A probabilistic framework for analysing the compositionality of conceptual combinations

Peter D. Bruza [a,*], Kirsty Kitto [a], Brentyn J. Ramm [b], Laurianne Sitbon [c]

[a] *Information Systems School, Queensland University of Technology, GPO Box 2434, Brisbane, Australia*
[b] *School of Philosophy, Australian National University, Australia*
[c] *School of Electrical Engineering and Computer Science, Queensland University of Technology, Australia*

HIGHLIGHTS

- Framing the principle of semantic compositionality in terms of probabilistic models.
- Formal methods for determining compositional and non-compositional semantics.
- Illustrates the formal methods on an empirical study 24 conceptual combinations.

ARTICLE INFO

ABSTRACT

Conceptual combination performs a fundamental role in creating the broad range of compound phrases utilised in everyday language. While the systematicity and productivity of language provide a strong argument in favour of assuming compositionality, this very assumption is still regularly questioned in both cognitive science and philosophy. This article provides a novel probabilistic framework for assessing whether the semantics of conceptual combinations are compositional, and so can be considered as a function of the semantics of the constituent concepts, or not. Rather than adjudicating between different grades of compositionality, the framework presented here contributes formal methods for determining a clear dividing line between compositional and non-compositional semantics. Compositionality is equated with a joint probability distribution modelling how the constituent concepts in the combination are interpreted. Marginal selectivity is emphasised as a pivotal probabilistic constraint for the application of the Bell/CH and CHSH systems of inequalities (referred to collectively as Bell-type). Non-compositionality is then equated with either a failure of marginal selectivity, or, in the presence of marginal selectivity, with a violation of Bell-type inequalities. In both non-compositional scenarios, the conceptual combination cannot be modelled using a joint probability distribution with variables corresponding to the interpretation of the individual concepts. The framework is demonstrated by applying it to an empirical scenario of twenty-four non-lexicalised conceptual combinations.

© 2015 Published by Elsevier Inc.

1. Introduction

Humans frequently generate novel associates when presented with unfamiliar conceptual combinations. For example, in free association experiments, subjects frequently produce the associate "slave" when cued with the compound "pet human" (Ramm, 2000), but neither "pet" nor "human" will have the same effect when presented individually (Nelson, McEvoy, & Schreiber, 2004). Such cases have sometimes been used to argue that conceptual combinations have a non-compositional semantics, as it is difficult to explain how the novel free associate "slave" can be recovered from its constituent concepts, however, this is a controversial position; within cognitive science, the question of how to represent even single concepts is still being debated.

Different positions have been put forward, including the prototype view, the exemplar view, and the theory view. Murphy (2002) contrasts these positions, asking which is most supported by the various aspects of cognition related to conceptual processing, e.g., learning, induction, lexical processing and conceptual understanding in children. He concludes, somewhat disappointingly,

* Corresponding author.
  E-mail addresses: p.bruza@qut.edu.au (P.D. Bruza), kirsty.kitto@qut.edu.au (K. Kitto), brentynramm@gmail.com (B.J. Ramm), laurianne.sitbon@qut.edu.au (L. Sitbon).

that "there is no clear, dominant winner". Moreover, there is a well documented tension in cognitive science between the compositionality and the prototypicality of concepts, which is difficult to reconcile (Fodor, 1998; Frixione & Lieto, 2012). Arguments in favour of compositionality centre around the systematicity and productivity of language; there are infinitely many expressions in natural language and yet our cognitive resources are finite. Compositionality ensures that this infinity of expressions can be processed because an arbitrary expression can be understood in terms of its constituent parts. Since compositionality is what explains systematicity and productivity, Fodor (1998) claims that concepts are, and must be compositional, however, such claims are at odds with well-known prototypicality effects (Fodor, 1998; Frixione & Lieto, 2012). For example, consider the conceptual combination PET FISH. A GUPPY is not a prototypical PET, nor a prototypical FISH, and yet a GUPPY is a very prototypical PET FISH (Hampton, 1997). Therefore, it is hard to imagine how the prototype of PET FISH can result from some composition of the prototypes of PET and FISH, which makes the characterisation of concepts in prototypical terms difficult to reconcile with compositionality (Fodor, 1998; Hampton, 1997). This supports a view put forward by the philosopher Weiskopf (2007) when he observed that conceptual combinations are "highly recalcitrant to compositional semantic analysis", but even this observation has garnered no general support.

