configuration, the details of this role might be clarified in future work.

To some extent, these questions are a natural consequence of these models' complexity – precisely the reason that much simpler models are so often preferred. As we saw in chapters 4 and 5, even comparatively small neural network architectures can be difficult to analyse effectively; with 135 models, that difficulty is naturally increased. On the other hand, it seems sensible to assume that, when single models are reported, their designers have gone through an analogous process to find the "best" parameters for it. By reporting the *process* as well as the result, this chapter can make at least some conclusions that could simply not be made with a single model. Nevertheless, the explanatory cost of this approach – the problem of analysing large spaces of complex computational models – can and should be addressed in future work.

Chapter 9

Digit String Processing with Single Fixations

In chapter 2, I mentioned that theories of numerical cognition – and the computational models that encode them – tend to minimise the role of lower-level perception in number processing tasks. MNL-based accounts of, for example, the Distance effect, depend on mutual interference at the level of complete number representations, so can often ignore the question of how those representations are derived from lower-level perception. That same agnosticism is also evident in the field's canonical computational models, almost all of which involve some kind "clamping", or fixing of the relevant input units to their designers' preferred representations (Zorzi, Stoianov, & Umiltà, 2005). And even when lower level processes *are* taken into account, their presence is not usually critical to the results that justify these models' behaviour; as we saw in chapter 8, learning models that employ the *compressed MNL* are perfectly capable of producing Size and Distance effects without any of the perceptual pre-processing that characterised the code's original implementation (Dehaene & Changeux, 1993). To a large extent, this bias reflects the intuition that visually presented numbers (even multi-digit numbers) are perceived in a largely parallel manner (e.g. McClelland & Rumelhart, 1981); to the extent that this is true, it seems reasonable to dissociate perception from representation in theories of number processing. But ever-longer digit strings must eventually outstrip the brain's ability to process digits in parallel, implying a shift to some sort of incremental, or sequential, strategy. At present, we know next to nothing about the structure of that sequence.

To begin to understand it, we need to establish some of the fundamental constraints on the way that people perceive numerical stimuli. The most natural way to do this is by borrowing from the tools and techniques that have emerged from over fifty years of research in reading. With the possible exception of the "thousands" string, discussed in chapter 7, digit strings do not support the kind of high-level semantics more commonly associated with words and sentences – so the best analogy to digit strings is probably a string of random letters. This chapter reports two experiments that apply paradigms developed for random letter strings to digit string perception. Specifically, the experiments were designed to answer two questions:

- How many digits can participants see with single fixations?
- How many digits can participants identify with single fixations?

Each of these questions addresses a fundamental stage in the processing of multi-digit numbers, and both are equally important. Digit identification plays an obvious role, but exact enumeration supplies the information needed to assign relative positions to each digit – and those positions define each digit's syntactic role.

In the terms used in prior work with letter strings, the first question addresses the *visual span* for digit strings (defined for letter strings as the number of letters that can be seen with a single fixation and without the help of any contextual or linguistic cues; e.g. O'Reagan, 1990). If digit strings are processed in the same way as letter strings (as has recently been suggested by Tydgat and Grainger, in press), then the visual span for digit strings should be about 10 digits (5 to the left and right of fixation if accuracies of at least 90% are required; O'Regan, Lévy-Schoen, & Jacobs, 1983). But because digit strings are so much less familiar as foci of reading than letter strings, we expected that the visual span for digits would be a somewhat smaller. The second question addresses what might be called the *perceptual span* for digit strings (defined for letter strings as the number of letters that can be identified in a single fixation; Rayner, 1998) – but in this case, the results of prior work are less clearly comparable to our current digit strings because, when letter strings form words and sentences (and even pronounceable non-words), they allow for rich top-down processing that can play no part in the processing of digit strings. In other words, we expected the perceptual span for digits to be very much smaller than the 18 or so characters (3-4 to the left, and 15-15 to the right of fixation; Rayner & Fisher, 1987; Underwood & McConkie, 1985) that subjects appear to be able to grasp with single fixations during reading.

9.1 Enumerating Digits

The structure of this experiment – a backward-masking paradigm – is analogous to the Reicher-Wheeler method (Reicher, 1969; Wheeler, 1970). In this case, the stimuli are briefly presented digit strings (with horizontal orientation), with subjects engaged in a forced-choice, two-alternative verification of each string's length. All stimuli were presented in Arabic notation, and digit strings were always composed of identical digits.

9.1.1 Method

9.1.1.1 Subjects

Sixteen naïve subjects participated; 7 men and 9 women. The youngest subject was 23 years old, and the oldest subject was 34 years old. The mean age was 25 years and 6 months. All but one of the subjects were right-handed, and all reported normal or corrected to normal vision.

9.1.1.2 Design

String length, string digit and answer digit were all manipulated; the range was the same in each case (2-8), leading to a 7x7x7 design. The exclusion of digit '0' reflects prior results suggesting that the concept of zero may emerge rather later in children than other, more concrete numerical concepts (Bialystock & Codd, 2000; Wellman & Miller, 1986), so might implicate different, cognitive processes. Digits '1' and '9' were excluded to minimise the combinatorial size of the experiment, and to keep the variance of string digits to the same range as the variance of the string lengths; single digits are not relevant to our current question, and we expected 8 digits to be sufficient to reveal the limits of the subjects' visual span. For each string stimulus digit and length, the matching probe occurred six times as often as any other single non-matching probe – ensuring that the probability of a matching trial was 0.5. We repeated each non-matching stimulus condition five times, leading to a total number of 2,940 trials[6].

9.1.1.3 Procedure

The experiment took place in a quiet, brightly lit room, with a Pentium III PC, a standard keyboard and a 15 inch colour screen. Subjects sat with their eyes approximately 50cm from the screen, but no chin rest was used.

The experiment was organised into 10 blocks, each lasting approximately 10 minutes; allowing for breaks between blocks, each experiment lasted approximately 2hrs; subjects completed 12 practice trials prior to beginning the first block, repeating them until they felt ready to proceed. The beginning of a new trial was signalled by the appearance of a dot in the centre of a blank image. This image was replaced by the "string stimulus" – a string of identical digits with horizontal orientation – after 1,000ms. The string stimulus was presented for 200ms, before being replaced by a string of eight hash marks. After a further 200ms, the probe stimulus – a single digit – was presented. The probe stimulus was removed when subjects pressed one of the two response keys, or after 2,000ms if no response was made.

Subjects had to decide if the probe stimulus was a valid report of the number of digits in the string stimulus, responding with the "up arrow" (right index finger) for matching trials, and the "down" arrow (left index finger) for non-matching trials. Feedback, in the form of a red image with a black exclamation mark at its centre, was provided whenever subjects either failed to respond

6 Six matching trials and six non-matching trials per combination of string length and string digit = 12 trials; 7 lengths and 7 digits = 49 combinations and 588 trials. 5 repetitions = 2,940 trials.

(misses) or responded incorrectly (errors). A schematic of the trial structure is displayed in Figure 37. Subjects were instructed to be accurate first, but also to try to be quick; we collected both reaction times and accuracy data.

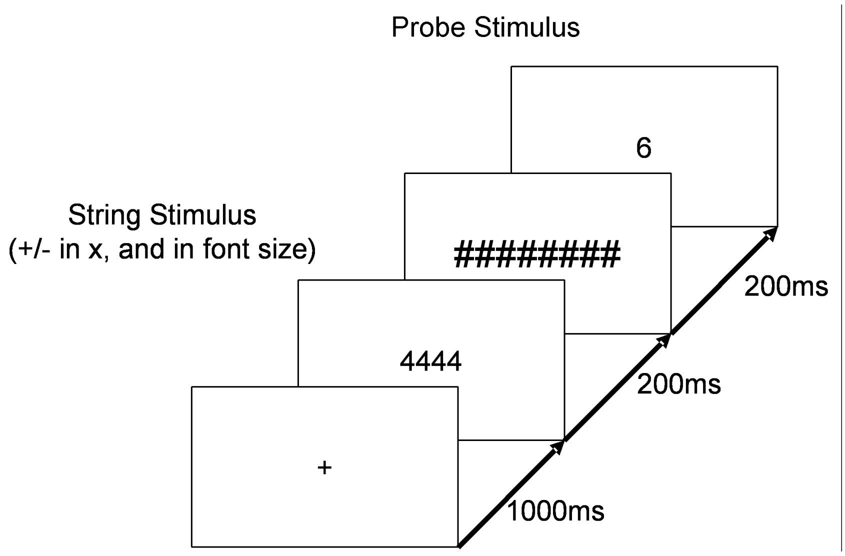


Figure 37: Schematic structure of Experiment 1 (judgement of digit string length). The example represents a non-matching trial; the matching probe in this case would be ‘4’.

With the exception of the error feedback – a black exclamation mark on a red background – all stimuli were presented as white characters (rgb = 63,63,63) on a black background (rgb = 0,0,0). All characters were presented in the Times New Roman font, with font sizes were as follows: 50 for the initial fixation cross, 60 for the mask, and 30 for the answer stimulus.

The string stimuli fonts were subject to random variation in the range 15-50, ensuring that apparent string width was not a reliable cue to actual string length; the shortest string length (two digits) in the largest font size was wider than the longest string length (eight digits) in the smallest font size. The horizontal position of the strings was also manipulated, with random adjustments of 7 pixels to the left and right of centre. The approximate visual angles of the string stimulus digits were in the range 0.8˚ (for the shortest strings in the smallest font size) to 10.1˚ (for the longest strings in the largest font size).

