



Published in final edited form as:

*J Math Psychol.* 2009 October ; 53(5): 363–377. doi:10.1016/j.jmp.2009.04.004.

## Is there something quantum-like about the human mental lexicon?

**Peter Bruza,**

Faculty of Science and Technology, Queensland University of Technology

**Kirsty Kitto,**

Faculty of Science and Technology, Queensland University of Technology

**Douglas Nelson,** and

Department of Psychology, University of South Florida

**Cathy McEvoy**

School of Ageing Studies, University of South Florida

Peter Bruza: p.bruza@qut.edu.au; Kirsty Kitto: kirsty.kitto@qut.edu.au; Douglas Nelson: dnelson@cas.usf.edu; Cathy McEvoy: cmcevoy@cas.usf.edu

### Abstract

Following an early claim by Nelson & McEvoy (35) suggesting that word associations can display 'spooky action at a distance behaviour', a serious investigation of the potentially quantum nature of such associations is currently underway. In this paper quantum theory is proposed as a framework suitable for modelling the human mental lexicon, specifically the results obtained from both intralist and extralist word association experiments. Some initial models exploring this hypothesis are discussed, and experiments capable of testing these models proposed.

### Keywords

quantum theory; contextuality; semantic structure; human memory experiments; quantum interaction

## 1 Introduction

Quantum theory (QT) is perhaps the most stunningly successful theory ever devised. At present, there are no known experimental deviations from its predictions. While often thought of as only applying to the sub-atomic realm, QT is also crucial for explaining why stars shine, how the universe formed and the stability of matter. Even our current age of information technology owes its origins to a quantum theory of matter. It is important to realise that QT is not only a physical theory in its own right, but also a framework in which theories can be developed. A quantum system is generally modelled using a set of procedural steps that transform a classical model to its quantum analogue, evolve it using particular time evolution, and then make predictions that can be tested by performing measurements upon the system itself. The time evolution and measurement equations are incompatible and one of the great challenges faced by those trying to interpret QT is to understand how and in what circumstances these two sets of equations should be applied (at present a set of ad hoc decisions are made by the person creating the model).

---

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Admittedly the models traditionally developed have been physical, but QT has increasingly been deployed outside of physics (7;8;21;40;1;39;6). This list of references is indicative; QT has been applied within a wide range of fields, including language, economics, artificial intelligence, complex systems science, organisational decision making, models of the brain and cognition etc.

The purpose of this article is to explore an initial quantum model of the human mental lexicon which is detailed enough to allow for experimental predictions to be formulated. We shall follow a long path to reach this destination, but this is to be expected in a work that attempts to apply QT to a field of cognitive psychology. The remainder of this section will consist of a brief review of the current state of the art understanding of the human mental lexicon as has been revealed by recall tasks. We shall see that an ad hoc model owing its origins to a quantum metaphor models experimental results better than the standard spreading activation models, which suggests that a more complete quantum model may be possible. Section 2 briefly introduces QT, before sections 3 and 4 present some early ruminations of how such a quantum model of the human mental lexicon could work. In section 5 we shall discuss the notion of entanglement in QT and show how this phenomenon leads to a natural model of the human mental lexicon which explains many of the effects seen in word association experiments. This suggests a number of experiments which, when they are performed will shed more light on the nature of a quantum model of the human mental lexicon, possibly ruling out such a model or perhaps instead providing strong support for a quantum approach.

### 1.1 A mental lexicon

A mental lexicon refers to the words that comprise a language, and its structure is defined here by the associative links that bind this vocabulary together. Such links are acquired through experience and the vast and semi-random nature of this experience ensures that words within this vocabulary are highly interconnected, both directly and indirectly through other words. For example, the word *planet* becomes associated with *earth*, *space*, *moon*, and so on, and within this set, *moon* can become linked to *earth* and *star*. Words are so associatively interconnected with each other they meet the qualifications of a ‘small world’ network wherein it takes only a few associative steps to move from any one word to any other in the lexicon (38). Because of such connectivity, we argue in this paper that individual words are not represented in long-term memory as isolated entities but as part of a network of related words. However, depending upon the context in which they are used, words can take on a variety of different meanings and this is very difficult to model (15).

Much evidence shows that for any individual, seeing or hearing a word activates words related to it through prior learning. As illustrated in Figure 1, seeing *PLANET* activates the associates *earth*, *moon*, and so on, because *planet-earth*, *planet-moon*, *moon-space* and other associations have been acquired in the past. This activation aids comprehension, is implicit, and provides rapid, synchronous access to associated words.

Understanding how such activation affects memory requires a map of links among known words, and free association provides one reliable means for constructing such a map (32). In free association, words are presented to large samples of participants who produce the first associated word to come to mind. The probability or strength of a pre-existing link between words is computed by dividing the production frequency of a response word by its sample size. For example, the probabilities that *planet* produces *earth* and *mars* are 0.61 and 0.10, respectively, and we say that *earth* is a more likely or a stronger associate of *planet* than *mars*. This attempt to map the associative lexicon soon made it clear that some words produce more associates than others. This feature is called ‘set size’ and it indexes a word’s associative dimensionality (27;34). Finally, mapping the lexicon also revealed that the associates of some words are more interconnected than others. Some words have many such connections (e.g.,

*moon-space, earth-planet*), whereas some have none, and this feature is called “connectivity” (31). Experiments have shown that link strengths between words, the set size and connectivity of individual words have powerful effects on recall which existing theories cannot explain.

## 1.2 Recall Tasks

Both physical and human memory experiments require very careful preparation of the state to be tested. Although a variety of preparations have been used in human memory experiments, we focus here on two: extralist and intralist cuing.

In extralist cuing, participants typically study a list of to-be-recalled target words shown on a monitor for 3 seconds each (e.g., *planet*). The study instructions ask them to read each word aloud when shown and to remember as many as possible, but participants are not told how they will be tested until the last word is shown. The test instructions indicate that new words, the test cues, will be shown and that each test cue (e.g., *universe*) is related to one of the target words just studied. These cues are not present during study (hence, the name extralist cuing). As each cue is shown, participants attempt to recall its associatively related word from the study list.

In intralist cuing the word serving as the test cue is presented with its target during study (e.g., *universe planet*). Participants are asked to learn the pairing, but otherwise the two tasks are the same.

These tasks allow for many variations in the learning and testing conditions and in the associative characteristics of the studied words and their test cues. For example, the to-be-recalled target words can be systematically selected for either task from the norms based on their individual associative structures. With other variables controlled, half of the targets in the study list could be high and half could be low in associative connectivity. Similarly, half could have small or large set sizes. The potential effects of some feature of the human mental lexicon are investigated by selecting words that systematically differ in that characteristic to determine how it affects recall. Extralist cuing experiments show that recall varies with the nature of the test cue and the target as individual entities and with the linking relationships that bind them together. The cue-target relationship can vary in strength in one or in all of four different ways. Figure 2 shows *planet* as a studied target, with *universe* as the test cue. As can be seen, cue-to-target strength is 0.18, and target-to-cue strength is 0.02. These two links directly connect the cue and target, and stronger links increase the probability of correct recall. Recall also varies with indirect links (24). Recall is higher when mediated links (*universe* → *space* → *planet*) and shared associate links are present (both *universe* and *planet* produce *star* as an associate). Finally, with cue-target strength controlled, other findings show that target words having higher levels of associative connectivity (more associate-to-associate links) are more likely to be recalled (33). In contrast, target words with greater dimensionality or set size, are less likely to be recalled (34). Figure 2 shows two associates, one linked to the cue (*eternity*) and one linked to the target (*mars*). These associates do not link the cue and the target together and are likely to compete with the target and hinder recall.

The positive effects of the target’s associative connectivity and the negative effects of its set size occur even though attention is never drawn to the associates at any time. Furthermore, the effects of connectivity and set size in the extralist cuing task are not produced by confounding word attributes, nor are they found only with particular types of participants or conditions (16;30;26). Both effects are evident regardless of target frequency, concreteness, and number of target meanings. The effects are found for young and old participants, under very fast and very slow presentation rates, as well as under incidental and intentional learning and testing conditions. In trying to understand how associative structure has such robust effects on recall,

we learned that standard psychological explanations failed, and that the quantum formalism offered a promising alternative (e.g., (6;14;2;3;35;9)).

### 1.3 Spooky Activation At a Distance

Figure 3 shows a hypothetical target having two target-to-associate links. There is also an associate-to-associate link between Associates 1 and 2, and an associate-to-target link from Associate 2 to the Target  $t$ . The values on the links indicate relative strengths estimated via free association. Nelson et al., (31) have investigated reasons for the more likely recall of words having more associate-to-associate links. Two competing explanations for why associate-to-associate links benefit recall have been proposed.

The first is the Spreading Activation equation, which is based on the classic idea that activation spreads through a fixed associative network, weakening with conceptual distance (e.g., (11)):

$$S(t) = \sum_{i=1}^n S_{ti} S_{ii} + \sum_{i=1}^n \sum_{j=1}^n S_{ti} S_{ij} S_{jt} \quad (1)$$

$$= (.10 \times .70) + (0.20 \times 0.60 \times 0.70) \quad (2)$$

$$= 0.154 \quad (3)$$

where  $n$  is the number of associates and  $i \neq j$ .  $S(t)$  denotes the strength of implicit activation of target  $t$  due to study,  $S_{ti}$  target-to-associate activation strength,  $S_{ii}$  associate-to-target activation strength (resonance), and  $S_{ij}$  associate-to-associate activation strength (connectivity). Multiplying link strengths produces the weakening effect. Activation ostensibly travels from the target to and among its associates and back to the target in a continuous chain, and the target is strengthened by activation that returns to it from pre-existing connections involving two- and three-step loops. More associate-to-associate links create more three-step loops and theoretically benefit target recall by increasing its activation strength in long-term memory. Importantly, note that the effects of associate-to-associate links are contingent on the number and strength of associate-to-target links because they allow activation to return to the target. If associate-to-target links were absent, even the maximum number of associate-to-associate links would have no effect on recall because activation could not return to the target.

In contrast, in the ‘Spooky Activation at a Distance’ equation, the target activates its associative structure in synchrony:

$$S(t) = \sum_{i=1}^n S_{ti} + \sum_{i=1}^n S_{ii} + \sum_{i=1}^n \sum_{j=1}^n S_{ij} \quad (4)$$

$$= 0.20 + 0.10 + 0.70 + 0.60 \quad (5)$$

$$= 1.60 \quad (6)$$

where  $i \neq j$ ;  $S_{ti}$ , target-to-associate  $i$  strength;  $S_{it}$ , associate  $i$ -to-target strength (resonance);  $S_{ij}$ , associate  $i$ -to-associate  $j$  strength (connectivity) This equation assumes that each link in the associative set contributes additively to the target's activation strength. The beneficial effects of associate-to-associate links are not contingent on associate-to-target links. Stronger target activation is predicted when there are many associate-to-associate links even when associate-to-target links are absent. In fact, associate-to-target links are not special in any way. Target activation strength is solely determined by the sum of the link strengths within the target's associative set, regardless of origin or direction.

Figure 4 shows the results of an extralist cuing experiment with cue-to-target strength set at a moderate level. Recall is more likely when target words have more associate-to-associate links and more associate-to-target links. Most importantly for theory, these variables have additive effects. The benefits of associate-to-associate links do not depend on associate-to-target links, indicating that the spreading activation rule provides an inadequate explanation for why the existence of associate-to-associate links facilitates recall. Spreading activation cannot explain why the links between *moon-space* and between *moon-earth* facilitate the recall of *planet* when given *universe* as a test cue.

This conclusion is reinforced by the results of a formal evaluation in which target activation strength was computed for both equations (1) and (4), and then used to predict the probability of correct recall for the 336 word pairs used in the experiments reported in (31).

The probability of recall for each pair was determined from the experimental data by doing item analyses. The results showed that each rule was significantly related to recall, but the correlation between predicted and obtained recall was stronger for the 'Spooky Activation at a Distance' rule (where (4) gives  $r = .57$ ) than for the spreading activation rule (where (1) gives  $r = .35$ ). However, when the rules were entered into a simultaneous multiple regression, only the distance rule was positively and significantly related to probability of correct recall as a predictor [ $F(2, 333) = 79.56$ ,  $MS_{res} = .042$ ,  $r = .57$ ,  $adj\ r^2 = .32$ ]. Overall mean correct recall for these pairs was .63 ( $SE = .01$ ), and the predicted mean recalls for the distance and spreading activation rules were .63 ( $SE = .01$ ) and .32 ( $SE = .01$ ). The findings indicate that the 'spooky rule' outperformed the spreading activation rule in accounting for the findings.

These results indicate that links between a target's associates affect its recall even when they are 'distant' from the target. This bears some resemblance to some of the measurement effects in standard QT. A 'particle' when given enough time can spread over very large distances, indeed it can be split into multiple different components.<sup>1</sup> However, when a measurement is performed upon the 'particle' at one location this inevitably effects other measurements performed upon that same 'particle' at a different location, even if the two locations are widely separated. The choice of a measurement setting for the 'particle' at one location effects the state of that same 'particle' at the other location, even when the two regions are distant.