Here, we approach the problem of non-compositionality from a novel perspective. We shall show that a suite of sophisticated tools have already been developed for analysing non-compositionality, albeit in another field of science. These tools can be naturally extended to the analysis of concepts, and provide theoretically justified grounds for deciding whether a particular conceptual combination can be considered in terms of the semantics of its constituent parts. Specific cases will be discussed where conceptual combinations can be shown to be non-compositional using these analytical methods. We begin with a brief review of conceptual combination as it is currently understood in cognitive science.

1.1. Cognitive theories, compositionality and conceptual combination

The principle of compositionality states that the meanings of higher order expressions such as sentences are determined from a combination of the meanings of their constituent parts (Costello & Keane, 2000; Mitchell & Lapata, 2010). This is a principle underlying many general theories of language, both natural and artificial, despite the fact that there is considerable debate about what the principle actually specifies (e.g., notions of both strong and weak compositionality appear in the literature (Pelletier, 1994)). A compositional account of conceptual combination is closely related to the notion that concepts are atomic in nature, but this assumption of atomicity is difficult to maintain when the full variety of possible semantic behaviour is considered.

Perhaps most supportive of the principle are those combinations that have an intersective semantics, e.g., the meaning of BLACK CAT is the intersection of black objects and objects that are cats. Here, it is possible to apply a conjunction operator between the two predicates referring to the constituent concepts, i.e., $black(x) \land cat(x)$. Such intersective semantics are compositional, as the semantics of BLACK CAT are determined solely in terms of the semantics of the constituent concepts BLACK and CAT. It is tempting to assume that most conceptual combinations can be modelled in this way, however, the study of intersective combinations in cognitive science has revealed that not all conceptual combinations display such intersective semantics (Hampton, 1997). For example, the intersection of ASTRONAUT and PEN in the combination ASTRONAUT PEN is empty, and therefore its semantics are

vacuous, despite its being a conceptual combination that humans can easily comprehend (Gärdenfors, 2000; Weiskopf, 2007).

A second type of conceptual combination arises when the first concept modifies the head concept, e.g., in CORPORATE LAWYER, CORPORATE modifies the more general head concept to give a sub-category of LAWYER. Schema-based theories of conceptual combination (Murphy, 1988; Wisniewski, 1996) propose that the head concept is a schema-structure made up of various property dimensions (e.g., colour, size, shape, etc.) and relational dimensions (e.g., habitat, functions, behaviours, etc.). Several studies have revealed that modification can produce emergent properties (e.g., in HELICOPTER BLANKET the modification of BLANKET by HELICOPTER generates associate properties such as "water proof", "camouflage", and "made of canvas"), a phenomenon which present theories struggle to account for (Wilkenfeld & Ward, 2001), such behaviour is sometimes viewed as evidence for non-compositional semantics (Hampton, 1997; Medin & Shoben, 1988).