9.1.1.4 Data Preparation

Accuracy data are reported after excluding all “miss” trials (0.04%); analyses of variance are conducted after applying an arcsin transform ($x = 2*\arcsin(\sqrt{y})$) to remove significant deviations from a normal distribution (skewness / standard error of skewness < 2 for all subjects). Reaction time data are reported after excluding first trials with responses faster than 200ms (0.01%: thought to indicate anticipations), then all trials in which the subjects failed to respond correctly (11.7%), and then, on a subject-by-subject basis (following the recursive method of Van Selst and Jolicoeur, 1994), all trials in which the reaction times were more than three standard deviations from the mean (10 iterations required, excluding a further 4.5%).

9.1.2 Results

The questions place a greater emphasis on accuracy than on reaction time data, but this chapter reports significant effects for both data types. Accuracy rates were significantly and negatively correlated with reaction times ($r = -0.101$, $p < 0.001$), implying that there was no speed-accuracy trade-off (15 of the 16 subjects display the same correlation, while in the other, no significant correlation was observed).

The data were analysed with a repeated-measures ANOVA using string length, string digit, and probe digit as within-subjects factors. Each factor had two levels; “small” (values 2-4) and “large” (values 6-8); the complete set of results is reported in Table 6. Two of the three factors appeared to significantly influence the subjects' responses; short strings were associated with significantly faster and more accurate responses than long strings (mean RT = 540.0ms for short strings vs. 599.5ms for long strings; $F(1,15) = 34.295$, $MSE = 2054.049$, $p < 0.001$; mean error rate = 4.1% for short strings vs. 17.8% for long strings; $F(1,15) = 119.683$, $MSE = 0.021$, $p < 0.001$), and the same pattern was also observed for small vs. large probe digits (mean RT = 545.3 ms for short strings vs. 592.9 ms for long strings; $F(1,15) = 59.076$, $MSE = 407.320$, $p < 0.001$; mean error rate = 4.1% for short strings vs. 17.9% for long strings; $F(1,15) = 141.598$, MSE = 0.021, $p < 0.001$). The interaction between between string length and probe stimulus magnitude was also significant (RT: $F(1,15) = 5.477$, MSE = 901.085, $p = 0.034$; Errors: $F(1,15) = 290.399$, MSE = 0.023, $p < 0001$; see Figure 38), indicating, as expected, that subjects found trials significantly more difficult when probe digits were numerically large (> 5) and the string stimuli were long (> 5 digits).

One other effect, revealed by secondary analysis, also deserve some mention; a numerical Stroop effect (e.g. Henik & Tzelgov, 1982; Foltz, Poltrock, & Potts, 1984; Girelli, Lucangeli, & Butterworth, 2000) caused, in this case, by the task-irrelevant digits that composed string stimuli. To find it, we compared matching trials (from the whole data set) in which the string digit was either identical, or numerically adjacent, to the string length (e.g. string stimulus: “4444”, probe stimulus “4” vs. string stimulus “3333”, answer stimulus “4”). These conditions should emphasise Stroop-like facilitation and interference respectively; with this restriction in place, a second ANOVA with numerical distance (2 levels) as the sole factor confirms that a numerical Stroop effect does significantly mediate both error rates and reaction times (Mean error rate for valid trials = 10.4% vs. 32.0% for invalid trials, $F(1,15) = 96.499$, MSE = 0.026, $p < 0.001$; Mean RT for valid trials = 559.0 ms vs. 583.7 ms for invalid trials, $F(1,15) = 8.914$, MSE = 547.815, $p < 0.001$). The emergence of this effect implies that, though the subjects' sensitivities to the digits in string stimuli did not significantly mediate their overall performance, those digits were nevertheless processed semantically.

Behaviour	Source	df	Error df	MSE	F	Sig.
Accuracy	dig	1	15	.011	.808	.383
	len	1	15	0.21	119.683	.000
	pro	1	15	0.21	141.598	.000
	dig * len	1	15	.006	1.104	.310
	dig * pro	1	15	.008	.046	.834
	len * pro	1	15	.023	290.399	.000
	dig * len * pro	1	15	.007	.333	.572
RT	dig	1	15	64.108	2.801	.115
	len	1	15	2054.049	34.295	.000
	pro	1	15	407.320	59.076	.000
	dig * len	1	15	87.530	3.135	.097
	dig * pro	1	15	118.656	.103	.752
	len * pro	1	15	901.085	5.477	.034
	dig * len * pro	1	15	66.350	.727	.407

Table 6: Repeated-measures ANOVA results, tracking the mediation of subjects' accuracies and reaction times by string length (len), probe digit (pro), and string digit (dig).

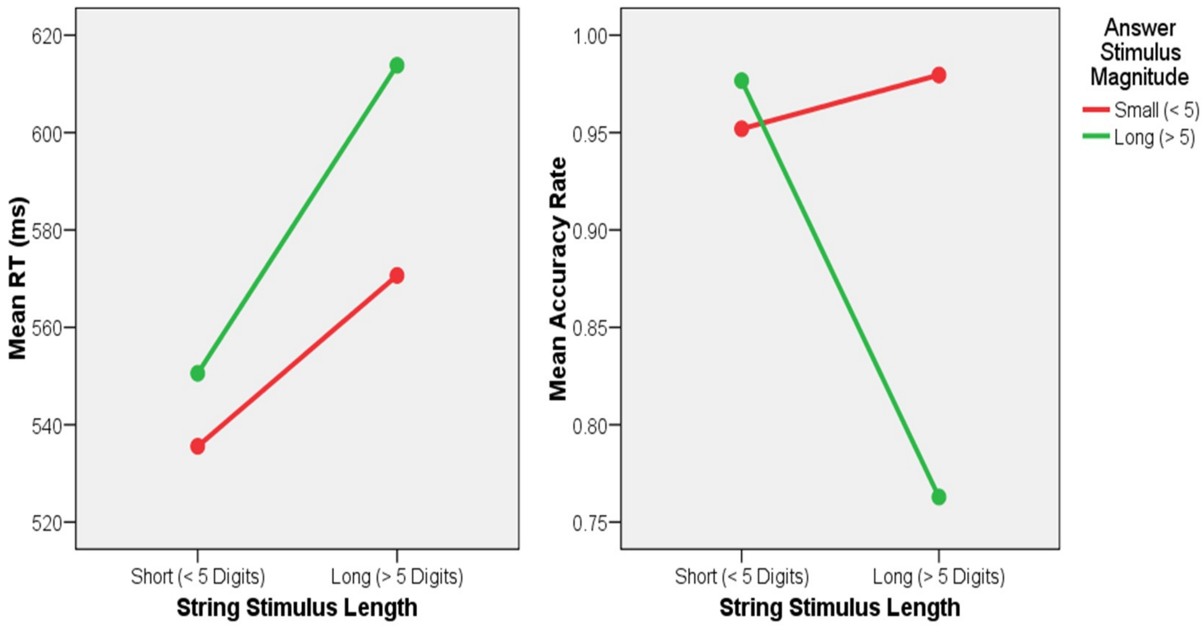


Figure 38: Interaction between string stimulus length and probe stimulus magnitude in judgements of the number of digits in a string.

9.1.2.1 How Many Digits Can Subjects See?

For each string stimulus length, the accuracy scores of subjects performing at chance levels should approximate a binomial distribution with a mean of 0.5 (the probability that a given trial is matching). Using binomial tests, we established that the error rates of every subject on every string length were significantly below that chance level (accurate responses > errors and $p < 0.001$ for all subjects and all string lengths). With single fixations, subjects seem to be sensitive to the presence of (at least) eight digits.

However, on closer inspection, this result is revealed to be somewhat over-optimistic, because it conflates exact perception with more general estimation; simply because subjects can tell you that a particular string does not have 2 digits, it does not follow that they know it has 8 digits. To establish a more credible limit, we need to restrict the analysis to a sub-set of the trials for which this kind of general estimation cannot be effective – to trials in which the probe stimuli were either identical or numerically adjacent to string stimulus lengths. To guard against response biases – for example the possibility that subjects make either "matching" or "non-matching" responses preferentially when they are unsure, we consider matching and non-matching trials separately in the analysis that follows.

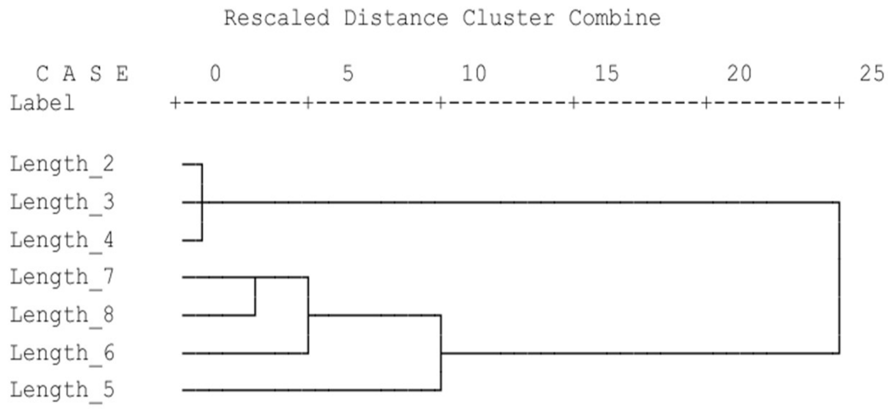
The restriction to matching trials only appears to have no significant impact on our fundamental conclusions; all subjects achieve above-chance levels of performance on all string lengths. However, the subjects are clearly subject to a bias in favour of "matching" responses when they are unsure; when only non-matching trials are considered (for which the numerical distance between probe stimulus magnitude and string length is exactly '1'), their performance clearly suffers. When considered as a group, they retain their above-chance performance on strings containing 6 digits (binomial test, mean = 0.5, Error rate = 43%, N = 1099, 2-tailed $p < 0.001$), and also – if only marginally – on strings containing 7 digits (binomial test, mean = 0.5, Error rate = 47%, N = 1099, 2-tailed $p = 0.046$). But for non-matching trials with strings of 8 digits, these subjects display significantly *below-chance* levels of performance (binomial test, mean = 0.5, Error rate = 78%, N = 549, 2-tailed $p < 0.001$).