For a specific example we might consider a photon (the fundamental quantum unit of light). Despite their apparently indivisible nature, these can be sent through a semitransparent mirror, or beamsplitter which 'splits' a photon into two different wave-packets travelling in two different directions. That is, the initial state  $|\psi\rangle$  might be sent through a beam splitter say with a 50% transmittance, which would mean that 50% of the time a 'particle' might instead be reflected. This would be represented as a *superposition* in QT, with the initial state evolving under standard Schrödinger dynamics to a state

<sup>1</sup>Which makes the use of the word 'particle', with its connotation of an indivisible unit, very suspect. This is often the case in QT, where words that were historically used unproblematically in classical mechanics take on very shady and difficult meanings in the quantum world leading to many of the ongoing debates and misunderstandings surrounding the interpretation of QT.

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|reflect\rangle + |transmit\rangle) \quad (7)$$

which represents a photon that is essentially in two places at once. These two different wavepackets might then travel to two different detectors based at two spatially separated points *A* and *B*. However, when the photon reaches a detector it will either be found only at that point, or it will not be found there (i.e., the photon will be found at *either A or B*, never both). The probability of it being found at one of the two points is given by the quantum mechanical projection postulate which allocates in this case a probability of  $(\frac{1}{\sqrt{2}})^2 = 50\%$  to either outcome (see below for more details of how this occurs). However, it is impossible to consider the probabilities as arising, as in classical probability theory, due to our lack of knowledge about which direction the photon actually travelled. Different experiments can be performed with the superposition state that forbid this interpretation which means that this quantum superposition state is somehow real, or real enough to have an effect, even if it is very difficult to understand *what* it might be, using our traditional classical understanding of the world. Thus, the photon that was indivisible somehow divides, and travels in *both* directions, before somehow being found in one position alone.

There are many different interpretations of the quantum formalism, each of which tells a different story about what actually happens to the photon,<sup>2</sup> but for our current purposes it is enough to simply highlight the similarities between this system and the behaviour of the associative sets of words as expressed by the ‘Spooky Activation at a Distance’ equation (4). Here, the instantaneous ‘collapse of word’ during testing to produce a given associate is very reminiscent of the instantaneous collapse of the photon to one or the other position. Also, it is likely that the number of associative links for a particular word could be modelled through a choice of weighting for the particular states involved. There is no reason why the beamsplitter must transmit at 50%, it might instead have a 30% transmission of the photon, in which case the state (7) would also change.

An analogous model for word associations will be gradually constructed and enhanced in what follows.

#### 1.4 The collapse of a word

In the intralist task, the test cue is presented during study with the target and serves as the sole context for the target (and the reverse). If studying a target word in the presence of an associatively related word causes its superposition state to collapse, then the effects of associative connectivity and dimensionality should be reduced and perhaps eliminated in this task. Such effects should be present in the extralist cuing task and absent in the intralist task. The meaning of the target is uncertain in the extralist cuing task because there is no specific semantic context made available during study to bias the meaning of the target. In contrast, the meaning of the target is more certain in the intralist cuing task because a meaningful and interactive context word is presented. Extant findings are consistent with this interpretation. Nelson, McEvoy et al. (30) compared recall in the two tasks where targets or context word-target word pairs were studied for 3 seconds, followed by an immediate cued recall test. Figure 5 shows effects of target connectivity were apparent in the extralist cuing task but not in the intralist cuing task. Studying the pair *universe-planet* completely eliminated the influence of links between *moon-space*, *moon-earth*, and so on. Figure 6 shows that the effects of target

<sup>2</sup>See (5) for a brief entertaining introduction or the more comprehensive review article (23).



dimensionality or set size are also present in the extralist cuing preparation but not in the intralist cuing task. In other words, in intralist cuing target competitor strength is no longer an issue.

Effects of the target's associative structure are essentially eliminated when it is studied in the context of an associatively related word. Other intralist cuing studies have shown that target competitor effects are eliminated even when the context cue present during study is not used to prompt recall during testing (34). Target competitor effects are not found when word pairs are studied, regardless of whether recall is prompted by intralist or extralist cues. The apparent collapse of the target's associative structure transcends the nature of the test cue. Hence, the collapse of a word's associative structure occurs during study, not during retrieval, and this collapse is brought about by the presence of the context word during study. Bruza & Cole (6) modelled context effects on word superpositions by assuming that context acts like a measurement leaving the word in a basis state. A basis state was equated with a particular sense of the word, a model that will be further developed here.

Intuitively, it seems likely that a target's superposition state would occur whenever the context is uninformative, ambiguous, or simply delayed (25), and the collapse of this state to a definite value becomes apparent as soon as context is informative, unambiguous, and simultaneously present. When this occurs, multiple senses of the target may no longer influence recall because the collapse brought about by context dampens their accessibility. When *bolt* is normed in isolation, it produces 11 associates, each reflecting three different senses of the words: 'Weather', 'fastener', and 'rapid movement'. However, senses of the target that are unrelated to the study context should not be produced even in free association, which allows any related response to be produced. For example, associates of the word *bolt* that are normally produced with a high probability when it is presented in isolation may not be produced when the pair *lightening bolt* is presented as the cue to free associate. To evaluate this possibility participants were asked to produce the first word to come to mind to pairs of words that captured the meaning of the relationship. The responses to pairs such as *lightening bolt* were then compared to the responses to *bolt* when it was presented in isolation.

The strength of each sense can be determined by summing the probabilities of the individual associates related to each sense. For example, the strength of the weather sense is determined by summing the strength of the associates 'lightening' (.18) and 'thunder' (.02), which equals .20. Using this procedure, the relative strengths of the 'weather', 'fastener', and 'rapid movement' senses are .20, .67, and .09, respectively. In contrast, free associations to *lightening bolt*, show that the relative probabilities of the three meanings are .94, 0.0, and 0.0. When context is absent, the fastener sense is strongest, but it is not represented in the set when the lightening context is present, nor is the rapid movement sense. Not surprisingly, the probability of the weather sense increases when weather is emphasized even though lightening is removed from the count because it was presented. Thus, relatively strong senses linked to a word disappear in the response set when the same item is processed in a meaningfully related context that biases processing towards a different sense.

The process of quantum measurement often generates very similar effects. Here we find that measuring a quantum system leaves it in a definite state (this is often referred to as a process of *quantum collapse*, see below), which means that performing another measurement in quick succession almost inevitably results in that same state being measured. Performing the *lightening bolt* measurement might be similarly collapsing the cognitive states of the subjects to the 'weather' sense, making it very hard for them to access the other senses of *bolt* and hence explaining the insignificantly small recall values for the 'fastener' and 'rapid movement' senses. Thus, the context in which a word is presented can have a profound effect upon the word associations that are generated, and there is some reason to believe that the quantum formalism might provide a good model for this form of effect.

We shall discuss this idea more fully in the next section, which will provide a brief introduction to the formalism of QT and explain why it is believed that quantum models of human word association experiments might be appropriate. Following this quick primer, we shall turn to some initial models of word meaning motivated by QT.

## 2 Concepts in quantum mechanics

According to standard QT there are two forms of time evolution exhibited by the wavefunction,  $|\psi(\mathbf{x}, t)\rangle$ , which represents the current state of a quantum system (in a complex, linear vector space known as a *Hilbert space*,  $\mathcal{H}$ ):

1. A continuous linear evolution represented by an equation of motion. This evolution occurs in all situations but that of *measurement*, when,
2. an instantaneous, nonlinear collapse occurs. After this collapse, the system is found in one of a set of possible states all of which are eigenvectors obtained from a combination of the measurement apparatus and the system itself. The result of the measurement is probabilistically determined from the associated eigenvalue of the eigenvector. This second form of time evolution is often called the *collapse*, or *projection postulate*, and the incompatibility between it and the first form of time evolution leads to one of the most vexing problems in QT, namely the quantum measurement problem.

The Schrödinger equation,  $i\hbar \frac{d}{dt}|\psi(\mathbf{x}, t)\rangle = H|\psi(\mathbf{x}, t)\rangle$ , is the generic dynamical equation of motion for standard quantum systems. In this equation  $H(\mathbf{x})$  is the Hamiltonian, a Hermitian (hence probability conserving) linear operator that can be derived from the Euler–Lagrange equations of motion of the associated ‘classical’ system. However, there is no a priori reason to suppose that this is the only equation that can govern the time evolution of quantum systems, especially those not normally covered by physics. The ‘ket’  $|\psi\rangle$  corresponds to the traditional column vector  $\vec{\psi}$  of unit length. This notation originates from the renowned quantum physicist Paul Dirac, and as we proceed, Dirac notation will gradually be introduced.

It is important to appreciate that because of the two forms of dynamical evolution exhibited by a quantum state it is not always possible to perform two different experiments upon the same physical system; the measurement context of a quantum system under study can have a significant effect upon the system itself, as well as upon any measurements performed upon it. This fundamental characteristic of quantum systems has led to some of the most profound results in the field, from Heisenberg’s Uncertainty relations, to the more recent results surrounding the contextuality and nonlocality that is apparently inherent in quantum models (23). Some of these results shall be introduced in this paper as they are required for the discussion. For now, it is worth asking *why* the quantum formalism might be considered as a good descriptor of the human mental lexicon.

Generally, in extracting the probability of some result, the wavefunction,  $|\psi\rangle$ , is written in terms of a set of basis states,  $\{|\varphi_i\rangle\}$ , which are chosen such that they correspond well with the variable to be measured. A representation of  $|\psi\rangle$  is obtained by expanding it as a linear superposition (i.e. an appropriately weighted sum) of one set of basis states (obtained through reference to the *choice* of apparatus and its orientation, state etc.). We find that  $|\psi\rangle = \sum_i c_i |\varphi_i\rangle$  where the weight terms  $c_i$  represent the contribution of each component of the basis to the actual state. The choice of basis states is governed by the observable to be measured and the quantization procedure that relates observable,  $A$ , to its counterpart in the quantum formalism,  $\hat{A}$ ; with a good choice we find that  $\hat{A}$  satisfies an eigenvalue equation  $\hat{A}|\varphi_i\rangle = a_i|\varphi_i\rangle$  (i.e. the superposition is nondegenerate) and that the spectrum of the operator representing the observable to be measured is real (i.e. the operator is self-adjoint, or Hermitian, for the choice of basis) (19). An example will help to clarify how the concept of an operator is used in QT. We might choose



to measure the position of a quantum particle. To do this we would make use of a *position operator*,  $\hat{\mathbf{X}} = (X_x, X_y, X_z)$ , which would correspond to an experimental setup capable of measuring the position of the quantum particle in each one of the three spatial dimensions ( $x, y, z$ ) (effectively performing the three experiments  $(X_x, X_y, X_z)$  at once). The experimental application of this operator would lead to an eigenvalue equation

$$\hat{X}_\alpha|\mathbf{x}\rangle = x_\alpha|\mathbf{x}\rangle, \quad (\alpha=x, y, z) \quad (8)$$

which has an unbounded continuous spectrum corresponding to the idea that space is a continuum and that each of the three position measurements commute (4). Other well-known quantum mechanical operators have been found for physical entities such as momentum, angular momentum, spin, and energy. It is anticipated that a similar set of operators might be found to describe the human cognitive state. One such operator, corresponding to a cue in a human memory experiment, will be introduced shortly.

The choice of basis that is made in formulating a quantum description is what leads to the expectation that QT might be used in the description of contextual systems. Right at the core of measurement in QT we see a recognition of the context of the system as important in extracting statements about its state. Thus context is represented in QT with the choice of basis; in choosing a suitable basis quantum theorists implicitly incorporate the context of the system under description. While this choice of basis does not lead to a discernable result in QT (4; 19) it does suggest that the mechanism of formally recognising context might be leveraged in describing systems outside of physics. This paper will explore this idea for the case of words.

The modelling of the human mental lexicon presents a very challenging problem. This is due to the spectrum of meanings that words themselves can take depending upon the context in which the word is used. For example, the word *bat* has a number of possible meanings, or senses, depending upon the context; it could mean the small furry creature generally found in caves (e.g. a vampire bat); it could be a sporting implement (e.g. a baseball bat); it could also take the more colloquial usage of a strange old lady (e.g. an old bat). A number of verb meanings are also possible: an idea might be batted around; someone might bat at a ball; they may even bat their eyelashes in order to attract attention and admiration. Some of these meanings are related. For example, verb meanings of bat can often be related to the sporting sense of bat in their etymology, but this does not affect how we extract meaning from a sentence. Clearly, we can only distinguish between this wide range of different senses by looking at the context in which the word occurs. Context is hard to define as it generally includes everything but the system under study. In the case of words, the sentence in which the words reside can be considered a very important contribution to that context, however, sometimes the context may need to be widened to include not only more of the text, but even genders, cultural groupings, historical periods etc.