Despite these tensions underlying the assumption of compositionality, virtually all researchers have at least assumed a weak form of compositionality in their analysis of human language, where for example, the initial combination process begins with separate meanings, but is supplemented later by external contextual information (Swinney, Love, Walenski, & Smith, 2007; Wisniewski, 1996). For example, in Wisniewski (1996)'s dual process theory of conceptual combination, a competition occurs between the processes of relation linking (e.g., ZEBRA CROSSING as a crossing for zebras), and property mapping (e.g., ZEBRA CROSSING as a striped crossing), as the meaning of the compound is decided upon. This process is affected by the similarity of the constituent concepts, because similar concepts share many facets and so are more likely to result in a property interpretation, whereas dissimilar concepts are more likely to be combined using a relational process. Thus, ELEPHANT HORSE is more likely to result in a property interpretation (e.g., a large horse), than ELEPHANT BOX, which is more likely to result in a relational interpretation (e.g., a box for holding elephants). This is because similar concepts share many dimensions (e.g. four legs, similar shape, etc. in the case of elephant and horse) and thus are easier to combine by mapping one property to another. However, it is important to note that these processes are all weakly compositional, in the sense that they rely almost exclusively on the properties of the individual concepts. It is only later that background knowledge is drawn upon to infer the possible emergent properties of the new concept. We see people making assumptions that an ELEPHANT BOX is likely to be made of a strong material such as wood, and hopefully to contain airholes. Swinney et al. (2007) found evidence for this form of weak compositionality in conceptual combination, when they showed that for adjectival combinations such as BOILED CELERY the properties of the individual words such as "green" are activated before emergent properties such as "soft". However, for the combination APARTMENT DOG, apartment modifies the "habitat" dimension of dog rather than its "size" (a dog the size of an apartment), which in turn shows that background knowledge also plays a role in early combinatory processes such as slot selection (Murphy, 1988).

Rather than entering this long running debate about the proper dividing line between weak and strong compositionality, it is our intention to provide a formal framework to analyse the (non-)compositionality of conceptual combinations, motivated by the analysis of composite systems in quantum physics. Importantly, this framework can be empirically tested. Thus, we feel that it is possible to shift the above largely theoretical debate onto a more experimental footing,[1] and this article is a step in that direction.

---

[1] In much the same way as the field of physics entered the realms of experimental testing with the work of Bell and Aspect, after decades of more philosophical debate as to the separability and completeness of the quantum formalism (Isham, 1995; Laloë, 2001).

In what follows we shall discuss the combination of concepts within a tiered model of cognition. This will provide a framework from which a (non-)compositional semantics can be developed in further sections.

2. Probabilistic approaches to modelling conceptual combinations

It is at the symbolic level of cognition where a significant portion of the work on compositional semantics can be placed, because this is where higher order symbolic structures and associated rules, such as grammar, are processed. A grammar specifies the parts of a sentence, and the manner in which they fit together. It makes sense that the semantics attributed to these primitive parts be intuitive, for example, a noun may be mapped to a set of entities. However, Zadrozny (1992) has suggested that it does not actually matter which components are chosen as primitive, a function can be found that will *always* produce a compositional semantics. In Zadrozny's own words, "..compositionality, as commonly defined, is not a strong [enough] constraint on a semantic theory". The consequence of this with respect to the compositional semantics of natural language, and hence conceptual combination, is that meaning need not be assigned to individual words, "we can do equally well by assigning meaning to phonemes or even LETTERS…" (Zadrozny, 1992). This raises the question about how to appropriately define the semantics of the language constructs being composed. It turns out that this is not a straightforward question to answer. In a well cited and detailed chapter about lexical semantics and compositionality, Partee (1995) highlights that at the outset there are disagreements about whether semantics can best be viewed from the point of view of mathematics or psychology, a debate that is yet to reach a resolution. For the purposes of this article we chose to approach the problem of how to semantically represent concepts from the point of view of psychology, using free association norms to ground our models.

Consider the concept BAT. One reliable way to seek an understanding of this concept is via free association experiments where subjects are cued with the word "bat" and asked to produce the first word that comes to mind (Nelson, Kitto, Galea, McEvoy, & Bruza, 2013). Over large numbers of subjects, probabilities can be calculated that a certain associate is produced. Fig. 1(a) depicts a subset of data taken from the University of South Florida word association norms (USF-norms) (Nelson et al., 2004). Upon examination of this table, we can see that these probabilities represent two clear senses for the cue "bat"; a SPORT sense (with relevant associates in bold) and an ANIMAL sense. Considering the full dataset[2] allows us to generate the total probability p_s of recall for the sport sense by summing the probabilities of the relevant associates: $p_s = 0.25 + 0.05 = 0.30$. The rest of the associates all happen to be relevant to the animal sense of bat, so $p_a = 0.70$.