If a subject's response bias can deflect their performance to below-chance levels in some circumstances, it makes sense not to trust their above-chance performance in others. For non-matching trials in which the numerical distance between probe digit and string length is exactly '1', no subjects fit that criterion for strings of 6 digits (and 6 subjects retain their above-chance performance in this case), 5 subjects exhibit below-chance performance for strings of 7 digits (and 3 retain their above-chance performance levels), and 14 display below-chance performance for strings of 8 digits (the other two display no significant deviation from chance). In other words, while we can be reasonably confident that the subjects could exactly perceive 6 digits, and that at least some could perceive 7, most could probably not perceive strings of 8 digits with any precision.

9.1.2.2 Subitizing vs. Enumeration?

Though the distinction is still contentious (e.g. Balakrhisnan & Ashby, 1991, 1992; Piazza, Mechelli, Butterworth, & Price 2002), a great deal of evidence suggests that the enumeration of small vs. large sets might implicate two different cognitive processes (e.g. Atkinson, Campbell, & Francis, 1976; Akin & Chase, 1978; Mandler & Shebo, 1982; Trick & Pylyshyn, 1994; Simon & Vaishnavi, 1996; Simon, Peterson, Patel, & Sathian, 1998). The subjects' ability to judge string length appears to exceed the subitizing range (usually thought to be 3 or 4 items), but their performance is still consistent with that distinction.

To capture these discontinuities, we ran hierarchical cluster analyses of the mean error rates and reaction times associated with each subject for each string stimulus length. Both analyses used values that were standardised by subject (transformed in the range 0-1), then clustered with intervals defined by squared Euclidean distance. The results confirm a clear distinction between subjects' performance on short strings (2-4 digits) versus long strings (6-8 digits); strings of length 5 are the last to be assigned (see Figure 39).



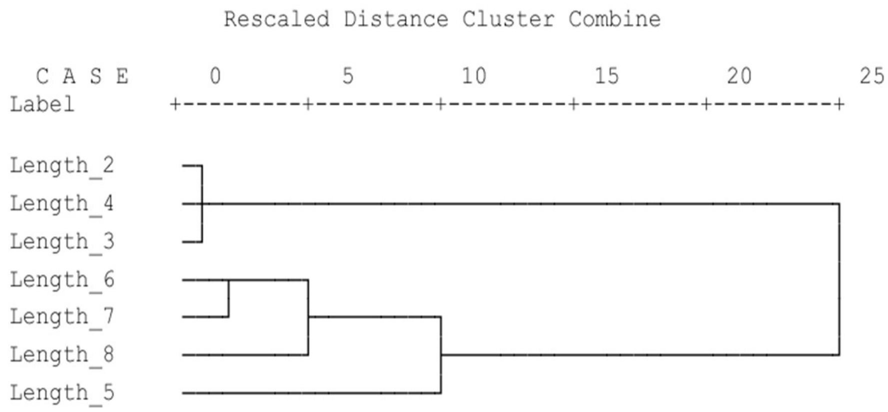


Figure 39: Dendrograms indicating the clusters to which string lengths were assigned. (Top) Error rates, (Bottom) Reaction times. In both cases, string lengths 2-4 are grouped together first, the string lengths 6-8. String length 5 is the last to be assigned, reflecting the intuition that some subjects appear to subitize strings of this length, while others appear to employ more general estimation to enumerate the digits in these strings.

9.2 Identifying Digits

The second experiment is similar to the first, but the goal in this case is to discover how many digits subjects can *identify* with single fixations. The approach is again analogous to that used by Reicher (1969) and Wheeler (1970); a forced choice, two-alternative verification of backward-masked items at different positions in strings of varying length.

9.2.1 Method

9.2.1.1 Subjects

Sixteen naïve subjects participated; 9 men and 7 women. The youngest subject was 22 years old, and the oldest subject was 45 years old. The mean age was 27 years. All but one of the subjects were right-handed, and all reported normal or corrected to normal vision.

9.2.1.2 Design, Stimuli, and Procedure

The structure of this experiment was very similar to that described previously, but in this case, the subjects' task was to decide if the probe stimulus had occurred in the string stimulus (now a string of different digits). String stimulus length was manipulated, and lengths in the range 3-6 were considered; strings were defined by selecting digits at random from the range 1-9, without repetition. Each string length occurred with equal probability, and matching probe stimuli were drawn from each string position at random, but also with equal probability. The probability of a matching trial was set to 0.5. As in the previous experiment, both reaction times and accuracy data were recorded, with subjects instructed to prefer accuracy over speed. Subjects were exposed to 750 trials, organised into 3 blocks, with a practice block of 12 trials. Allowing for breaks between blocks, the experiment lasted approximately 30min.

Visual presentation conditions were identical to those used in the previous experiment, with the exception that the widest and narrowest viewing angles of string stimuli were changed in line with the smaller range of string lengths under consideration. The shortest string length was '3', corresponding to a minimum viewing angle of ~1.4°, and the longest string length was '6', corresponding to a maximum viewing angle of ~9.4°.

The data were also prepared as before; accuracy data exclude only "missed" trials (0.2%), and are subjected to an arcsin transform (skewness / standard error of skewness < 2 for all subjects). Reaction time data exclude all inaccurate trials (18.5%), and following the recursive method used previously (Van Selst & Jolicoeur, 1994), all trials more than 3 standard deviation from the mean (6 iterations, excluding a further 2.6%).

9.2.2 Results

No significant speed-accuracy trade-off was observed; accuracy rates were significantly and negatively correlated to reaction times ($r = -0.103$, $p < 0.001$; this group-wide result is also reflected at the level of particular subjects, 14 of whom display the same negative correlation, and none of whom display a positive correlation). We considered two factors in the analysis: string length and string position. The first of these two corresponds to a factor used in Experiment 1 – but in this case, string lengths 3-4 and 5-6 were grouped together. String position refers to the side (left half vs. right half) in which matching probe stimuli appeared in the string stimulus (this analysis is restricted to matching trials only) In strings of odd length (3 and 5), the middle position was excluded.

A repeated measures ANOVA, confirms that both accuracy and reaction time data were significantly mediated by string length (see Table 7 for the complete set of results). For short strings, the mean error rate is 13.8%, and the mean response latency is 745.0 ms, while for long strings, the mean error rate is 29.5% and the mean response latency is 788.6 ms; as expected, longer strings make for more difficult trials. The string position factor also exerts a main effect on both sets of data; digits in the left of the strings (mean error rate = 17.7%, mean RT = 704.8 ms) appear to be significantly easier to identify than digits on the right side of the strings (mean error rate = 26.7%, mean RT = 788.6 ms). For reaction times only, we also observed an interaction between string position and string length; right-sided digits appear more difficult to identify in longer strings (mean RT = 762.7 ms) than they are in shorter strings (mean RT = 705.5).

The significance of string length here is consistent with the results of the first experiment; if subjects find it harder to enumerate the digits in longer strings, they should also find it harder to identify the digits in those longer strings. The significance of the "string position" factor is consistent with the other Reicher-Wheeler type experiments on letter-string perception (e.g. Wheeler, 1970; Tydgat & Grainger, in press), with centrally presented strings. This effect might be an artefact of that central presentation, but the leading figures in a multi-digit string contribute the most to the number's magnitude, so it seems natural that subjects should process these digits preferentially.

Dependent	Source	df	Error df	MSE	F	Sig.
Accuracy	pos	1	15	.050	60.054	.003
	len	1	15	.056	12.101	<.001
	pos * len	1	15	.039	2.520	.133
RT	pos	1	15	1408.948	12.712	.003
	len	1	15	2061.873	12.168	.003
	pos * len	1	15	759.196	9.587	.007

Table 7: Repeated-measures ANOVA results, tracking the effect of digit position (pos) and string length (len) on the subjects' accuracy rates and reaction times.

9.2.2.1 How Many Digits Can Subjects Identify?

As in the previous experiment, the probability that a given trial is matching, is 0.5. For each position of each string stimulus length, each subject's performance was compared to a binomial distribution with a mean of 0.5. The result was a tally, by subject and string length, of the string positions for which each subject was able to identify digits at levels that were significantly above chance. The maximum number of digits that subjects were able to identify corresponds to the maximum number of positions in which above-chance levels of performance were achieved. These maxima are overwhelmingly in the range 3-4 (see Figure 40); only one subject was able to identify 5 digits, and no subjects could identify six digits.