There are very few formalisms capable of modelling such contextual dependencies exhibited by words when we try to extract their meanings, QT is one. In (22) two important factors behind the success of QT are identified. Each factor is linked to a long held assumption about the nature of reality that is implicitly rejected by the quantum formalism. These are the assumptions of:

**Objects**, or pre-existing entities with clearly definable boundaries. This assumption makes the standard techniques of reductionism available in the construction of any model. Thus, a system can be cleanly separated into its constituents and modelled according to their separate behaviour alone, with the model's individual elements eventually being synthesised into a larger picture. This assumption is not always valid

for quantum systems. As was mentioned above, allegedly indivisible ‘particles’ can be somehow divided into states of superpositions in the quantum picture of reality, with the results of later measurements unexplainable if the particle is considered to have been in a single location at each particular time instant.

**Objectivity** is another assumption rejected by QT. This is the assumption that the natural world is one consisting of objects that we simply measure, but do not in any way influence in such a way as to change the results of those measurements (although the state of the object itself might be changed). According to the objectivity assumption our measurements are merely extracting information about pre-existing ‘elements of reality’ (13), they are not in any way creating the results of their measurements. This assumption has been proven wrong in one of the most celebrated debates surrounding QT (see (23) for a good review of this debate). Experiments have consistently favoured QT (43), showing that reality does not always consist of objects behaving objectively.

Many of the ‘complex systems’ currently defying our reductive techniques display similar behaviour (22). That is, they cannot be modelled using assumptions of objects displaying objective behaviour. Many of the quantum models currently being proposed in the new emerging field of quantum interaction (7;8) are aimed at modelling precisely these systems. For a specific example we might return to the the word *bat*. This cannot be cleanly separated from other words used concurrently; and neither can meaning be attributed to the word without a consideration of the context in which it occurs. Even when *bat* is studied in isolation during an extralist cuing task, the large body of memory research cited above suggests overwhelmingly that it cannot be considered as isolated from its associative structure. Both the assumptions of objects and of objectivity appear to be somehow invalid when applied to the meaning of words. Since QT itself appears to somehow invalidate these assumptions, there is reason to suppose that a quantum model might perform more adequately than the more standard ‘classical’ ones. In what follows, as we gradually introduce more quantum theoretic concepts when they become necessary to the discussion, we shall argue that this is indeed the case.

### 3 Words, Context and Hilbert Space

Up until this point no explicit quantum model of the human mental lexicon has been presented. In this section we shall start to construct a quantum-like model of an *individual* subject’s mental lexicon and the way in which this structure then interacts with a memory experiment. The next section will show how such a model can be extended to the groups of individuals tested in human memory experiments.

Taking advantage of the implicit consideration of context in QT, the starting point of this model of words and their associations will be a set of basis states similar to those used in the measurement of a quantum system. A basis  $\mathcal{B}$  is a set of linearly independent vectors  $\{|v_1\rangle, \dots, |v_n\rangle\}$ . For our purposes, the basis is assumed to be orthonormal, meaning the vectors in  $\mathcal{B}$  are pairwise mutually orthogonal and of unit length. This is also usually the case in QT although it need not be.

A basis defines a vector space, and the vector spaces employed in QT are vector spaces over complex numbers. A Hilbert space is a complete inner product space. In the formalization to be presented here, an  $n$ -dimensional Hilbert space over the field of real numbers will be employed using Euclidean scalar product as the inner product. A Hilbert space defined by a basis  $\mathcal{B}$  will be denoted  $\mathcal{H}_{\mathcal{B}}$ . The state of a quantum system is represented in  $\mathcal{H}_{\mathcal{B}}$  using the basis that defines it. Just like in QT, the application of quantum formalism to memory experiments involves a choice of basis. This choice depends to a large degree on how context is to be brought into the picture in relation to a word.

In order to find an appropriate basis, we shall start with the free association probability  $\Pr(w|q)$  of a word  $w$  being recalled in relation to a cue  $q$ . In conditional probabilities, the symbols to the right of the “|” can be viewed as context for what is left of the “|”. The basis in this case takes the form  $\{|0_q\rangle, |1_q\rangle\}$ , where the basis vector  $|0_q\rangle$  represents the basis state ‘not recalled’ and represents the basis state ‘recalled’ in relation to the cue  $q$ . The word  $w$  is assumed to be in a state of superposition reflecting its potential to be recalled, or not, in relation to the given cue  $q$ :

$$|w\rangle = b_0|0_q\rangle + b_1|1_q\rangle, \quad (9)$$

where  $b_0^2 + b_1^2 = 1$  implies that the state is normalised (which itself implies that the probability of a given outcome upon measurement lies between 0 and 1). This quantum superposition of ‘recalled’ and ‘not recalled’ is called a  $q$ -bit in the physics community. The connection between choice of basis and the context of  $w$  becomes clearer by considering that the same vector  $|w\rangle$  can be generated by a different choice of basis. Thus, the same word  $w$ ’s state in relation to another cue  $p$  is accordingly:

$$|w\rangle = a_0|0_p\rangle + a_1|1_p\rangle. \quad (10)$$

This idea of a word in the context of two different cues, represented via two different bases is depicted in figure 7. Word  $w$  is represented as a unit vector and its representation is expressed with respect to the bases  $\mathcal{P} = \{|0_p\rangle, |1_p\rangle\}$ , and  $\mathcal{Q} = \{|0_q\rangle, |1_q\rangle\}$ .

In QT superposed states like  $w$  are never ‘seen’ as such, but rather a certain outcome is observed. Thus, when a subject is presented with a cue  $q$ , a word  $w$  is either recalled, or it is not recalled. This probability of a given outcome is obtained via the projection postulate, which considers some observable  $A$ , represented by a self-adjoint linear operator,  $\hat{A}$ , and a (normalised) state,  $w \in \mathcal{H}_p$ , and returns an expected value. In both probability and QT the *expectation* of a discrete random variable is the sum of the probability of each possible outcome of the experiment multiplied by the outcome value. This value provides an elegant notation for the probability of a certain outcome being observed, framed in terms of the observable and the state of the system. The *expectation value* associated with performing a measurement of some observable  $A$  upon a system in state  $|w\rangle$  is:

$$E_A = \langle A \rangle_w = \langle w | \hat{A} w \rangle \quad (11)$$

$$= \sum_{i,j=1}^N w_i^* A_{ij} w_j \quad (12)$$

$$= \begin{pmatrix} w_1^* & w_2^* & \dots & w_N^* \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1N} \\ A_{21} & A_{22} & \dots & A_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N1} & A_{N2} & \dots & A_{NN} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix} \quad (13)$$

If  $\hat{A}$  has a complete set of eigenvectors  $\varphi_j$ , with eigenvalues  $a_j$ , (i.e.  $\hat{A}|\varphi_j\rangle = a_j|\varphi_j\rangle$ ), where the basis  $\mathcal{B} = \{|\varphi_1\rangle, \dots, |\varphi_n\rangle\}$ , is assumed,  $w \in \mathcal{H}_{\mathcal{B}}$  and  $\hat{A}$  is an operator on  $\mathcal{H}_{\mathcal{B}}$ , then (11) can be

expressed as  $\langle \hat{A} \rangle_w = \sum_j a_j \langle w | \phi_j \rangle^2 = a_1 (w_1 \phi_1)^2 + \dots + a_n (w_n \phi_n)^2$ . In this situation we can quickly see the way in which the probability,  $|\langle w | \phi_j \rangle|^2$ , of obtaining a particular measurement outcome,  $a_j$ , relates to the expectation value of the operator  $\hat{A}$  which is the sum of all possible outcomes.

A further understanding of probability in QT arises from taking account of its geometrical nature; it derives from taking the cosine between  $\vec{w}$  and  $A\vec{w}$ , which means that probabilities in QT are related to the geometry of the space representing the state of the system. In the case of a single word  $w$  and a cue operator  $A_q$ , this corresponds to a situation in which the outcomes are modelled as Boolean “recalled” ( $|1_q\rangle$ ) or “not recalled” ( $|0_q\rangle$ ) target words. Here, the probability of a target word being recalled in relation to the cue  $q$  relates to its projection down onto the  $|1_q\rangle$  axis in figure 7. These probabilities are of a geometric nature (which is a unique characteristic of quantum probability (19)), and due to Pythagoras’ theorem  $b_0^2 + b_1^2 = 1$  (with the assumption that  $w$  is normalised). When a subject is presented with a cue  $q$ ,  $w$  collapses onto the basis state  $|0_q\rangle$  ( $w$  not recalled) or  $|1_q\rangle$  ( $w$  recalled) with probabilities  $b_0^2$  or  $b_1^2$  respectively. Thus, after the experiment, the subject’s cognitive state has changed from a quantum superposition state to one single ‘classical’ state; word  $w$  has been produced as an associate, or it has not.

The operator  $\hat{A}$  has a number of interesting characteristics resulting from its self-adjoint nature. In particular, the eigenvalues  $a_j$  satisfying the eigenvalue equation  $\hat{A}|\phi_j\rangle = a_j|\phi_j\rangle$ , for self-adjoint operator are real numbers, and the eigenvectors corresponding to two different eigenvalues of a self-adjoint operator are orthogonal (19) which assists in the identification of a suitable basis to represent measurement. In particular, the *spectral theorem* can be applied to any self-adjoint operator (19) which means that it can be expanded into a linear combination of the 1-D projectors that correspond to its eigenvectors. Hence, it is always possible to represent the general operator  $\hat{A}$  as a weighted sum of projection operators,

$$\hat{A} = \sum_{m=1}^M \lambda_m \hat{P}_m \quad (14)$$

where the projectors  $\hat{P}_m$  are a set of pairwise orthogonal operators (19). A real matrix is self-adjoint if  $\hat{A}^T = \hat{A}$ , which implies that the matrix must be symmetric.

To illustrate this formalism, consider once again word  $w$  in relation to the cue  $q$  represented in an associated basis  $\{|0_q\rangle, |1_q\rangle\}$ , with  $|w\rangle = b_0|0_q\rangle + b_1|1_q\rangle$ , where  $b_0^2 + b_1^2 = 1$ . In this scenario  $w$  can be considered a target word forming part of a subject’s cognitive state. When the subject is presented with  $q$ ,  $w$  will, or will not, be recalled. At the end of the experiment, the subject will be in a new cognitive state, they will either have recalled the word (and be in a cognitive state  $|1_q\rangle$ ) or they will not have recalled the word (and be in the state  $|0_q\rangle$ ). This is akin to a quantum measurement which ‘collapses’ the superposition onto the corresponding basis state  $|0_q\rangle$  ( $w$  not recalled) or  $|1_q\rangle$  ( $w$  recalled). Leaving aside the cognitive state of the subject for a moment, the two experimental outcomes are represented as  $\lambda_i$ ,  $i \in \{0, 1\}$ , where  $\lambda_0$  corresponds to the experimental result of  $w$  not being recalled and  $\lambda_1$  corresponds to the experimental result of  $w$  being recalled. For this simple scenario, we represent the measurement of the subject’s cognitive state using the cue  $q$  as two projection operators (19;4),  $\hat{A}_0 = |0\rangle\langle 0|$  corresponds to the word not being recalled, and  $\hat{A}_1 = |1\rangle\langle 1|$  to the converse scenario. Hence, the probability of some outcome  $\lambda_i$ , for each of the two different scenarios is given by

$$\Pr(\lambda_i) = E_{A_i} = \langle w | \widehat{A}_i | w \rangle^2 \quad (15)$$

$$= (b_0 \langle 0 | + b_1 \langle 1 |) | i \rangle \langle i | (b_0 | 0 \rangle + b_1 | 1 \rangle) \quad (16)$$

$$= b_i^2, \quad (17)$$

where we have made use of the orthonormality of the basis to extract the  $b_i$  values, and we note that as these two values sum to 1 it is quite possible to think of them as probabilities.

Modelling a free association recall task can be done as follows. The cue word  $|q\rangle$  as a unit vector in a Hilbert space where the basis vectors  $\mathcal{B} = \{|x_1\rangle, \dots, |x_n\rangle\}$  correspond to the  $n$  potential associates of  $q$ :

$$\begin{aligned} |q\rangle &= a_1 |x_1\rangle \\ &+ \dots + a_n |x_n\rangle \\ &, \text{ where } \sum a_i^2 = 1. \end{aligned} \quad (18)$$

Thus, a potentially very high-dimensional Hilbert space represents the word  $w$  in the context of all of its associates, just as planet is cognitively stored with its associates in figure 1. Observe how in this case the Hilbert space naturally models the associative dimensionality of a word, an important aspect mentioned in the introduction. When a quantum state  $|\psi\rangle$  is measured the projection postulate implies that it ‘collapses’, thus, after the measurement it is no longer in a superposition state, but rather it is in *one* of the possible states exemplified by the eigenstates of the operator depicting the measurement. In free association, the eigenstate corresponds to a particular associate being recalled. In other words, measurement of a property to a high degree of accuracy erases all information about other properties of the state. ‘Measurement’ of word senses appears to behave in the same manner; a sufficiently strong context erases all information about the other senses. An appropriate analogy is the Necker cube<sup>3</sup>, which is an ambiguous line drawing. The human perceptual mechanism will switch between alternate interpretations of the drawing, but both interpretations *cannot* be perceived simultaneously.<sup>4</sup> In QT, measuring a quantum system unavoidably disturbs it leaving it in a basis state determined by the outcome. This phenomenon carries across to memory experiments in the sense that recalling a word, or not, unavoidably disturbs the cognitive state of the subject in question. Evidence can be found in the free associates of *lightning bolt* as described earlier. Twenty pairs such as lightning bolt were normed and compared to the target word presented alone. Across the 20 pairs the mean probability for the biased sense (e.g. the weather sense) when context was absent was .52 (SD = .31) and when the biasing context was present the mean probability was .95 (SD = .07). The alternative meaning (e.g., fastener) had a mean probability of .40 (SD = .28) when no context was present and .04 (SD = .06) when a biasing context was present. In other words, after collapse, the other senses are hardly available. These findings echo the context dependent model of lexical ambiguity access (37). If the context is weak or ambiguous, then multiple senses may be activated, whereas when overall biasing

<sup>3</sup>[http://en.wikipedia.org/wiki/Necker\\_cube](http://en.wikipedia.org/wiki/Necker_cube)

<sup>4</sup>Interestingly, Conte et al. (12) have proposed a quantum-like model for Gestalt phenomena.

context is strong, then the effect upon activation is assumed immediate, and thus only the relevant sense will be accessible. We shall now sketch out a simple toy model of this process.

Recall from the introduction, in the extralist cuing task, subjects are asked to study  $m$  target words  $t_1, \dots, t_m$  which are usually carefully chosen so as to be independent of each other. The study period leaves the subject in a certain cognitive state,  $|\psi\rangle$  say, and they are then cued with a word not in the target list and recall a particular target word. We model this situation with QT by assuming that after the preparation phase the cognitive state of the subject is in a superposition of the eigenstates that pertain to the possible outcomes, i.e., the set of words presented during study. The target words correspond to the experimentally measured outcomes; they are the eigenvalues of the measurement that was conducted by exposure to the cue  $q$ . Thus, to each measurement outcome (a word  $t_i$  recalled by the subject) there corresponds a cognitive state which is represented by the eigenvector satisfied by the equation

$$\hat{q}|t_i\rangle = t_i|t_i\rangle. \quad (19)$$

In a quantum picture the cue would act like a measuring device which collapses a subject's potentially very complex initial cognitive state onto one value; a particular target word is recalled. Thus, subjects are presented with a cue that was not in the preparation phase, but has a relationship to the words that a subject studied during that phase. The subject's cognitive state after preparation is represented by some wavefunction of form (18), which could be described by *any number of different* basis states. It would make sense however, to represent this state with reference to the target words that the subject is instructed to associate with the forthcoming cue. Thus, there is a *natural* choice of basis, that can be chosen to represent  $|\psi\rangle$ :

$$\begin{aligned} |\psi\rangle &= b_1|t_1\rangle \\ &+ b_2|t_2\rangle \\ &+ \dots + b_m|t_m\rangle \\ &, \text{ where } \sum b_i^2 = 1. \end{aligned} \quad (20)$$

Such a cognitive state represents a subject who is currently thinking of the  $m$  target words, but with some words contributing more strongly than others. For example, returning to the word *bat* there are two main sets of associates surrounding the word: the 'vampire', 'blind', etc. associates that are related to the flying mammal, and the 'ball', 'baseball', etc. associates which relate to the sport sense of the word. All of these potential associates might be listed as target words in an extralist experiment (although this would admittedly be a poor experimental design). However, a particular subject might have a strong association with 'vampire' hence the weighting attributed in equation (20) to this target would be larger than the more weakly associated target word 'blind' etc. There is no reason to assume orthonormality holds between the quantum representations of cue and target, which makes the derivation of values like (17) difficult. All we can surmise is that when we perform a measurement, represented by a cue operator  $\hat{q}$  upon the subject's cognitive state (20), the expectation value (average value) would be given by

$$E_{\hat{q}} = \langle \psi | \hat{q} | \psi \rangle \quad (21)$$

which is not an easy entity to either manipulate (due to the lack of orthonormality) or relate to experimental outcomes (since the subject's cognitive state collapses onto a target).



Despite the illustrative nature of this toy model, it fails to incorporate the connectivity and set-size effects on recall mentioned above. (A quantum model taking these aspects into account is presented in section 5.1). Another immediate problem presents itself, and its solution will make this model, while conceptually simple and hence easier to understand, less important to the eventual development of a full quantum model of the human mental lexicon. This problem arises because of the fundamental nature of psychology experiments; they consist of numerous individual tests of a large number of subjects and there are no data that can be compared with the cognitive state of an individual. In order to compare a quantum model with experimental data we must turn to a model capable of talking about the cognitive state in relation to ensembles of subjects. The first steps in that direction will be made in section 4. First however, we will briefly consider the case of intralist cueing.

In the intralist scenario, a cue and its associated target are presented together to the subject during the training phase. This results in the meaning of the word in the subjects cognitive state ‘collapsing’ to the one presented during study. Returning to our ‘bat’ example, if presented with *bat ball* during study, the subject’s cognitive state will have been already collapsed to the sporting sense of ‘bat’. Such a subject should respond ‘ball’ when presented with *bat* during the experiment itself. However, the time interval involved between training and testing is of interest. We might after all expect that some portion of subjects would give a response pertaining to the mammal sense. A quantum approach could be expected to model such an effect very elegantly with the addition of a time evolution operator, but unfortunately the lack of data makes the development of such an operator difficult. This is a promising area for future experimental and theoretical work.

As a brief example of this, we propose an experiment where a subject during is presented with an intralist word pair where the cue has an ambiguous meaning, i.e., they might see the pair *bat cave*, where ‘bat’ is the subsequent cue. We might expect that a larger proportion of false recalls would be exhibited with a longer time between study and cueing of the subject. Indeed, we might even predict that in some cases a subject would respond with ‘ball’ say, instead of the expected target ‘cave’. This would be due to the time evolution of the subject’s cognitive state back to a situation of some ambiguity, where it could be considered as a superposition of the two different meanings. This is a line of research that we will pursue in future work.

#### 4 Modelling recall experiments across multiple subjects

The previous section illustrated how the state of a single quantum particle (a word) can be modelled as a state vector in Hilbert space. This model applied to one subject in a psychology experiment, and showed how the subject’s cognitive state might ‘evolve’ during study from a quantum superposition representing all possible word associations to one word association alone; the one that was produced when exposed to the cue. An immediate problem arises in the application of this model to the experimental data mentioned in section 1. These data are gathered over many different subjects. It is a classical ensemble of data obtained from many subjects that provides these data. This problem is not insurmountable however, because QT has a formalism for dealing with such *mixed states* rather than the simpler *pure states* discussed above in section 3. In QT density matrices are usually employed for this purpose, as they are able to represent statistical mixtures of many different quantum states.

In such a scenario the ensemble of cognitive states of many different subjects after study is modelled as a density matrix  $\pi$ . A density matrix is a self-adjoint and linear operator with the trace, defined as the sum of its diagonal elements, equal to unity:  $\text{tr}(\pi) = 1$ . We shall approach this concept by first looking at the density matrix for an individual subject, and then extending the concept to multiple subjects.

Figure 8 represents a  $q$ -bit register, modelling the potential of one individual subject to recall a word from a set of studied targets. The subject's cognitive state is represented as a superposition of  $q$ -bits, that is, a superposition of 'recalled' and 'not recalled' target words with a weighting related to the probability of each target to actually be recalled. In the case of  $m$  study words, the cognitive state of this one subject could be represented by the pure state

$$|\psi\rangle = |t_1\rangle \otimes \dots \otimes |t_m\rangle \quad (22)$$

where each  $q$ -bit can be written, analogously to (9), as  $|t_i\rangle = (a_0)_i|0\rangle + (a_1)_i|1\rangle$ . However, instead of using this pure state approach, we could use the density matrix formalism which is defined for the pure state of the individual as:

$$\pi = |\psi\rangle\langle\psi|. \quad (23)$$

If we are to represent *all* subjects then we must sum the contribution that each subject's cognitive state makes to the statistics regarding the likelihood of recall of a particular target by one subject alone. In this situation it is not appropriate to describe the combination of all states via the pure state  $|\psi\rangle$ , rather, the collection (or ensemble) of all  $n$  subjects should be represented using the density matrix:

$$\pi = \frac{1}{n} \sum_{i=1}^n (|t_1\rangle \otimes \dots \otimes |t_m\rangle)_i \langle t_1| \otimes \dots \otimes \langle t_m|_i. \quad (24)$$

Here the subscript  $i$  is used to denote the fact that each subject's cognitive state will most likely consist of a  $q$ -bit register in a different state (see figure 9).

Figure 9 illustrates a collection, or ensemble of subjects, each similar to the individual subject of figure 8. That is, a many particle quantum system consisting of a number of subjects, each of which have a cognitive state that could be represented as a  $q$ -bit register.

As was the case in section 3, the probability of obtaining a result from a particular subject in response to a particular cue is represented using an expectation value, but here a new cue operator,  $\hat{Q}$ , is defined and used to calculate the expectation value with reference to the ensemble of subject's cognitive states. This is calculated using the relationship

$$E_{\hat{Q}} = \langle \hat{Q} \rangle = \text{tr}(\pi \hat{Q}). \quad (25)$$

As was the case previously the cue  $q$  can be represented as an operator  $\hat{Q}$  which 'measures' the 'cognitive state'  $\pi$ . The cue operator  $\hat{Q}$  has eigenvalues  $\lambda_1, \dots, \lambda_n$ , and a basis can be found such that each eigenvalue  $\lambda_i$  corresponds to the experimental outcome of target word  $t_i$  being recalled.

To briefly summarise the above, the subject initially represented after having studied the targets by the density matrix  $\pi$  will, upon being presented by a cue  $\hat{Q}$ , respond with some target  $t_i$ . The probability of the particular target  $t_i$  being recalled can be extracted from the expected value (25) in a similar manner to that of section 3. In particular, the diagonal elements of the matrix  $\pi \hat{Q}$  yield the probabilities of each particular target being recalled, in the limit of large numbers of subjects.

The above example could further be refined to include ‘backward strengths’ (i.e., target-cue free association probabilities) and other associates contributing to the cue activation level. Another possible benefit from moving to a density state picture can be found in the matrix representation of the subject’s cognitive state; since it is of matrix form, it is easy to represent associate-associate links. This is an area for future investigation.

## 5 Entangling words and meaning

*I would not call [entanglement] one but rather the characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought,*

Schrödinger, p 807 (36)

How should we represent the combination of words in the human mental lexicon? QT uses the tensor product,  $\otimes$ , to denote composite systems. We shall build up this concept through the use of a series of examples. Consider the case of  $m = 2$  study words:  $u$  and  $v$  presented to a subject. Let us assume that, when cued, the subject recalls neither target word. In this case we could represent the cognitive state of the subject after the experiment as:

$$|u\rangle \otimes |v\rangle = |0\rangle \otimes |0\rangle = |00\rangle \quad (26)$$

where the notation  $|00\rangle$  is just shorthand for the tensor product state  $|0\rangle \otimes |0\rangle$  describing the the state corresponding to neither  $u$  nor  $v$  being recalled. If word  $u$  alone was recalled then we would write  $|u\rangle \otimes |v\rangle = |10\rangle$ , whereas in the converse case we would write  $|01\rangle$  and finally, if both words were recalled then the tensor product would yield the state  $|11\rangle$ .

However, this straightforward scenario is not the only form of situation possible in the quantum formalism. We have seen a number of scenarios where superposition states can occur, and these are important as they can represent the situation where the words  $u$  and  $v$  may be more likely to be recalled in one context than another. Assume, as we did in section 3, that we can represent one subject’s cognitive state with reference to the combined targets  $u$  and  $v$  as a 2  $q$ -bit register that refers to their states of ‘recalled’ and ‘not recalled’ in combination. Thus, if we represent the target words using the standard superpositions  $|u\rangle = a_0|0\rangle + a_1|1\rangle$  and  $|v\rangle = b_0|0\rangle + b_1|1\rangle$  that were discussed in section 3, then it is possible to denote the state of the combined system by writing the tensor product

$$|u\rangle \otimes |v\rangle = (a_0|0\rangle + a_1|1\rangle) \otimes (b_0|0\rangle + b_1|1\rangle) \quad (27)$$

$$\begin{aligned} &= a_0b_0|00\rangle \\ &\quad + a_1b_0|10\rangle \\ &\quad + a_0b_1|01\rangle \\ &\quad + a_1b_1|11\rangle, \end{aligned} \quad (28)$$

where  $|a_0b_0|^2 + |a_1b_0|^2 + |a_0b_1|^2 + |a_1b_1|^2 = 1$ . This is the most general state possible. It represents a quantum combination of the above four possibilities, obtained using a tensor multiplication between the states  $|u\rangle$  and  $|v\rangle$ . In contrast to the simple cases discussed above, here no state of recall is ‘the’ state, rather, we must cue the subject and elicit a response from them before we can talk about a word being ‘recalled’ or ‘not recalled’. Indeed, a different cue might elicit a

very different response, and the quantum formalism could deal with this via a change of basis. Just as occurred in section 3, the coefficients of the states are related to the probability that a cue will elicit that response.