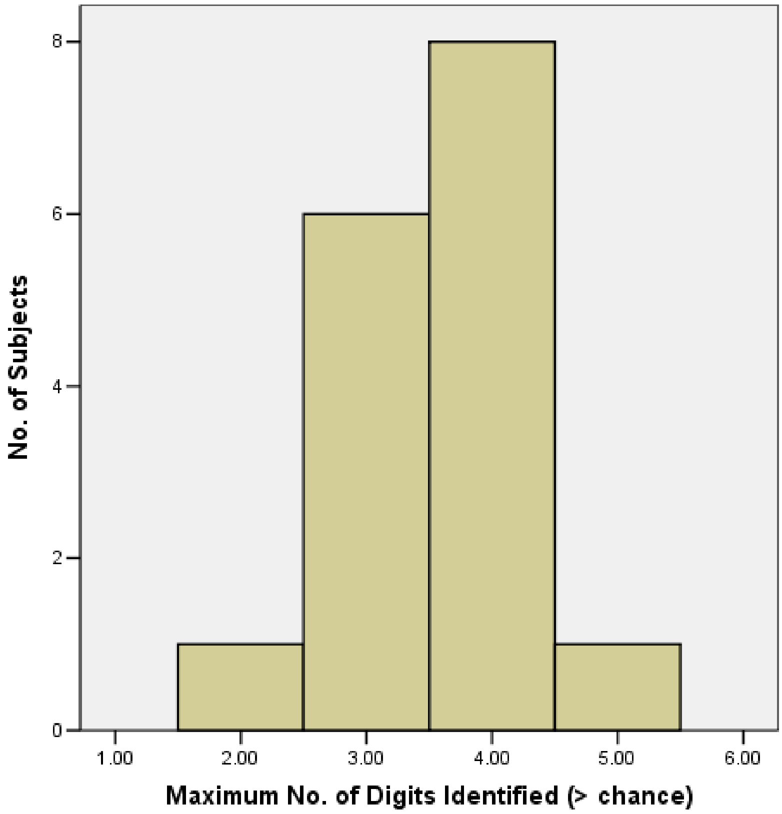


Figure 40: A histogram illustrating the maximum range of above-chance digit identification, by subject. The results are based on binomial tests, comparing each subjects' accuracy data to a binomial distribution with a mean of 0.5

The subjects' performance tends to be degraded on strings that exceed their maximum range of identification; all but one of the subjects achieve their best performance by successfully identifying digits at every position in strings of 3 or 4 digits. One feature that makes longer strings more difficult is lateral masking – typically invoked (e.g. Grainger & Holcombe, in press) to explain the emergence of the classical W-shaped performance curve (see Figure 41) in Reicher-Wheeler experiments with centrally presented strings (e.g. Hammond & Green, 1982, Lefton, Fisher, & Kuhn, 1978; Mason, 1982; Stevens & Grainger, 2003). To try to capture the significance of lateral masking in degrading the subjects' performance, we can conduct a further analysis designed to exclude it.

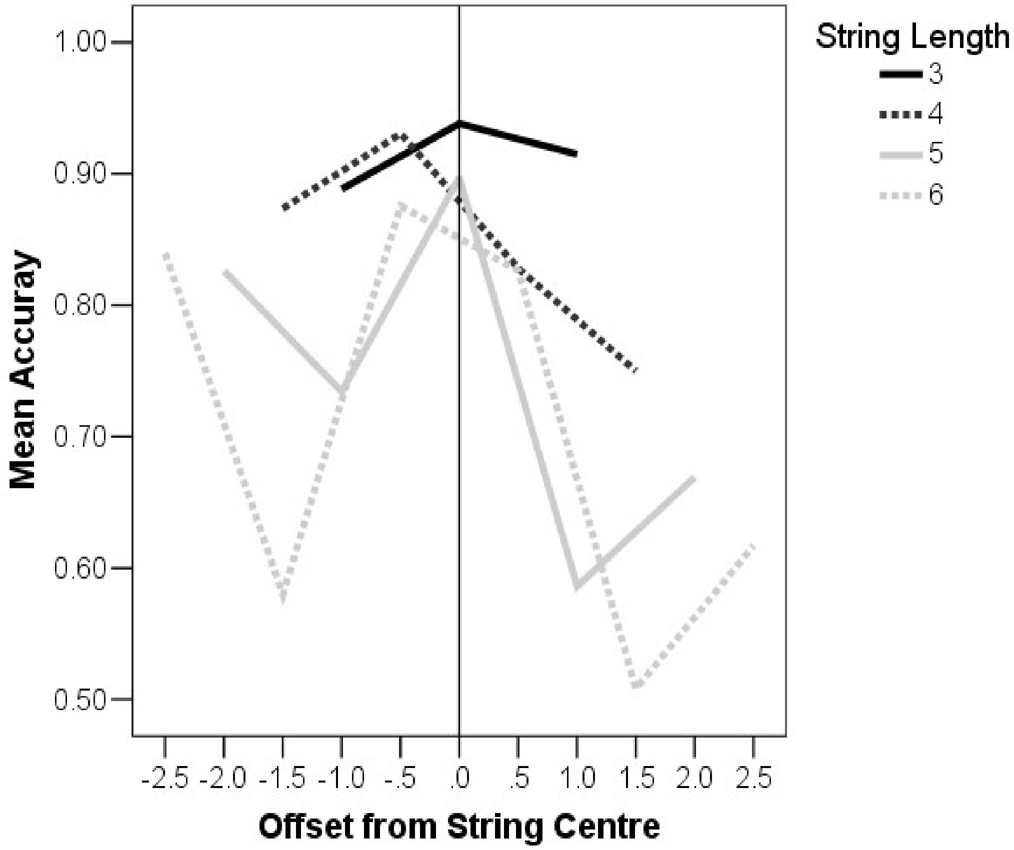


Figure 41: Subjects' accuracy by string position and string length; offset positions are expressed in terms of the number of digits from the string centre (strings of even length require half-digit distances).

The W-shape emerges because lateral masking has a reduced effect on leading and trailing digits, as well as digits at fixation (in this case, in central positions). To capture the role that string length plays independently of this masking, the analysis was restricted to those string positions that were least affected – to the positions that permitted the best performance for each string length. In order of increasing string length, these were positions 2, 2, 3 and 3 (i.e. central positions for strings of odd length, and just left of centre for strings of even length). We can then perform a repeated measures ANOVA tracking the impact of string length (2 levels) on the subjects' performance in these restricted cases. The results suggest that string length does not significantly mediate the subjects' performance when lateral masking is excluded. The implication is that – in this case at least – lateral masking is the principal source of degraded performance on longer digit strings.

9.3 Interim Discussion

The results suggest that, with single fixations, subjects can see up to 6 or 7 digits, but reliably identify only 3 or 4. From Experiment 1, the visual span for digits appears to be rather smaller than that for letters – this result seems inconsistent with the assertion that letters and numbers implicate the same "alphanumeric" perceptual processing architecture (Tydgat & Grainger, in press). The difference may be a function of familiarity; if letter and digit detectors can be adapted to processing dense strings (at least relative to other symbols; Tydgat & Grainger, in press), then that adaptation may be dependent on the subjects' experience of reading (which might favour letters over digits). The results of Experiment 2 are consistent with those of Tydgat and Grainger (in press); with strings of 5 digits, they observed a "W-shaped" position function for accuracies that is very similar to that displayed in Figure 41. My explanation for the effect – based on lateral masking (or crowding) – is also consistent with their interpretation. Indeed, since all but one subject achieved their best performance by identifying digits at every position in strings of at most 4 digits, it seems reasonable to suggest that this masking is a key limiting factor on their perceptual span for digits.

Alongside the main results, it is also clear that there are discontinuities in subjects' digit enumeration performance (Experiment 1), consistent with claims that the enumeration of small vs. large sets implicates two different processes. Further, the range in which subitizing occurs is similar to the range in which subjects can identify digits (Experiment 2), consistent with the intuition that the subitizing range is connected to subjects' capacity for opening "object files" (e.g. Carey 1998) – that, with single fixations, subjects can identify precisely as many items as they can subitize. The current work lacks the within-subject data that would be required to claim strong conclusions of this sort – but this weakness could be rectified in future work.

Though clearly rather removed from natural number processing, the tasks employed here have nevertheless provided a critical foundation for the work that must be done to understand how very large numbers are perceived; to interpret saccadic strategies in more natural number processing tasks, one must first identify the constraints that (visual) perception imposes on them. These data suggest that the most efficient saccadic strategy for digit strings processing is to divide them into discrete “chunks” (of 3 or 4 digits) that can then be processed in parallel – but this conclusion assumes that above-chance recognition performance is sufficient for everyday number processing tasks. Observed deviations from that strategy can therefore tell us something about the – probably rather more stringent, and possibly task-dependent – level of perceptual confidence that subjects actually require.

Following the logic of prior work on reading, a natural next step might be to assess the interaction between the constraints observed here and the details of the problem that subjects must solve. From the current results, we can be reasonably sure that at least some digits in both experiments were processed semantically; numerical Stroop effects emerged in Experiment 1, while a W-shaped position function emerged in Experiment 2 (consistent with that observed for letter and digit strings by Tydgat and Grainger, in press, but not with that observed for strings containing other symbols). However, the current tasks did not force the subjects to process digit strings as *integrated* numbers; the latter goal might be achieve by asking them to identify what digit played a specified syntactic role (e.g. tens digit, hundreds digit, and so on) in visually presented digit strings. Another approach to confirming these constraints' generality would be to use a paradigm that gives subjects more freedom to move their eyes, perhaps by employing the *moving window* technique (McConkie & Rayner, 1975; Reder, 1973) to selectively control the visual information available to subjects while they manipulate long digit strings. For digit string reading tasks, the prediction would be that subjects need only be able to see 3 or 4 digits clearly during each fixation to recognise the numbers that these strings represent.