It is important to realise however, that (27) is *not* the only form of state that can be obtained from combination of  $|u\rangle$  and  $|v\rangle$  in the quantum formalism.

The other form of state, an *entangled state* is one that it is impossible to write as a product. As an example of an entangled state, we might consider the the state  $\psi$  where the words  $u$  and  $v$  are either both recalled, or both not recalled in relation to a cue  $q$ . One representation of this scenario is given by the following state:

$$\psi = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \quad (29)$$

This seemingly innocuous state is one of the so-called *Bell states* in QT. It is impossible to write as a product state, thus it differs markedly from (27). The fact that entangled systems cannot be expressed as a product of the component states makes them non-separable. More specifically, there are no coefficients which can decompose equation (29) into a product state exemplified by equation (27) which represents the two components of the system,  $u$  and  $v$ , as independent of one another. For this reason  $\psi$  is not written as  $|u\rangle \otimes |v\rangle$  as it can't be represented in terms of the component states  $|u\rangle$  and  $|v\rangle$ .

Entangled states do not occur in classical physics. They are responsible for many of the weird results of QT, including quantum nonlocality, Schrodinger cat states, the measurement problem and quantum cloning (23;43). Entangled states are widely held responsible for specific experimental outcomes that violate the predictions made by classical theories, such as the Bell inequalities (23), which means that entangled states lead to predictions that can be experimentally violated or verified in actual experiments. While there are classical theories capable of mimicing some of these results, no single classical theory exists that can predict all of the effects now seen in quantum experiments.

The existence of entanglement in quantum systems leads to a natural question for a quantum model of the human mental lexicon: can we find evidence of entanglement in human memory experiments? A positive answer to this question would have significant ramifications for models of human cognitive structures, as it would place the more standard models of cognitive behaviour in the difficult position of having to explain what is in essence a quantum effect. The fact that entangled states do not occur in classical physical models, suggests that they would be very difficult to create in cognitive models of this variety. Despite the importance of such a conclusive test, there are significant difficulties involved in constructing such experiments (some of which we have briefly discussed in section 4). Indeed, one of the main contributions of this paper is to develop the quantum model to the point where it can be tested. The remainder of this discussion will focus first upon an analysis of the Spooky-activation-at-a-distance formula in terms of entanglement and thereafter upon particular experimental scenarios that can be potentially deployed to test for the existence of quantum-like entanglement of words in human memory.

### 5.1 An analysis of Spooky-activation-at-a-distance in terms of entanglement

Nelson and McEvoy have recently begun to consider the Spooky-activation-at-a-distance formula in terms of quantum entanglement, "The activation-at-a-distance rule assumes that the target is, in quantum terms, entangled with its associates because of learning and practicing language in the world. Associative entanglement causes the studied target word to

simultaneously activate its associate structure” (29, p3). The goal of this section is to formalize this intuition. At the outset, it is important that the quantum model be able to cater for the set size and connectivity effects described at length in the introduction. Recall that both set size and associative connectivity have been demonstrated time and again as having robust effects on the probability of recall. Because the Spooky-activation-at-a-distance formula sums link strengths irrespective of direction, it expresses the intuition that a target with a large number of highly interconnected associates should translate into a high activation level during study.

Table 1 is a matrix representation of the associative network of the hypothetical target  $t$  shown in Figure 3. The bottom line of the matrix represents the summation of free association probabilities for a given word. Hence, free association probabilities may be added

$$p_{a_2} = \Pr(a_2|t) + \Pr(a_2|a_1) \quad (30)$$

$$= 0.1 + 0.6 \quad (31)$$

$$= 0.7 \quad (32)$$

as it is assumed that each free association experiment is independent.

The assumption behind both Spreading Activation and Spooky-activation-at-a-distance is that free association probabilities determine the strength of activation of a word during study, they only differ in the way this activation strength is computed. Viewing free association probabilities in matrix form allows us to consider the system in figure 3 as a many bodied quantum system modelled by three qubits. Figure 10 depicts this system as a set of qubits, each word is in a superposed state of being activated  $|1\rangle$  or not  $|0\rangle$ . Note how each summed column in table 1 with a non-zero probability leads to a qubit. For ease of exposition in the following analysis, we shall change variables.

The probabilities depicted in table 1 are related to the probability densities of figure 10 by taking their square root: e.g.  $\pi_t^2 = p_t$ . Using such a change of variables, the state of the target word  $t$  would be written as:

$$|t\rangle = \sqrt{\bar{p}_t}|0\rangle + \sqrt{p_t}|1\rangle = \bar{\pi}_t|0\rangle + \pi_t|1\rangle, \quad (33)$$

where the probability of recall due to free association is  $p_t = \pi_t^2$ , and  $\bar{p}_t = 1 - p_t = \bar{\pi}_t^2$  represents the probability of a word not being recalled. Thus, the states of the individual words are represented as follows in order to avoid cluttering the analysis with square root signs:

$$|t\rangle = \bar{\pi}_t|0\rangle + \pi_t|1\rangle \quad (34)$$

$$|a_1\rangle = \bar{\pi}_{a_1}|0\rangle + \pi_{a_1}|1\rangle \quad (35)$$

$$|a_2\rangle = \bar{\pi}_{a_2}|0\rangle + \pi_{a_2}|1\rangle \quad (36)$$

where  $\bar{\pi}_t = 1 - \pi_t$ ,  $\bar{\pi}_{a_1} = 1 - \pi_{a_1}$  and  $\bar{\pi}_{a_2} = 1 - \pi_{a_2}$ .

As detailed in the previous section, tensor products are used to model many bodied quantum systems. The state  $\psi_t$  of the most general combined quantum system is given by the tensor product of the individual states:

$$\psi_t = |t\rangle \otimes |a_1\rangle \otimes |a_2\rangle \quad (37)$$

$$\begin{aligned} &= (\bar{\pi}_t|0\rangle \\ &\quad + \pi_t|1\rangle) \otimes (\bar{\pi}_{a_1}|0\rangle \\ &\quad + \pi_{a_1}|1\rangle) \otimes (\bar{\pi}_{a_2}|0\rangle \\ &\quad + \pi_{a_2}|1\rangle) \end{aligned} \quad (38)$$

$$\begin{aligned} &= \bar{\pi}_t \bar{\pi}_{a_1} \bar{\pi}_{a_2} |000\rangle + \pi_t \bar{\pi}_{a_1} \bar{\pi}_{a_2} |100\rangle + \bar{\pi}_t \pi_{a_1} \bar{\pi}_{a_2} |010\rangle + \pi_t \pi_{a_1} \bar{\pi}_{a_2} |110\rangle \\ &\quad + \bar{\pi}_t \bar{\pi}_{a_1} \pi_{a_2} |001\rangle + \pi_t \bar{\pi}_{a_1} \pi_{a_2} |101\rangle + \bar{\pi}_t \pi_{a_1} \pi_{a_2} |011\rangle + \pi_t \pi_{a_1} \pi_{a_2} |111\rangle \end{aligned} \quad (39)$$

The introduction detailed show free association probabilities can be used to compute the strength of activation of target  $t$  during study. Hence,  $|111\rangle$  represents the state in which all respective qubits collapse onto their state  $|1\rangle$ . In other words,  $|111\rangle$  denotes the state of the system in which words  $t$ ,  $a_1$  and  $a_2$  have all been *activated* due to study of target  $t$ . The probability of observing this is given by the taking the square of the product  $\pi_t \pi_{a_1} \pi_{a_2}$ . Conversely, the state  $|000\rangle$  corresponds to the basis state in which none of the words have been activated.

The state  $\psi_t$  of the three-bodied system does not capture Nelson & McEvoy's intuition that the studied target word  $t$  simultaneously activates its entire associative structure. This intuition suggests target  $t$  activates its associates in synchrony; when target  $t$  is studied it activates *all* of its associates, or none at all. In order to model this intuition, the the state  $\psi_t$  is assumed to evolve into a Bell entangled state  $\psi'_t$  of the form:

$$\psi'_t = \sqrt{p_0}|000\rangle + \sqrt{p_1}|111\rangle \quad (40)$$

This formula expresses a superposed state in which the entire associative structure is activated ( $|111\rangle$ ) or not at all ( $|000\rangle$ ). The question remains how to ascribe values to the probabilities  $p_0$  and  $p_1$ . In QT these values would be determined by the unitary dynamics evolving  $\psi_t$  into  $\psi'_t$ . As such a dynamics is yet to be worked out for cognitive states, we are forced to speculate. One approach is to assume the lack of activation of the target is determined solely in terms of lack of recall of any of the associates. That is,

$$p_0 = \bar{p}_t \bar{p}_{a_1} \bar{p}_{a_2} \quad (41)$$



Consequently, the remaining probability mass contributes to the activation of the associative structure as a whole. Departing from the assumption of a Bell entangled state, the probability  $p_1$  corresponds to Nelson & McEvoy's intuition of the strength of activation of target  $t$  in synchrony with its associate:

$$p_1 = 1 - \bar{p}_t \bar{p}_{a_1} \bar{p}_{a_2} \quad (42)$$

$$= 1 - (1 - p_t)(1 - p_{a_1})(1 - p_{a_2}) \quad (43)$$

$$= 1 - (1 - p_t - p_{a_1} - p_{a_2} + p_t p_{a_1} + p_t p_{a_2} + p_{a_1} p_{a_2} - p_t p_{a_1} p_{a_2}) \quad (44)$$

$$= \underbrace{p_t + p_{a_1} + p_{a_2}}_A + \underbrace{p_t p_{a_1} p_{a_2}}_B - \underbrace{(p_t p_{a_1} + p_t p_{a_2} + p_{a_1} p_{a_2})}_C \quad (45)$$

Term  $A$  corresponds to the summation of the free association probabilities in the above matrix. In other words, term  $A$  corresponds exactly to the Spooky-activation-at-a-distance formula (See equation (4)). At such, the assumption of a Bell entangled state provides partial support for the summation of free association probabilities which is embodied by the Spooky-activation-at-a-distance equation. Ironically perhaps, term  $B$  corresponds to free association probabilities multiplied according to the directional links in the associative structure. This is expressed in the second term of the spreading activation formula (See equation (1)). In other words, departing from an assumption of entanglement leads to an expression of activation strength which combines aspects of both Spooky-activation-at-a-distance and Spreading Activation.

The third term  $C$  is more challenging to interpret. It arises because of the underlying structure of the tensor space and how probabilities are amassed in the initial product state and then in the Bell entangled state. When seen in the context of actual values, term  $C$  has a significant compensating effect:

$$p_1 = A + B - C \quad (46)$$

$$= (0.7 + 0.2 + 0.7) + (0.7 \times 0.2 \times 0.7) - (0.7 \times 0.2 + 0.7 \times 0.7 + 0.2 \times 0.7) \quad (47)$$

$$= 0.928 \quad (48)$$

It is interesting to note that the strength of activation  $p_1$  lies between Spreading Activation and Spooky-activation-at-a-distance. Based on a substantial body of empirical evidence, Nelson

and McEvoy have argued persuasively that spreading activation *underestimates* strength of activation. On the other hand, when departing from an assumption that the associative structure is Bell entangled, a preliminary hypothesis emergent from the above analysis is Spooky-activation-at-a-distance overestimates the strength of activation and it is term  $C$  which compensates for this.

We conclude this section with some remarks about how the entanglement model above accounts for the previously mentioned aspects of connectivity and set size. The more associates a target has, the more qubits are needed to model it. When these are tensored the resulting space will have higher dimensionality. Therefore a large set size is catered for by a tensor space of higher dimensionality. Conversely, interconnectivity is catered for by larger probabilities in the initial superposed states of the respective qubits, and this is represented in table 1. In the general case, when the associative structure is highly interconnected the sums of probabilities in the last row will tend to be higher. These will contribute to higher activation strength as they are summed (as defined by probability  $p_1$ ) in the same way as in Spooky-activation-at-a-distance. So it is possible to have a high activation strength even though the tensor space has high dimensionality. When the associative structure is not interconnected these probabilities will be low and hence lead to lower strength of activation.

In the remainder of this article we shall look at ways in which it might be possible to experimentally test for the existence of entangled cognitive states.

## 5.2 Bell inequalities for general systems

Tests for entanglement generally revolve around an assumption of separability. This assumption is used to generate testable criteria; if the system of interest is indeed separable then it will satisfy them, if not then a violation will ensue.

One of the most widely known criteria is given by the set of Bell inequalities (23). Here we shall look at a separability assumption involving two different experiments, which can be performed separately, or at the same time. We shall look for indications suggesting whether it is possible to consider the two different experiments individually, or if it might instead be necessary to consider them together, which would imply that the system is entangled and should be treated as an undivided whole.

Following (1) we shall construct Bell inequalities by considering some entity  $\psi$  and four experiments  $e_1, e_2, e_3, e_4$  that can be performed upon it. Each experiment has two possible outcomes, which we shall denote  $|1\rangle$  and  $|0\rangle$  respectively in keeping with the above formalism. Some of the different experiments can be combined, which leads to a number of *coincidence experiments*  $e_{ij}, i, j \in \{1, 2, 3, 4\}, i \neq j$ . Combinatorially, there are four outcomes possible from performing such a coincidence experiment,  $|11\rangle, |10\rangle, |01\rangle$ , and  $|00\rangle$ , just as occurred in the combination of the two quantum systems above. Note that there is no a priori reason to expect the results of the coincidence experiments to be compatible with an appropriate combination of the results of the two *separate* experiments, although the utility of such reductive assumptions in science has generally been considerable.

Expectation values sum the probabilities for the different coincidence experiments, weighted by the outcome itself, and they have already been introduced in equation (11). In the current scenario an expectation value is the sum of all the values that the two experiments can yield, weighted by the (experimentally obtained) probabilities of obtaining them. For the current setup, if a  $|11\rangle$  is obtained the value of the experiment is +1 (denoting a correlation between the two results), and the result is the same if a  $|00\rangle$  results, whereas if either  $|10\rangle$  or  $|01\rangle$  anticorrelation results are obtained then a -1 is recorded. Thus, for the experimental combination  $e_{ij}$  (with the  $ij$  denoting the choice of coincidence experiment):

$$\begin{aligned}
 E_{ij} = & +1.P(|11\rangle) \\
 & +1.P(|00\rangle) \\
 & - 1.P(|10\rangle) - 1.P(|01\rangle).
 \end{aligned}
 \tag{49}$$

These expectation values allow us to test the appropriateness of the reductive separability assumption that the combination  $e_{ij}$  of the two experiments  $e_i$  and  $e_j$  can be appropriately described by a product  $e_i e_j$ . This is done with the use of a Bell-type inequality:

$$|E_{13} - E_{14}| + |E_{23} + E_{24}| \leq 2 \tag{50}$$

which assumes in its derivation that the outcome of one sub-experiment in a coincidence pair does not affect the other and vice versa (1;23;43). If (50) is violated then this implies that the assumption is invalid, and that the coincidence experiment is fundamentally different from the combination of the two separate sub-experiments. This particular inequality is due to (10) and is known as the Clauser–Horne–Shimony–Holt (CHSH) inequality.

The preparation of an appropriate entangled state is key to all of the experimental procedures that follow in this section. To this end, subjects would take part in an extralist cueing procedure where a subject is first asked to study a list of words with the intention of the experiment not being disclosed. Each word is studied in isolation for a couple of seconds before the next study word is produced. Returning to the Bell state as exemplified in equation (29) we face the thorny question of whether it is possible to actually create a Bell-type state for words in human memory. One possibility for potentially creating such a state is to alter the preparation phase of the extralist cuing experimental framework. It was shown above how the senses of a word are reflected in the free associates of the word. Is it possible to turn this around and use the free associates to prepare a Bell-type entangled word state that could exhibit a violation of (50)? By way of illustration, consider once again the words *bat* and *boxer*. The study list would comprise two parts - each part relating to one of the senses of these two cue words. For example, the first part would contain those associates of *boxer* and *bat* related to the animal sense such as *dog, cave, vampire, night, blind*. The intention here is to correlate *boxer* and *bat* only within the animal sense, represented by the basis vector  $|11\rangle$  in equation (50). The second part of the study list is made up of associates related to the sport sense such as *gloves, punch, fight, ball, baseball*. The intention of the second part of the list is to correlate the words only within the sport sense, represented by the basis vector  $|00\rangle$ . By studying the associates of both words *within* a given sense, the hope is that the anti-correlations  $|10\rangle$  and  $|01\rangle$  will be negligible.

After all words are studied, the subject is presented with a cue word (not in the original preparation list) and asked to recall a single word, or words, from the list just studied. It is supposed that in some carefully controlled situations an entangled state is created by studying the target words. It can then be probed in a number of different ways. This section will now discuss two different experimental procedures that might be conducted to investigate such an entangled state.

### 5.3 Non-separability in semantic space

The first experimental procedure is intended to highlight the nonseparable nature of word meanings via a direct violation of a Bell-type CHSH inequality. We shall choose four sub-experiments, each of which arise from exposing a subject to a different cue ( $q$ ):

$$\begin{aligned}
 e_1:q &= \text{bat}, e_2:q \\
 &= \text{stick}, e_3:q \\
 &= \text{lady}, e_4:q = \text{animal}
 \end{aligned}
 \tag{51}$$

and then asking them to give a set of associations from the list of priming words that they have seen during the preparation phase. In the preparation phase for this experimental procedure a subject would be suitably prepared such that an entangled state of all three senses of the word *bat* (animal, sporty, strange old lady) might be expected, using the method depicted in the previous section.

The subject will then be told that they will see two cue words *simultaneously* in the same way as they were exposed to the priming words and asked to recall as many words as possible from the priming list, hence, one of the four  $e_{ij}$  experiments will be run. Thus, four coincidence experiments will be conducted, based upon the following four coincidence possibilities:

$$e_{13}:(\text{bat}, \text{lady}) \tag{52}$$

$$e_{14}:(\text{bat}, \text{animal}) \tag{53}$$

$$e_{23}:(\text{stick}, \text{lady}) \tag{54}$$

$$e_{24}:(\text{stick}, \text{animal}) \tag{55}$$

After the experiment has concluded each association word that was recalled will be examined for its agreement with the meaning of the experimental setup. By way of illustration, consider experiment  $e_{13}$ . A recalled word, e.g., *crazy* that correlates with both of the cuing words will be recorded as returning an  $|11\rangle$ , a word, e.g., *tree* that correlates with neither of the cuing words will be considered as returning as  $|00\rangle$ , a recalled word, e.g., *ball* appropriate to the first cue but not the second will be considered as a  $|10\rangle$ , and the converse case will lead to the recording of a  $|01\rangle$ , e.g., *perfume*.

It is expected that equation (50) may be violated when these results are counted and the appropriate expectation values obtained. This would show that it is not always possible to consider the combined experiment  $e_{ij}$  as equivalent to the product of the two separate experiments  $e_i e_j$  when one is combining word association data from human memory experiments.

#### 5.4 A direct entanglement effect

Consider once again the word *bat*, and assume for simplicity that it has two senses: the animal sense, and a sporting sense. As indicated earlier, this would mean that *bat* can be represented as the superposition:

$$|b\rangle = a_0|0\rangle + a_1|1\rangle \tag{56}$$

where  $|0\rangle$  now corresponds to the sport sense being recalled and  $|1\rangle$  to the animal sense being recalled in response to a given cue  $q$ . Some knowledge of the *bat* superposition can be recovered by studying free association norms using the word *bat* as cue. It turns out the the majority of the free associates are related to the animal sense of *bat*. However, the most probable associate *ball* ( $p(\text{ball}|\text{bat}) = 0.25$ ) relates to the sport sense (32). The word *boxer* also has a sport and animal sense, i.e., the breed of dog. Examination of the free association norms reveals the sport sense is heavily favoured via associates such as *fighter*, *gloves*, *fight*, *shorts*, *punch*. The sole free associate related to the animal sense ( $p(\text{dog}|\text{boxer}) = 0.08$ ) (32).

The question being probed in this section is whether the words *bat* and *boxer* can be prepared as an entangled cognitive state in way that is some ways analogous to entangled twin state photons. (Here *bat* and *boxer* are assumed to be ‘twins’ for the purpose of the experiment). The exhibition of a direct entanglement effect between words in memory relies on the hypothesis that the collapse of one such entangled word onto one sense in a subject’s cognitive state results in the collapse of the other word onto that same sense.

Assuming that something akin to a Bell state can be prepared in human memory, the next step is to devise an experiment so the CHSH inequality of equation (50) can be applied. Unlike the experiment of the previous section, the coincidence experiments are realised by cuing a subject *twice* after the study period. Example cues are as follows:

$$\begin{aligned} e_1:q=\text{black, bat}, e_2:q \\ =\text{bat}, e_3:q \\ =\text{boxer}, e_4:q=\text{black, boxer} \end{aligned} \quad (57)$$

The subjects are asked to recall the first word from the study list that comes to mind in response to a cue. Four coincidence experiments are set out as follows:

$$e_{13}:(\text{black bat}), (\text{boxer}) \quad (58)$$

$$e_{14}:(\text{black bat}), (\text{black boxer}) \quad (59)$$

$$e_{23}:(\text{bat}), (\text{boxer}) \quad (60)$$

$$e_{24}:(\text{bat}), (\text{black boxer}) \quad (61)$$

A given subject would take part in one experiment and be ‘measured’ by two cues, one followed by the other. For example, in a coincidence experiment  $e_{13}$ , a given subject would be first cued with *black bat* and a word then recalled, say ‘vampire’. In this case observe how the recalled word reflects the animal sense of *bat*. This is not surprising as the context word *black* could easily promote the animal sense, though note this is not a certainty as baseball bats are often black. This is a deliberate design which is akin to setting a polarizer in a certain direction. What happens next is potentially more interesting. The same subject would then be presented with the cue *boxer* and a word recalled, say ‘dog’. In this example, the recalled word reflects the animal sense of *boxer*. The supposition here is that the collapse of *bat* onto the animal sense has influenced the collapse of *boxer* onto the animal sense, and this influence has occurred in

the face of an a priori strong tendency towards the sport sense of *boxer*. It is this influence which is central to entanglement and is the essence of why it is considered so weird and surprising. In this connection, it is important that the context word associated with a second cue be chosen more or less neutrally so the influence of the first cue is not washed out by a local context effect when the second cue is activated. Note that the second cue *black boxer* is ambiguous and leaves room for both the animal and sport senses to manifest.

The above experiment would be run using  $n$  subjects. The subjects would be divided into four groups G1, G2, G3, G4 of equal size. All subjects would be prepared in the same fashion. Subjects in G1 would be given experiment  $e_{13}$ , subjects in G2 given  $e_{14}$ , subjects in G3 given  $e_{23}$  and subjects in G4 given  $e_{24}$ . The expectations of the experiments can be calculated using equation (49). For example, calculating  $E_{13}$  requires the estimation of  $P(|11\rangle)$ . This can be computed by counting the number of subjects in G1 which collapsed the state onto the animal sense in response to the cues *black bat* and *boxer*. This value is then divided by  $\frac{n}{4}$ , the size of G1. Similarly for the other three cases:  $|00\rangle$ ,  $|10\rangle$ , and  $|01\rangle$ .

What would it mean if the inequality of equation (50) is violated? As described in the previous section, it suggests that *bat* and *boxer* are inseparable in the cognitive state resulting from the preparation. Certainly the preparation proposed above is highly artificial, however, it is likely that in certain circumstances context acts like the preparation procedure above, yielding something like a Bell state in memory. For example, in general the words *Reagan* and *North* are semantically distant the human mental lexicon, however, according to the intuition above, seeing *Reagan* in the context of *Iran* leads the collapse of *Reagan* onto a basis state (sense) of President Reagan dealing with the Iran-Contra scandal which in turn may influence the collapse of *North* onto the Iran-Contra basis state, i.e., ‘Oliver North’, who was a central figure in the scandal.

## 6 Summary and Outlook

This article attempts to take the initial steps in developing a model of the human mental lexicon based on QT. At the outset it is important to note it makes no claim in regard to quantum processes in the brain, rather, QT is seen as an abstract framework which has encouraging potential to model the human mental lexicon. It has been shown how words in memory can be viewed as superposed states in a Hilbert (vector) space. Cues are akin to a measuring device the orientation of which is formalized as basis vectors. Different cues are modelled as measuring devices that have different orientations. Interactions between words are modelled as tensor products of their respective Hilbert spaces. In this way, QT is used as a metaphor, and from a formal perspective it is similar to vector based models of the human mental lexicon, such as tensor based models(17;18;42). The main differences of this work in relation to previous vector based models of memory reside in the issues of contextuality and non-separability.

In the memory literature, context is sometimes represented as a vector such as in the matrix model of memory (17) and more recently in a holographic model of the human mental lexicon (20). In the quantum model presented in this article, context is essentially modelled just like it is in quantum physics; the context of the experiment is a particular choice and orientation of a measuring apparatus. This boils down to a particular choice of basis. We have shown a spectrum of bases, for example, a two dimensional basis corresponding to ‘recalled’, ‘not recalled’ to a bases whereby the basis vectors correspond to associates, or senses, of the given word being modelled. QT is contextual and this is a subtle issue. By way of illustration, if we observe the hair colour of a person, it is not dependent on other other observations we may make, such as height, or weight. Generally, in QT, the outcome of an observation will depend on other observations that are made. In other words, the context of an observation is crucial.