IV

Summary, Discussion and Conclusions

Chapter 10

Summary, Discussion and Conclusions

10.1 Summary

We began with a series of chapters that introduced (chapter 3), developed (chapter 4) and applied (chapter 5) a Dynamicist model-building method to the problem of number comparison. From the perspective of numerical cognition, the key result was the discovery of a novel format for semantic number knowledge – the *single-unit accumulator code* – in agents that had been "evolved" to forage effectively (chapter 5). But the work also makes the much more general point that, perhaps contrary to expectations, Dynamicism *can* be employed to create models that interact with conventional debates on the structure of cognitive representation. Chapter 6 shifted the focus away from representations and toward the processes that employ them – describing another novel model-building method which employs the assumption that neural information processing will be optimal or near optimal for any specific task. Like chapter 5, chapter 6 made a methodological contribution by demonstrating its method's utility. And like chapter 5, chapter 6 also provided results of scientific interest – yielding a model that captures both reaction time and accuracy data, as well as the neural dynamics that primates seem to employ.

Chapters 7 to 9 all considered problems associated with the processing of larger, multi-digit numbers. Chapter 7 presented an experiment that dissociates subjects' sensitivities to these numbers' integrated values from their sensitivity to the numbers' single-digit components – and also revealed an apparent dissociation between the Size and Distance effects, which is not consistent with the predictions of three of the four popular theories of MNL format (the *compressed MNL*, the *noisy MNL*, and the *numerosity code* all imply that both effects stem from the same causal source:

representation-level interference). Chapter 8 presented a model-space for multi-digit number comparison, which confirmed that model quality is mediated by the interaction between MNL formats and MNL structures, so implies an interaction between the format and structure debates. The analysis also confirmed the plausibility of the *single unit accumulator code*; the best, discovered models employed this format with a *decomposed* structure. Finally, chapter 9 reported two experiments designed to capture some constraints on the perceptual processing of very large numbers. The results imply that, with single fixations, human subjects are sensitive to the presence of up to eight digits, can exactly enumerate up to 6 or 7, but reliably identify only 3 or 4.

10.2 General Discussion

This thesis has reported two kinds of work; computational models in chapters 4-6 and 8, and more direct, empirical experiments in chapters 7 and 9. The comparison between these two groups can be instructive. As compared to the computational work, the key advantage of the experiments is their concreteness – though their interpretation can be criticised, the data that these chapters report is much harder to dismiss. That concreteness is much more difficult for computational models to achieve because of the sense that they are simply too arbitrary to be trusted – because there may be infinitely many distinct ways to implement the same, cognitive behaviour.

To a large extent, that distinction can explain the overwhelming preference for experiments over computational models in cognitive science. Since our knowledge of the neurophysiology underlying cognition is still too coarse to verify the details of most computational cognitive models, the preference reflects a sense that these models might be rather premature. To some extent, that doubt can – and should – be addressed by empirical experiments; chapters 7 and 9 reflect that need by contributing new data to that can inform our knowledge of numerical cognition. But the criticism also implies a rather limited stance on the role that computational models can play in cognitive science – specifically, that they must implement their designers' favoured cognitive theories. None of the models that this thesis reports were designed to play that role.

The clearest connection between the computational projects in this thesis is their dual focus. Each addresses a problem of relevance to numerical cognition, but each also introduces a new methodology for model-building. And though different in detail, all of these methods embodies the common intuition that model-building can – and should – answer empirical questions. Cast in that light, the computational models that this thesis reports are just as concrete – just as “experimental” – as the experiments themselves; the difference is simply that the different kinds of work ask different kinds of question.

In chapter 5, the question is: what number representations might selective evolution encourage in simple organisms? The model-building method attempts to answer it by “evolving” simple organisms in an artificial ecosystem, with an ecologically plausible definition of evolutionary fitness. The result's utility demonstrates that this kind of question can be answered – or at least that sensible answers can be proposed – by computational means. And by taking as many constraints as possible away from the designer's explicit control, the Dynamicist method also minimises the contingent, arbitrary content of the answers that it can provide.

The same logic – motivating a preference for “minimal” methodology – is also employed by chapter 6. In this case, the question is: what neural network architecture implements an optimal (or near-optimal) information processing strategy for making decisions under sensory uncertainty? As in chapter 5, the results are more believable because so few of the models' parameters are explicitly defined. But in this case, much more explicit justifications are also available, because the search-based approach directly reflects the popular assumption that the models' referent (the relevant, biological architecture) is at least close to optimal, and because the dynamics of that referent have already been so well-reported. From the latter, we could confirm that the resulting models did actually reflect the biology that they were intended to capture. And though the model-building method does not (cannot) guarantee optimality, its deviations from optimality also appear to be reflected in the biology it was intended to capture.

Chapter 8 was designed to answer the question: what combinations of MNL format and structure can drive effective, associative models of multi-digit number comparison? The approach in this case is not so much minimal as it is combinatorial – but the results are no less empirical for that distinction. Like chapters 5 and 6, chapter 8 takes the focus of its question away from the designer, and like those two previous chapters, this method's utility is confirmed by the relevance of the results that it supplies – results that are particularly interesting because they favour the novel MNL format (the *single unit accumulator code*) that was discovered in chapter 5. The question that this chapter asks has a limited scope, because its results might not be robust to different learning systems, and because associative memory may simply be the wrong approach for modelling multi-digit number comparison. But within its restricted scope, the answers that this chapter supplies have a definite, empirical character.

In all of the computational chapters, this empirical content carries an analytical cost. That cost is most explicit in chapters 4 and 5, which introduced extensive, novel methodology (the BMM) that was specifically designed to address it. The analysis in chapter 6 is simplified because the models that result are less structurally complex – but they still illustrate the same essential problem; when the structural specification of a model is automated, rather than explicitly designer-driven, there is no guarantee that we will understand how the finished result actually works. Chapter 8 illustrates a different kind of cost, because its models are all designed by a much more familiar, much more constrained method. But the results' complexity – the sheer number of models that they include – makes a detailed analysis quite difficult to complete.

That limitation – and the work required to address it – can justify a preference for the more restricted role that computational models are more commonly employed to play. When models explicitly encode their designers' preferred cognitive theories, their implementation implies prior knowledge of their functional architecture. That prior knowledge ensures that the results are at least usually clear, even if their scope might be restricted. This thesis has been motivated by the belief that computational models can, and should, play a much richer role in cognitive science. Chapters 3-5, 6, and 8 have proposed three routes toward that goal. Each approach can be criticised, but each also yield conclusions of cognitive interest – conclusions that could not be made by more conventional work. The *thesis* of this thesis is that this benefit outweighs any of the associated costs.

10.3 Conclusions

Brains are fundamentally machines, so any even remotely believable account of cognition must be expressed at the level of causal / functional architecture. Computational models are the most natural way to express cognitive theory at that level of detail, but since our knowledge of the relevant neurophysiology is still too coarse to verify most of those models directly, they can often seem rather premature. Both the minimal and combinatorial approaches to model-building can be construed as attempts to find a way to begin to *trust* the models that we design – by making their unverified content the focus of structured, empirical investigation. And in that respect, this work appears to be successful. Like the more directly empirical experiments that this thesis reports, the interpretation of the computational work can clearly be criticised – but the data itself, the emergence of *this* model architecture, or *that* model behaviour, under *those* circumstances, is much more concrete.

The model-building methods that this thesis reports are also valuable because they have the capacity to *surprise* – to produce unexpected answers to questions of cognitive interest. Chapter 3 illustrated that, despite over 50 years of research, many of the field's most basic foundations are still extremely contentious. In some cases at least, our natural – and even informed – intuitions about how cognition works are both difficult to ignore and quite probably misleading. To the extent that computational methods can exceed our expectations, they are also released from the constraints those expectations impose.

Perhaps the biggest surprise of the experience that this thesis reports is that very few people – very few of my own colleagues, or of the other academics in the field – appear to believe that a complete, causal-level understanding of cognition can ever be achieved. But I *do* believe it – more strongly now than ever before. Whatever else may be required, it seems clear that computational methods must play some role in achieving that goal. The methodological content of this thesis reflects a sense that many of the obstacles in the path toward it stem from a poverty of perspective – that if the problems seem too difficult to solve, it is only because we are approaching them in the wrong way. This thesis has proposed several new computational methods, one or more of which may bring us closer to that hypothetically “right” approach. Only time – and further work – can tell us how successful they can be.

References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A Learning Algorithm for Boltzmann Machines, *Cognitive Science*, *9*, 147-169.

Aharanov, R., Segev, L., Meilijson, I., & Ruppin, E. (2003). Localization of function via lesion analysis. *Neural Computation*, *14*(4), 885-913.

Ahmad, K., Casey, M. C., & Bale, T. (2002). Connectionist Simulation of Quantification Skills. *Connection Science*, *14*(3), 165-201.

Akin, O. & Chase, W. (1978). Quantification of three-dimensional structures. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 397–410.

Aleksander, I. (1989). *Neural Computing Architectures: the design of brain-like machines*. Kogan Page.

Ashcraft, M. H. (1992). Cognitive arithmetic: a review of data and theory. *Cognitio*n, *4*4, 75-106.

Ashcraft, M., & Christy, K. (1995). The frequency of arithmetical facts in elementary texts: addition and multiplication in grades 1–6. *Journal for Research in Mathematical Educatio*n, *2*6, 396-421.

Atkinson, J., Campbell, F., & Francis, M. R. (1976). The magic number 4 ± 0: A new look at visual numerosity judgements. *Perception*, *5*, 327–334.

Bächthold, D., Baumüller, M., & Brugger, P. (1998). Stimulus-response compatibility in representational space. *Neuropsychologia*, *3*6, 731–735.