Contextuality, in turn is related to non-separability. Bodies in a quantum system may become entangled which, broadly speaking, means the respective bodies cannot be treated as separate systems. A key contribution of this article is to put forward the speculation that words in memory may be non-separable, which to our knowledge, no existing models in the literature have taken into account. In this connection, it is also important to bear in mind that QT is the only theory which models non-separability. We have shown by assuming entanglement of words in memory, there is some theoretical justification for the Spooky-activation-at-a-distance-model (31). In addition, the analysis places the the Spreading Activation Model and Spooky-activation-at-a-distance model in theoretical perspective regarding their respective ability to estimate activation strength of words during study. This perspective has led to the hypothesis that Spooky-activation-at-a-distance may be overestimating activation strength whereas Spreading Activation underestimates it.

This article has also proposed two experimental frameworks as a means of testing for the existence of entanglement effects between words. The existence of such effects would strongly support the claim that the quantum formalism can be used to model human memory experiments. Should strong evidence of the entanglement of words appear, it is hard to predict what the consequences of such a discovery might be. From a philosophical point of view, current reductionist models of human memory would be seriously undermined; words just cannot be considered as isolated entities in memory. More practically, one wonders whether entanglement of words may be somehow leveraged, just as in quantum computing where entanglement is seen as a resource to be exploited. Some preliminary thoughts in this direction centre around whether entanglement of words may be leveraged for knowledge discovery (41).

Clearly, the application of the QT beyond its standard domain will be controversial, but it is hoped that this article has been able to identify some intriguing new directions for modelling the human mental lexicon.

## Acknowledgments

This project was supported in part by the Australian Research Council Discovery grant DP0773341 to P. Bruza and K. Kitto, and by grants from the National Institute of Mental Health to D. Nelson and the National Institute on Aging to C. McEvoy.

## References

1. Aerts D, Aerts S, Broekaert J, Gabora L. The Violation of Bell Inequalities in the Macroworld. *Foundations of Physics* 2000;30:1387–1414.
2. Aerts, D.; Broekaert, J.; Gabora, L. A Case for Applying an Abstracted Quantum Formalism to Cognition. In: Bickhard, MH.; Campbell, R., editors. *Mind in Interaction*. John Benjamins; Amsterdam: 2005.
3. Aerts D, Gabora L. A Theory of Concepts and Their Combinations II: A Hilbert Space Representation. *Kybernetes* 2005;34:176–205.
4. Ballentine, L. *Quantum Mechanics: A modern development*. World Scientific; 1998.
5. Bell, JS. *Speakable and unspeakable in quantum mechanics*. Cambridge University Press; Cambridge: 1987. *Six possible worlds of quantum mechanics*; p. 181-195.
6. Bruza, P.; Cole, R. Quantum Logic of Semantic Space: An Exploratory Investigation of Context Effects in Practical Reasoning. In: Artemov, S.; Barringer, H.; Sd'Avila Garcez, A.; Lamb, LC.; Woods, J., editors. *We Will Show Them: Essays in Honour of Dov Gabbay*. Vol. 1. London: College Publications; 2005. p. 339-361.
7. Bruza, PD.; Lawless, W.; van Rijsbergen, CJ.; Sofge, D., editors. *Proceedings of the AAAI Spring Symposium on Quantum Interaction*; March 27–29; Stanford University; AAAI Press; 2007.

8. Bruza, PD.; Lawless, W.; van Rijsbergen, CJ.; Sofge, D., editors. Proceedings of the second conference on Quantum Interaction; March 26–28; London. College Publications; 2008.
9. Bruza, PD.; Widdows, D.; Woods, JA. Quantum Logic of Down Below. In: Engesser, K.; Gabbay, D.; Lehmann, D., editors. Handbook of Quantum Logic and Quantum Structures. Vol. 2. Elsevier; 2007.
10. Clauser JF, Horne MA, Shimony A, Holt RA. Proposed experiment to test local hidden-variable theories. *Physical Review Letters* 1969;23:880–884.
11. Collins AM, Loftus EF. A spreading-activation theory of semantic processing. *Psychological Review* 1975;82(6):407–428.
12. Conte E, Todarello O, Federici A, Vitiello F, Lopane M, Khrennikov A, Zbilut J. Some remarks on an experiment suggesting quantum-like behaviour of cognitive entities and formulation of an abstract quantum mechanical formalism to describe cognitive entity and dynamics. *Chaos, Solitons and Fractals* 2007;33:1076–1088.
13. Einstein A, Podolsky B, Rosen N. Can quantum-mechanical description of physical reality be considered complete? *Physical Review* 1935;47:777–780.
14. Gabora L, Aerts D. Contextualizing concepts using a mathematical generalization of the quantum formalism. *Journal of Experimental and Theoretical Artificial Intelligence* 2002;14:327–358.
15. Gabora L, Rosch E, Aerts D. Toward an ecological theory of concepts. *Ecological Psychology* 2008;20:84–116.
16. Gee NR, Nelson D, Krawczyk D. Is the connectedness effect a result of underlying network interconnectivity? *Journal of Memory and Language* 1999;40(4):479–497.
17. Humphreys MS, Bain JD, Pike R. Different ways to cue a coherent memory system: A theory for episodic, semantic and procedural tasks. *Psychological Review* 1989;96:208–233.
18. Humphreys MS, Pike R, Bain JD, Tehan G. Global matching: A comparison of the SAM, Minerva II, Matrix and TODAM models. *Journal of Mathematical Psychology* 1989;33:36–67.
19. Isham, CJ. Lectures on Quantum Theory. Imperial College Press; London: 1995.
20. Jones MN, Mewhort DJK. Representing word meaning and order information in a composite holographic lexicon. *Psychological Review* 2007;114(1):1–37. [PubMed: 17227180]
21. Khrennikov, A. Interpretations of Probability. VSP; Utrecht: 1999.
22. Kitto K. High End Complexity. *International Journal of General Systems* 2008;37(6):689–714.
23. Laloë F. Do we really understand quantum mechanics? Strange correlations, paradoxes, and theorems. *American Journal of Physics* 2001;69(6):655–701.
24. Nelson D, Bennett DJ, Leibert T. One step is not enough: Making better use of association norms to predict cued recall. *Memory & Cognition* 1997;25(6):785–796.
25. Nelson D, Gee NR, Schreiber TA. Sentence encoding and implicitly activated memories. *Memory & Cognition* 1992;20(6):643–654.
26. Nelson D, Goodman LB. Experiencing a word can prime its accessibility and its associative connections to related words. *Memory & Cognition* 2002;30:380–398.
27. Nelson D, McEvoy CL. Encoding context and set size. *Journal of Experimental Psychology: Human Learning and Memory* 1979;5(3):292–314.
28. Nelson, D.; McEvoy, CL. Implicitly activated memories: The missing links of remembering. In: Izawa, C.; Ohta, N., editors. *Learning and Memory: Advances in Theory and Applications*. Erlbaum; New Jersey: 2005. p. 177-198.
29. Nelson, D.; McEvoy, CL. Entangled associative structures and context. In: Bruza, PD.; Lawless, W.; van Rijsbergen, CJ.; Sofge, D., editors. Proceedings of the AAAI Spring Symposium on Quantum Interaction. AAAI Press; 2007.
30. Nelson D, McEvoy CL, Janczura GA, Xu J. Implicit memory and inhibition. *Journal of Memory and Language* 1993;32:667–691.
31. Nelson D, McEvoy CL, Pointer L. Spreading activation or spooky action at a distance? *Journal of Experimental Psychology: Learning, Memory and Cognition* 2003;29(1):42–52.
32. Nelson D, McEvoy CL, Schreiber T. The University of South Florida, word association, rhyme and word fragment norms. *Behavior Research Methods, Instruments & Computers* 2004;36:408–420.

33. Nelson D, McKinney VM, Gee NR, Janczura GA. Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review* 1998;105(2):299–324. [PubMed: 9577240]
34. Nelson D, Schreiber TA, McEvoy CL. Processing implicit and explicit representations. *Psychological Review* 1992;99(2):322–348. [PubMed: 1594728]
35. Nelson, DL.; McEvoy, CL. Entangled Associative Structures and Context. Bruza, et al., editors.
36. Schrödinger E. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften* 1935;23:807.
37. Simpson, G. Context and the processing of ambiguous words. In: Gernsbacher, MA., editor. *Handbook of Psycholinguistics*. Academic Press; 1994. p. 359-374.
38. Steyvers M, Tenenbaum JB. Graph theoretic analysis of semantic networks: Small worlds in semantic networks. *Cognitive Science* 2005;29(1):41–78.
39. Vitiello, G. *My Double Unveiled*. John Benjamins Publishing Company; Amsterdam: 2001.
40. Widdows, D. *Geometry and Meaning*. CSLI Publications; 2004.
41. Widdows, D.; Bruza, PD. Quantum information dynamics and open world science. In: Bruza, PD.; Lawless, W.; van Rijsbergen, CJ.; Sofge, D., editors. *Quantum Interaction*. AAAI Press; 2007. AAAI Spring Symposium Series
42. Wiles, J.; Halford, GS.; Stewart, JEM.; Humphreys, MS.; Bain, JD.; Wilson, WH. Tensor Models: A creative basis for memory and analogical mapping. In: Dartnall, T., editor. *Artificial Intelligence and Creativity*. Kluwer Academic Publishers; 1994. p. 145-159.
43. Zeilinger A. Experiment and the foundations of quantum physics. *Reviews of Modern Physics* 1999;71:288–297.

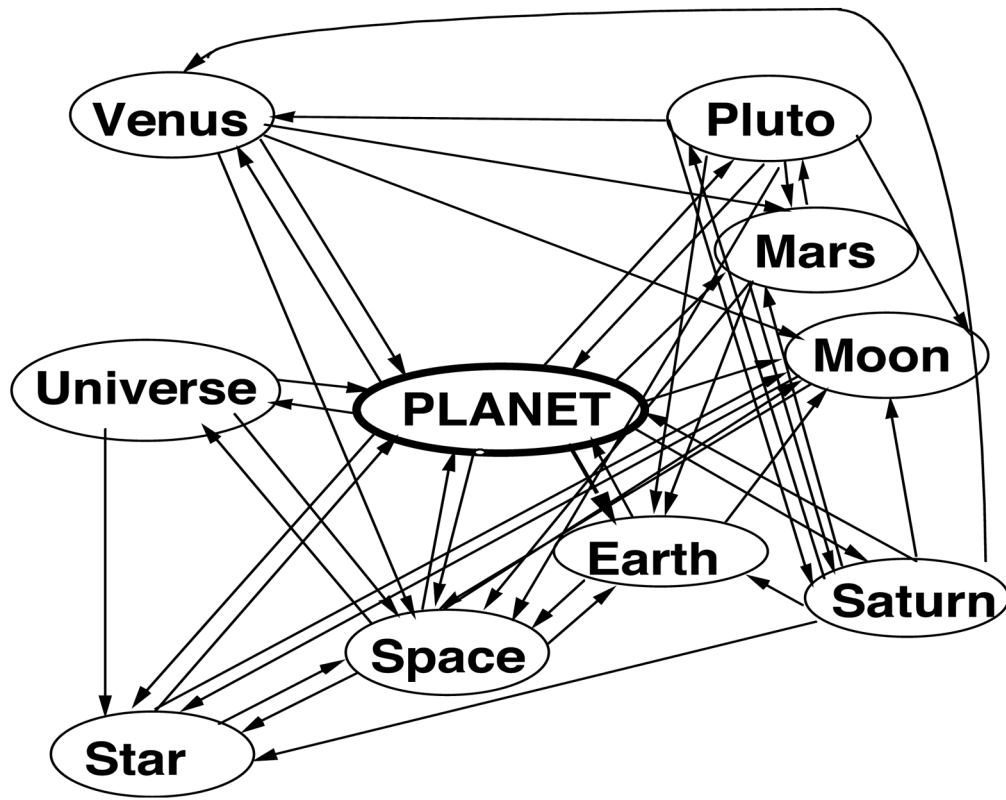
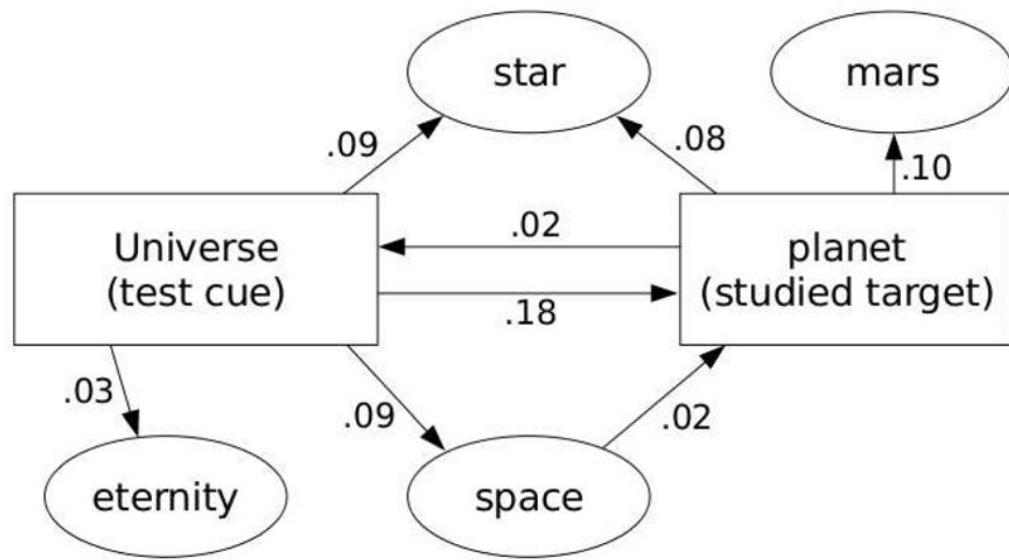
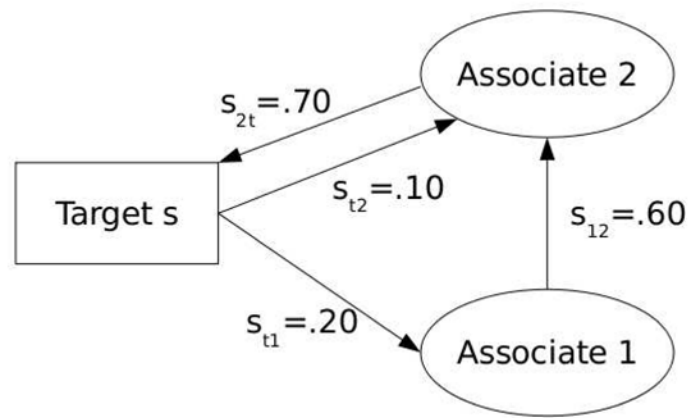