Balakrishnan, J. D. & Ashby, F. G. (1991). Is subitizing a unique numerical ability? *Perceptual Psychophysics*, *50*(6), 555–564.

Balakrishnan, J. D. & Ashby, F. G. (1992). Subitizing: Magical numbers or mere superstition?

Psychological Review, *54*, 80–90

Bakker, B. (2000). The adaptive behavior approach to psychology. *Cognitive Processing*, 1, 39-70.

Bechtel, W. & Abrahamsen, A.(1991). *Connectionism and the Mind* (chapter 6) Oxford: Blackwell.

Beer, R. D. (1996). Towards the Evolution of Dynamical Neural Networks for Minimally Cognitive Behavior. In P. Maes, M. Mataric, J. Meyer, J. Pollack & S. Wilson (Eds.), *From animals to animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior* (p. 421-429). MIT Press.

Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences, 4*(3), 91-99.

Beer, R. D. (2003). The Dynamics of Active Categorical Perception in an Evolved Model Agent. *Adaptive Behavior, 11*, 209-243.

Berkeley, I., S., N. (1997). *Some Myths of Connectionism*, http://www.ucs.louisiana.edu/~isb9112/dept/phil341/myths/myths.html.

Bialystok, E. & Codd, J. (2000). Representing quantity beyond whole numbers: Some, none and part. *Canadian Journal of Experimental Psychology*, *54*, 117-128.

Biro, D., & Matsuzawa, T. (2001). Use of numerical symbols by the chimpanzee (Pan troglodytes): Cardinals, ordinals, and the introduction of zero. *Animal Cognition*, *4*, 193-199.

Blumberg, B. (1995). Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments. Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, 47-54, ACM: New York.

Bonato, M., Fabbri, S., Umiltà, C., & Zorzi, M. (2007). The Mental Representation of Numerical Fractions: Real or Integer? *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1410-1419.

Botvinick, M. & Plaut, D. C. (2004). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113*, 201-233.

Brannon, E.M., & Terrace, H.S. (2003). The evolution and ontogeny of ordinal numerical ability. In M. Beckoff (Ed.), *The Cognitive Animal*, 197–204, MIT Press.

Brannon, E., Wusthoff, C., Gallistel, C. R., & Gibon, J. (2001). Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychological Scienc*e, *1*2, 238-243.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139-159.

Butterworth, B. (1999). *The Mathematical Brain*, Macmillan

Butterworth, B. (2005). Developmental Dyscalculia. In J. I. D. Campbell (Ed.) *The Handbook of Mathematical Cognition*, Psychology Press, 455-467.

Butterworth, B., Zorzi, M., Girelli, L., & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *Quarterly Journal of Experimental Psychology*, *54A*, 1005-1029.

Bynner, J. & Parsons, S. (1997). *Does Numeracy Matter? Evidence from the National Child Development Study on the Impact of Poor Numeracy on Adult Life.* Report to the UK Basic Skills Agency, London.

Campbell, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, *1*, 121-164.

Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General, 130*, 299-315.

Cantlon, J. & Brannon, E.M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, *17*(5), 401-406.

Cappelletti, M., Butterworth, B., & Kopelman, M. (2001). Spared numerical abilities in a case of semantic dementia. *Neuropsychologia*, *39*, 1224-1239.

Cappelletti, M., Kopelman, M., & Butterworth, B. (2002). Why semantic dementia drives you the dogs (but not to the horses): A theoretical account. *Cognitive Neuropsychology*, *19*(6), 483–503.

Carey, S. (1998). Knowledge of Number: its Evolution and Ontogeny. *Science*, *282*, 641-642

Carey, S. (2004). *On the Origin of Concepts*, Daedulus.

Chomsky, N. (1959). Review of Verbal Behavior by B.F. Skinner. *Language*, *35*, 26-58.

Ciaramitaro, V. M. & Glimcher, P. W. (2001). Stimulus probability redirects spatial attention: an enhancement of sensitivity in humans and monkeys. *Vision Research*, *41*, 57-75.

Cipolotti, L., Butterworth, B., & Denes, G. (1991). A specific deficit for numbers in a case of dense acalculia. *Brain*, *114*, 2619-2637.

Clark, A. (1999). Embodied, situated, and distributed cognition. In W. Betchel and G. Graham (Eds), *A Companion to Cognitive Science*, Malden, MA: Blackwell Publishing.

Cockcroft, W. H. (1986). *Mathematics Counts.* London: Her Majesty's Stationery Office.

Colby C. L., Duhamel, J. R., & Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of Neurophysiology, 76*(5), 2841-2852.

Cooper, R. & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology, 17,* 297–338.

Cooper, R., & Shallice, T. (2006). Structured representations in the control of behavior cannot be so easily dismissed: A reply to Botvinick and Plaut. *Psychological Review, 113*, 887-916.

Crick, F. H. C. (1989). The current excitement about neural networks. *Nature*, *337,* 129-132.

Crowther-Heyk, H. (1999). George Miller, language, and the computer metaphor. *History of Psychology, 2*, 37-64.

Crutchfield, J. P. (1998). Dynamical embodiments of computation in cognitive processes, *Behavioral and Brain Sciences*, *21*(5), 635-637.

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*, Oxford University Press

Dehaene, S. (2001) Subtracting pigeons: Logarithmic or linear ? *Psychological Science, 1*2, 244-246.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends in Cognitive Neurosciences, 7*(4), 145-147.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*, 371–396.

Dehaene, S. & Changeux, J.P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal Cognitive Neuroscience*, *5*, 390-407.

Dahaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in Cognitive Science*, *9*(7), 335-341.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital ? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 626-641

Dehaene, S, Piazza, M, Pinel, P, & Cohen, L. (2003). Three parietal circuits for number processing. Cognitive Neuropsycholy 2003, 20, 487-506

Den Heyer, K., & Briand, K. (1986). Priming single digit numbers: Automatic spreading activation dissipates as a function of semantic distance. *American Journal of Psycholog*y, 99, 315-340.

DfES. (2001). *Guidance to support pupils with dyslexia and dyscalculia (No. DfES 0512/2001)*. London: Department of Education and Skills.

Di Luca, S., Grana, A., Semenza, C., Seron, X., & Pesenti, M. (2006). Finger-digit compatibility in Arabic numeral processing. *Quarterly Journal of Experimental Psychology*, *5*9(9), 1648-1663.

Durgin, F. H. (1995). Texture Density Adaptation and the Perceived Numerosity and Distribution of Texture, *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 149-169.

Eliasmith, C. (1996). The third contender: A critical examination of the dynamicist theory of cognition. *Journal of Philosophical Psychology*, *9*(4), 441-463

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7*(2/3), 195-226.

Feigenson, L., Carey, S., & Hauser, M. (2002). Infants spontaneous representations of more and less. *Psychological Science*, *12*(2), 150-156.

Feldman, A. G., (1966). Functional Tuning of the Nervous System with Control of Movement of Maintenance of a Steady Posture of Movement or Maintenance of a Steady Posture-II. Controllable Parameters of the Muscles. *Biofizika*, *11*, 498-508.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist Models and their Properties. *Cognitive Science 6*, 205-254.

Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognitio*n, *2*, 95-110.

Fias, W., & Fischer, M.H. (2005). Spatial Representation of Numbers. In J.I.D. Campbell (ed.) *Handbook of Mathematical Cognition*. Hove: Psychology Press.

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

Foltz, G. S., Poltrock, S. E. & Potts, G. R. (1984). Mental comparisons of size and magnitude. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 442–453

Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., & Evans, A.C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, *1*, 214-220.

Funahashi S., Bruce C. J., & Goldman-Rakic P. S. (1993). Dorsolateral prefrontal lesions, and oculomotor delayed-response performance: evidence for mnemonic "scotomas." Journal of Neuroscience, 13, 1479-1497.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43-74.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*, 59-65.

Gallistel, C. R., & Gelman, R. (2005). Mathematical Cognition. In K. Holyoak & R. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*, 559-588, Cambridge University Press.

Ganor-Stern, D., Tzelgov, J., & Ellenbogen, R. (2007). Automaticity of two-digit numbers. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 483-496.

Gazzellini, S. & Laudanna, A. (2005). Are Two-Digit Numerals Decomposed or Holstically Represented? Evidence from Number Comparison. Abstract for a poster at the *27th Annual Conference of the Cognitive Science Society*, available at http://www.psych.unito.it/csc/cogsci05/post_sessions.html.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discrimination. *Psychological* Review, *96*, 267-314.

Gelman, R. & Butterworth, B. (2005). Number and language: how are they related? *Trends in Cognitive Neurosciences*, *9*(1), 6-10.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, *87*(3), B87-95.

Gevers, W., Reynvoet, B., & Fias, W. (2004). The mental representation of ordinal sequences is spatially organized: evidence from days of the week. *Cortex*, *40*(1), 171-172.

Girelli, L., Lucangeli, D. & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology*, *76*, 104–122.

Gnadt, J. W. & Mays, L. E. (1995). Neurons in monkey parietal area LIP are tuned for eye movement parameters in three-dimensional space. *Journal of Neurophysiology*, *73*(1), 280-297.

Gold, J. I. & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, 404, 390 – 394.

Gold, J. I. & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. Trends in Cognitive Science, 5, 10-16.

Gold, J. I. & Shadlen, M. N. (2003). The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *Journal of Neuroscience*, *23*, 632-651.

Goldberg D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison Wesley Publishing Company.

Gonzalez, R. C. & and Richard, E. W. (1992). *Digital image processing*. Addison Wessley Publishing Company.

Grainger, J. & Holcomb, P. J. (in press). Neural constraints on a functional architecture for word recognition. In P. Cornelissen, P. Hansen, M. Kringelbach & K. Pugh (Eds.), *The neural basis of reading*. Oxford University Press: Oxford

Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition, *Psychological Review*, *79*, 329-343.

Gross-Tsur, V., Manor, O., & Shalev, R. S. (1996). Development dyscalculia: Prevalence and demographic features. *Developmental Medicine and Child Neurology*, *38*, 25-33.

Grossberg, S. & Repin, D. (2003). A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes. *Neural Networks*, *16*, 1107-1140.

Hammond, E. J. & Green, D. W. (1982). Detecting targets in letter and non-letter arrays. *Canadian Journal of Psychology*, *36*, 67-82

Harvey, I. (1996). Untimed and misrepresented: connectionism and the computer metaphor. *Newsletter of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB Quarterly), 96*, 20-27.

Harvey, I. (2001). Artificial Evolution: A Continuing SAGA. In G. Takashi (Ed.) *Evolutionary Robotics: From Intelligent Robots to Artificial Life, Proceedings of the 8th International Symposium on Evolutionary Robotics*. Springer-Verlag Lecture Notes in Computer Science LNCS 2217.

Harvey, I., Husbands, P., Cliff, D., Thompson, A., & Jakobi, N. (1996). Evolutionary Robotics at Sussex. In *Robotics and Manufacturing: Recent trends in research and applications (Proc. World Automation Conf. WAC'96)*, p 293-298. New York: ASME Press.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory and Cognition*, *10*, 389–395

Henschen, S. E. (1920) *Klinische und Anatomische Beitrage zu Pathologie des Gehirns*, Nordiska Bokhandeln.

Hodgkin, A. & Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology, 117,* 500-544.

Honavar, V. & Uhr, L. (1989). Brain-structured connectionist networks that perceive and learn. *Connection Science, 1*(2), 139-159.

Joanisse, M. & Seidenberg, M. S. (1999). Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Sciences, 96*, 7592-7597.

Jordan, K. & Brannon E.M. (2006a). A common representational system governed by Weber's Law: Nonverbal numerical similarity judgments in 6-year-olds and rhesus macaques. *Journal of Experimental Child Psychology*, *95*, 215-229 .

Jordan, K. & Brannon, E.M. (2006b) The influence of Weber's law on the numerical representations of rhesus macaques (Macaca mulatta). *Animal Cognition*, *9*, 159-172.

Johnson-Laird, P. (1988). *The Computer and the Mind.* Harvard University Press.

Keinan, A., Sandbank, B., Hilgetag, C. C., Meilijson, I., & Ruppin, E. (2004a). Fair attribution of functional contribution in artificial and biological networks. *Neural Computation, 16*(9), 1887-1915.

Keinan, A., Hilgetag, C. C., Meilijson, I., & Ruppin, E. (2004b). Causal localization of neural function: The Shapley value method, *Neurcomputing*, *58-60C*, 215-222.

Kelso, J.A.S. (1995). *Dynamic Patterns*, MIT Press.

Kim, J-N & Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience*, *2*(2), 176-185.

Koechlin, E., Naccache, L., Block, E., & Dehaene, S. (1999). Primed numbers: Exploring the modularity of numerical representaions with masked and unmasked priming. *Journal of Experimental Psychology: Human Perception and Performance, 25*,1882-1905.

Kohonen, T. (1995). *Self-Organizing Maps*, Berlin: Springer-Verlag.

Korvorst, M., & Damian, M. F. (2007). The differential influence of decades and units on multi-digit number comparison. *Quarterly Journal of Experimental Psychology*, available at http://www.informaworld.com/smpp/content~content=a782895981~db=all~order=author.

Lefton, L. A., Fisher, D. F. & Kuhn, D. M. (1978). Left-to-right processing of alphabetic material is independent of retinal location. *Bulletin of the Psychonomic Society*, *12*, 171-174

LeFevre, J., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 216-230.

Lengyel M., Kwag J., Paulsen O., & Dayan P. (2006). Matching storage and recall: hippocampal spike timing-dependent plasticity and phase response curves. *Nature Neuroscience*, *8*, 1677-1683.

Lewis, C., Hitch, G., & Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9-and 10-year-old boys and girls. J*ournal of Child Psychology and Psychiatry*, *35*, 283 –292.

Lewin, R. (1992). *Complexity: Life at the edge of chao*s. Macmillan.

Lorch, R. F. Jr. & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. Journal of Experimental Psychology: Learning, Memory, & *Cognition*, *16*(1), 149-157.

Lories, G., Aubrun, A., & Seron, X. (1994). Lesioning McCloskey & Lindemann's 1992 Mathnet: the effect of damage location and amount. *Journal of Biological Systems*, *2*(3), 335-356.

Mandler, G. & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *11*, 1–22

Marr, D. (1982) *Vision*. *A computational investigation into the human representation and* *processing of visual information.* New York: W.H. Freeman.

Mason, M. (1982). Recognition time for letters and nonletters: Effects of serial position, array size, and processing order. *Journal of Experimental Psychology*, *8*, 724-738

McCloskey, M. & Lindemann, A. M. (1992). MATHNET: Preliminary results from a distributed model of arithmetic fact retrieval. In J. I. D. Campbell (Ed.), *The Nature and Origins of Mathematical Skills*. Elsevier Science Publishers B.V.

McConkie, G. W. & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading, *Perception & Psychophysics*, *17*, 578-586.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, Part 1: An account of basic findings. *Psychological Review, 88*, 375-405.

McLeod, P., Plunkett, K., & Rolls, E. T. (1998), *Introduction to Connectionist Modelling of Cognitive Processes*, Oxford University Press.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behaviour Processes*, *9*, 320-324.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227-234.

Miller, A., & Baker, G. (1968). The effects of shape, size, heterogeneity and instructional set on the judgment of visual number. *American Journal of Psychology*, *81*, 83-91.

Miller, E. K. & Desimone, R. (1991). A neural mechanism for working and recognition memory in the inferior temporal cortex. *Science*, *254*, 1377-1379.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002) *Quantitative Development in Infancy and Early Childhood*, Oxford University Press.

Moyer, R. S. & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519-1520

Naccache, L., & Dahaene, S. (2001). The priming method : imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex*, *11*, 966-974.

Najemnik, J. & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*, 387-391.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner, & G. W. Humphreys (Eds.), *Basic processes in reading* , 264-336. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. Science, *297*, 1708.

Nieder, A. & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate pre-frontal cortex. *Neuron*, *37*, 149-157.

Nieder, A. & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Science (USA)*, *101*, 7457-7462.

Norris, D. & McQueen, J.M. (in press). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*.

Nowack, A. (2004). Dynamical Minimalism: Why Less is More in Psychology. *Personality and Social Psychology Review*, *8*(2), 183-192.

Nürk, H. C., Bauer, F., Krummenacher, J., Heller, D., & Willmes, K. (2005). The power of the mental number line. How the Magnitude of Unattended Numbers affects Performance in an Eriksen Task. *Psychology Science: Special Issue Brain and Number*, *42*, 34-50.

Nürk, H.C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. Cognition, 82, B25-B33.

Oja, E. (1989). Neural networks, principal components, and subspaces. *International Journal of Neural Systems*, *1*(1), 61-68.

O'Regan, J. K. (1990). Eye movements and reading. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes*. (Vol. 4 of *Reviews of Oculomotor Research*), 395-453, Elsevier, Amsterdam.

O’Regan, J. K., Lévy-Schoen, A., & Jacobs, A. M. (1983).The effect of visibility on eye-movement parameters in reading. *Perception & Psychophysics*,*34* ,457-64.

O'Reilly, R. C. (1998). Six Principles for Biologically-Based Computational Models of Cortical Cognition. *Trends in Cognitive Sciences*, *2*, 455-462.

Ostad, S. E. (1998). Developmental progression of subtraction studies:a comparison of mathematically normal and mathematically disabled childre. *European Journal of Special Needs Education*, *14* (1), 21-36.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*, 376-404.

Peterson, S. & Simon, T. J. (2000). Computational evidence for the subitizing phenomenon as an emergent property of the human cognitive architecture. *Cognitive Science*, *24*, 93–122

Pesenti, M., Thioux, M., Seron, X., & De Volder, A. (2000). Neuroanatomical substrates of Arabic number processing, numerical comparison and simple addition: A PET study. *Journal of Cognitive Neuroscience*, *12*(3), 461-479.

Phattanasri, P., Chiel, H.J., & Beer, R.D. (submitted). The dynamics of associative learning in evolved model circuits. Submitted to the *Journal of Adaptive Behaviour*, available at http://vorlon.case.edu/~beer/Papers/LearningPaper2.pdf

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus.. *Neuron*, *44*(3), 547-55.

Piazza, M., Mechelli, A., Butterworth, B. & Price, C. J. (2002). Are Subitizing and Counting Implemented as Separate or Functionally Overlapping Processes? *NeuroImage*, *15*, 435-446.

Platt, M. and Glimcher, P.W. (1998) Neurons in LIP carry information correlated with movement probability and reward magnitude. *Investigative Opthalmolology and Visual Science*, *39*, S326.

Plaut, D. C. & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*, 377-500.

Port, R.F. & van Gelder, T. (1995). *Mind as Motion*, MIT Press.

Ramsey, W., Stich, S. P., & Garon, J. (1991). Connectionism, eliminativism, and the future of folk psychology. In W. Ramsey, S. P. Stich, & D. E. Rumelhart (Eds.), *Philosophy and connectionist theory*, 199-228, Hillsdale, NJ: Lawrence Erlbaum.