Fig. 1.  
Planet's associative structure

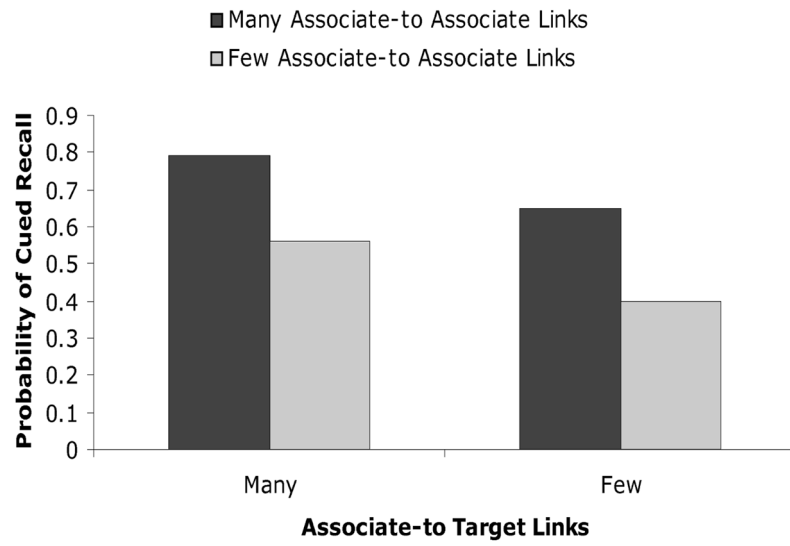


**Fig. 2.** Links that join the test cue and target and competing associates that do not. Adapted from (28).

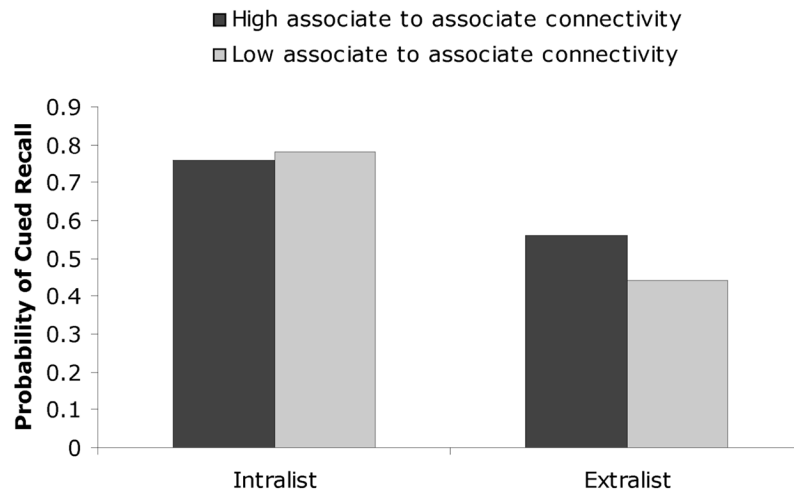


**Fig. 3.** A hypothetical target with two associates and single associate-to-target and associate-to-associate links. From Nelson, McEvoy, and Pointer (31).

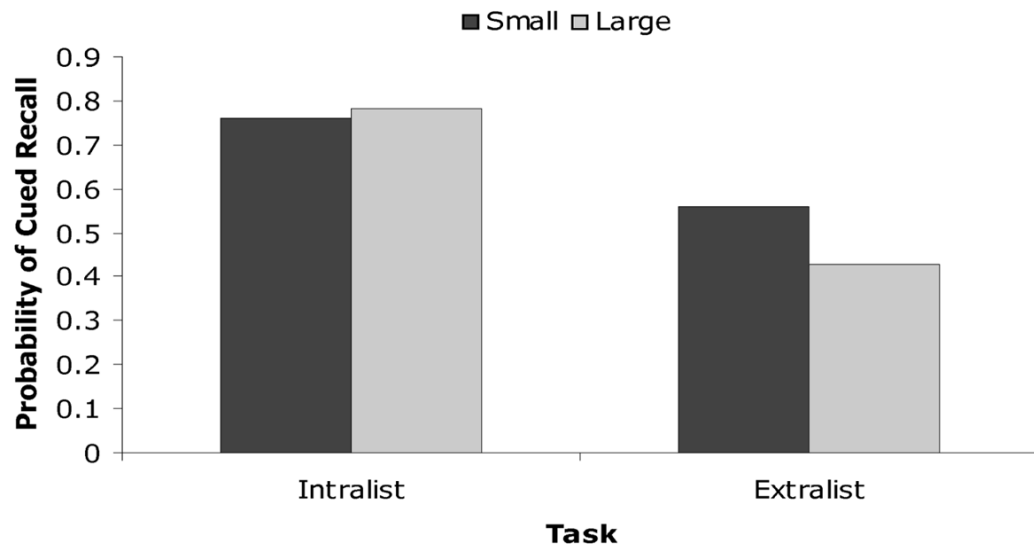




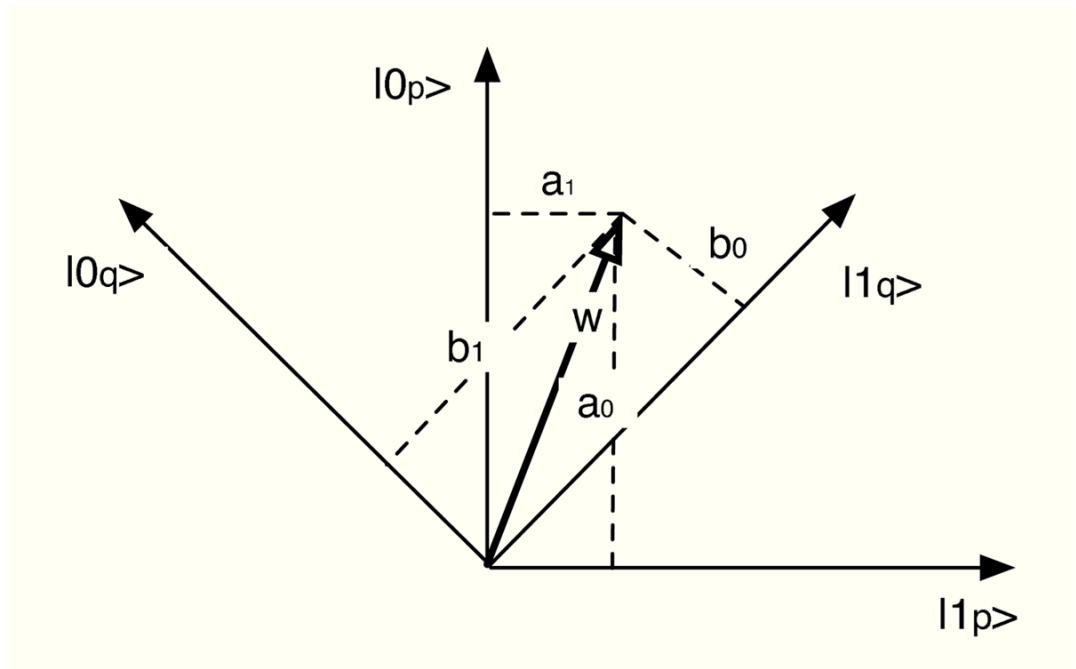
**Fig. 4.** Probability of cued recall as a function of the numbers of associate-to-target and associate-to-associate links.



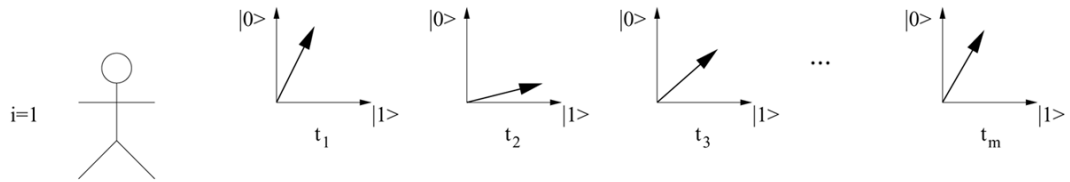
**Fig. 5.** Effects of associate-to-associate connectivity as a function of cuing task (adapted from (30))



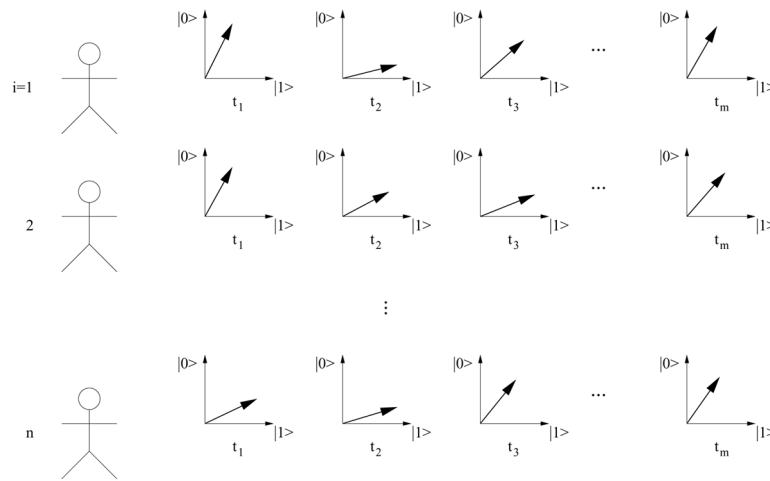
**Fig. 6.**  
Target set size effects as a function of cuing task (adapted from (30))



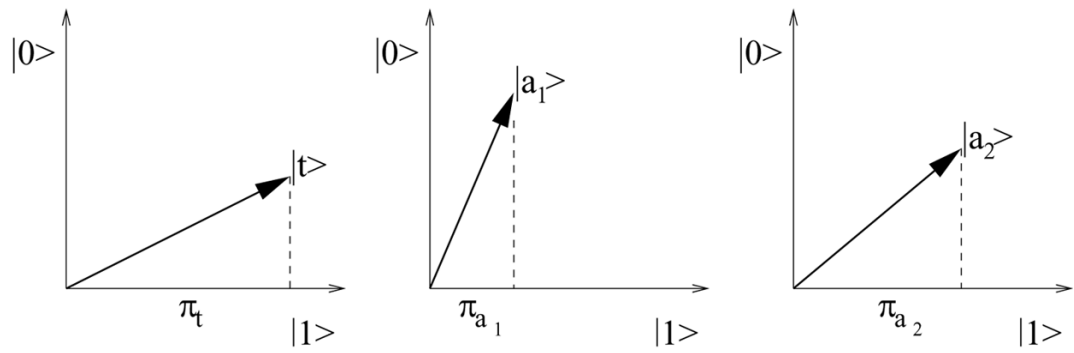
**Fig. 7.**  
Word vector  $w$  with respect to two different bases



**Fig. 8.** One subject's cognitive state modelled as a set of studied targets each modelled as a  $q$ -bit which is in a superposition of 'recalled' and 'not recalled' where the direction of the arrows identifies how likely each of these outcomes are.



**Fig. 9.** An ensemble of subjects each with a different cognitive state arising from their historical context. Upon summing appropriately this combined state can be represented as a density matrix  $\pi$ .



**Fig. 10.** Three bodied quantum system of words. The projection of the qubit onto the  $|1\rangle$  basis relates to the probabilities in table 1 via the change of variables below.



**Table 1**

Matrix corresponding to hypothetical target shown in Figure 3. Free associations probabilities are obtained by finding the row of interest (the cue) and running across to the associate word obtained.

	$t$	$a_1$	$a_2$
$t$		0.2	0.1
$a_1$			0.6
$a_2$	0.7		
	$p_t = 0.7$	$p_{a_1} = 0.2$	$p_{a_2} = 0.7$