Rao, C. R. (1964). The Use and Interpretation of Principal Component Analysis in Applied Research. *Sankhya A 26*, 329-358.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59-108.

Ratcliff, R. (2001). Diffusion and random walk processes. *International encyclopedia of the social and behavioral sciences*, Oxford, England: Elsevier, *6*, 3668-3673.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873-922.

Ratcliff, R., & Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347-356.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261-300.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, *124*, 372-422.

Rayner K. & Fisher D.L. (1987). Eye movements and the perceptual span during visual search. In J. K. O'Regan & A. Levy-Schoen (Eds) *Eye movements: From Physiology to Cognition*. Amsterdam: North-Holland, 293-302.

Reder, S. M. (1973). On-line monitoring of eye position signals in contingent and noncontingent paradigms. *Behaviour Research Methods & Instrumentation*, 5, 218-228.

Reicher, G.M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 275-280.

Reynvoet, B., & Brysbaert, M. (1999). Single-digit and two-digit Arabic numerals address the same semantic number line. *Cognition, 72,* 191-201.

Reynvoet, B., Brysbaert, M., & Fias, W. (2002). Semantic priming in number naming. *The Quarterly Journal of Experimental Psychology*, *55A*(4), 1127-1139.

Reynvoet, B., Caessens, B., & Brysbaert, M. (2002). Automatic stimulus–response associations may be semantically mediated. *Psychonomic Bulletin & Revie*w, *9*, 107-112.

Roitman, J., Brannon. E.M. & Platt, M.L. (2007). Monotonic Coding of Numerosity in Macaque. *PLoS Biology*, 5(8).

Roitman J. & Shadlen M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, *22*, 9475-9489.

Rumelhart, D. E. & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volumes 1 and 2*. Cambridge, MA: MIT Press.

Schacter, D. L., Alpert, N. M., Savage, C. R., Rauch, S. L., & Albert, M. S. (1996). Conscious recollection and the human hippocampal formation: evidence from positron emission topography. *Proceedings of the National Academy of Science (USA)*, *93*, 321-325.

Schacter, D. L. & Buckner, R. L. (1998). Priming and the brain. *Neuron*, *20*, 185-195.

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523-568.

Sejnowski, T.J., & Rosenberg, C.R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems*, *1*, 145-168.

Seth, A.K. (1998). Evolving action selection and selective attention without actions, attention, or selection. In R. Pfeifer, B. Blumberg, J. Meyer, & S. Wilson (Eds.), *From animals to animats 5: Proceedings of the Fifth International Conference on the Simulation of Adaptive Behavior,* 139-147, Cambridge, MA. MIT Press.

Seth, A.K. (2008). Causal networks in simulated neural systems. *Cognitive Neurodynamics*, *2*, 49-64.

Shadlen, M. N., Hanks, T. D., Churchland, A. K., Kiani, R., & Yang, T. (2006). The speed and accuracy of a simple perceptual decision: a mathematical primer. In K. Doya, S. Ishii, R. Rao, and A. Pouget (Eds.) *Bayesian Brain: Probabilistic Approaches to Neural Coding*, MIT Press: Cambridge.

Shalev, R. S., Auerbach, J., Manor, O. & Gross-Tsur, V. (2000). Developmental dyscalculia: prevalence and prognosis. *European child & adolescent psychiatry*, *9*(2), 58-64.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Shapley, S. L. (1953). A Value for n-person Games. In H.W. Kuhn & A.W. Tucker (Eds.), *Contributions to the Theory of Games*, *volume II; Annals of Mathematical Studies*, *28*, 307-317, Princeton University Press.

Siegler, R. S. & Opfer, J. E. (2003). The Development of Numerical Estimation: Evidence for Multiple Representations of Numerical Quantity. *Psychological Science*, *14*(3), 237-243.

Simon, O., Mangin, J. F., Cohen, L., Le Bihan D., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, *33*, 475-487.

Simon, T. J., Peterson, S., Patel, G. & Sathian, K. (1998). Do the magnocellular and parvocellular visual pathways contribute differentially to subitizing and counting*? Perceptual Psychophysics*, *60*(3), 451–464

Simon, T. & Vaishnavi, S. (1996). Subitizing and counting depend on different attentional mechanisms: Evidence from visual enumeration in afterimages. *Perceptual Psychophysics*, *58*(6), 915–926.

Smolensky, P. (1987). On Variable Binding and the Representation of Symbolic Structures in Connectionist Systems. *Technical report CU-CS-355-87*, Department of Computer Science, University of Colorado at Boulder.

Spieler, D.H., & Balota, D.A. (1997). Bringing computational models of word naming down to the item level. Psychological Science, 8, 411-416.

Squire, L. R., Ojemann, J. G., Miezin, F. M., Petersen, S. E., Videen, T. O., & Raichle, M. E. (1992). Activation of the hippocampus in normal humans: a functional anatomical study of memory. *Proceedings of the National Academy of Science (USA)*, *89*, 1837-1841.

Steen, L. (1990). Numeracy. *Daedalus Special Issue on "Literacy in America"*, *119*(2), 211-231.

Stevens, M. & Grainger, J. (2003). Letter visibility and the viewing position effect in visual word recognition. *Perception & Psychophysics*, *65*, 133-151

Stich, S., & Ravenscroft, I. (1994). What is Folk Psychology? *Cognition*, *50*, 447-68.

Tomko, G., & Crapper, T. (1974). Neuronal variability: non-stationary responses to identical visual stimuli. *Brain Research*, *79*, 405-418.

Trick, L. M. & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*(1), 80-102.

Tydgat, H. & Grainger, J. (in press). Serial Position Effects in the Identification of Letters, Digits, and Symbols. *Journal of Experimental Psychology: Human Perception and Performance*.

Tzelgov, J., Meyer, J. & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 166-179.

Uller, C., Jaeger, R., Guidry, G., & Martin, C. (2003). Salamanders (Plethodon cinereus) go for more: rudiments of number in an amphibian. *Animal Cognition*, *6*, 105-112.

Underwood, N. R. & McConkie, G. (1985). Perceptual span for letter distinctions during reading. *Reading Research Quarterly*, *20*, 153-162.

Usher, M. & McClelland, J. L. (2001). On the Time Course of Perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*, 550-592.

Van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, *91*(7), 345-381.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioural Brain Science*, *21*, 615-665.

Van Gelder, T. (1999). Dynamic approaches to cognition. In Wilson, R., Keil, F. (Eds.), *The MIT Encyclopedia of Cognitive Sciences*, 244-246, Cambridge MA: MIT Press.

Van Selst, M. & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology*, *47A*, 631-650.

Verguts, T. & De Moor, W. (2005). Two Digit Comparison: Decomposed, Holistic or Hybrid? *Journal of Experimental Psychology*, *52*(3), 195-200

Verguts, T. & Fias, W. (2004). The representation of number in humans and animals: a neural model. *Journal of Cognitive Neuroscience*, *16*(9), 1493-1504

Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin and Review, 12,* 66-80.

Verguts, T. & Van Opstal, F. (2005). Dissociation of the distance effect and size effect in one-digit numbers. *Psychonomic Bulletin & Review*, *12*, 925-930.

Viscuso, S. R., Anderson, J. A., & Spoehr, K. T. (1989). Representing simple arithmetic in neural networks. In G. Tiberghien (Ed.), *Advances in Cognitive Science (Vol. 2)*. Chichester (UK): Ellis Horwood.

Wagenmakers, E-J., van der Maas, H., & Grasman, R. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin and Review*, *14*, 3-22.

Watson, R. A. & Pollack, J. B. (2001). Coevolutionary Dynamics in a Minimal Substrate. In L. Spector et al (Eds.), *Proceedings of the 2001 Genetic and Evolutionary Computation Conference*, Morgan Kaufmann.

Wellman, H. M. & Miller, K. F. (1986). Thinking about nothing: Development of concepts of zero. *Journal of Developmental Psychology*, *4*, 31042.

Whalen, J. (1997). The influence of semantic magnitude representations on arithmetic: Theory, data, and simulation. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, 814-819, Mahwah, NJ: Erlbaum.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal Counting in Humans: The Psychophysics of Number Representation. *Psychological Science*, *10*(2), 130-137.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1*, 59–85

Wilson, F. A., O'Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science, 260,* 1955-1958.

Wong, K.F. & Wang, X.J. (2006). A Recurrent Network Mechanism of Time Integration in Perceptual Decisions. *Journal of Neuroscience*, *26*, 1314-1328.

Zbrodoff, N. J., & Logan, G. D. (2005). What everyone finds. The problem size effect. In J. Campbell (Ed.), *Handbookof Mathematical Cognition*, 331-345, New York and Hove: Psychology Press.

Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing. The SNARC effect and the reverse SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *Journal of Cognition & Culture*, *5*, 165-190.

Zorzi, M., & Butterworth, B. (1999). A computational model of number comparison. In: M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (p. 778-783). Mahwah (NJ): Erlbaum.

Zorzi, M. and Priftis, K., & Umilta, C. (2002) Brain damage: neglect disrupts the mental number line. *Nature*, *417*(6885), 138-139.

Zorzi, M., Stoianov, I., Becker, S., Umiltà, C., & Butterworth, B. (under revision). The numerosity code for the representation of numerical magnitude. *Psychological Review*.

Zorzi, M., Stoianov, I., & Umiltà, C. (2005) Computational modeling of numerical cognition. In J. Campbell (Ed.), *Handbook of mathematical cognition,* chapter 5, 67-83, London: Psychology Press.