The same can be said for the concept BOXER (see Fig. 1(b) where, once again, the associates relevant to the sport sense of BOXER are in bold).

Having constructed a model of the individual concepts, we might ask how the conceptual combination BOXER BAT will be interpreted by a subject. Four interpretations are possible within this scenario. For example, when BOXER is interpreted as a sport and BAT as an animal, the corresponding interpretation of the combination might be something along the lines of a "furry black animal with boxing gloves on", or perhaps BOXER could be interpreted as a sport and BAT as a sport leading a subject to interpret the compound as "a fighter's implement".

a	*Associate*	*Probability*
	ball	**0.25**
	cave	0.13
	vampire	0.07
	fly	0.06
	night	0.06
	baseball	**0.05**
	bird	0.04
	blind	0.04
	animal	0.02
	. . .	. . .

b	*Associate*	*Probability*
	fighter	**0.14**
	gloves	**0.14**
	fight	**0.09**
	dog	0.08
	shorts	**0.07**
	punch	**0.05**
	Tyson	**0.05**
	. . .	. . .

Fig. 1. (a) Free association probabilities for the word "bat" (a) and the word "boxer" (b).

Conceptual combinations usually have more than one possible interpretation. This may arise from a range of factors, including the meaning of the concepts themselves (e.g. BOXER can be interpreted as a dog, a sportsperson, a pair of shorts, someone who puts things in boxes, etc.), the sentence in which they appear, the background of the subject, etc. Different human subjects will often interpret the same conceptual combination differently, indeed, the same human subject, if placed in a new context may very well provide a different interpretation for the same concept. Thus, it is sensible to approach the analysis of compositionality probabilistically.

In what follows each concept is assumed to have a *dominant* sense and one or more *subordinate* senses. The distinction between the two can be inferred from free association norms such as those discussed above. For example, the dominance of the sport sense of BOXER is clearly evident in Fig. 1(a), where the probability associated with the sport sense is greater than the animal sense, which leads us to designate the sport sense as "dominant" and the animal sense as "subordinate". It should be noted, however, that the distinction between "dominant" and "subordinate" senses is not necessary for the theory presented below, rather it is an explanatory aid.

Standard probabilistic reasoning suggests that if two ambiguous concepts A and B have behaviour that can be considered as compositional, then it should be possible to describe this behaviour in terms of sets of dichotomous random variables. In the dominant/subordinate scenario introduced above this would lead to four random variables {**A1**, **A2**} and {**B1**, **B2**}, ranging over two values {+1, −1}, where we have used the numbers 1 (dominant) and 2 (subordinate) to refer to the senses that are used to prime to their respective concepts A and B in an experiment.

Now, if a human subject is first primed with the word "vampire" and subsequently asked to interpret the compound BOXER BAT, then they may be oriented towards giving an animal interpretation of the concept BAT (which could in turn influence their interpretation of BOXER). This suggests a minimal natural extension of the model where **A1** = +1 represents a situation where the subject was first primed with a word representing the dominant sense of concept A (e.g. for BAT this could be "vampire") and concept A was indeed subsequently interpreted in this dominant sense by a human subject when they are asked to give an interpretation for a conceptual combination. Conversely, **A1** = −1 represents the case where the subject was primed with the dominant sense of concept A but A was not subsequently interpreted in that sense. Similarly, **A2** = ±1 represents a situation where a subordinate sense of concept A was primed, and concept A was (+ 1) or was not (−1) interpreted in this sense. Note that a concept may have more than one subordinate sense. For example, the concept BOXER could be considered to have an extra subordinate sense indicative of clothing, namely "boxer shorts". This point can be incorporated into the above formalism based on a primary and subordinate senses by allowing that the concept A can be interpreted in a third (or more) senses. Thus, **A1** = −1
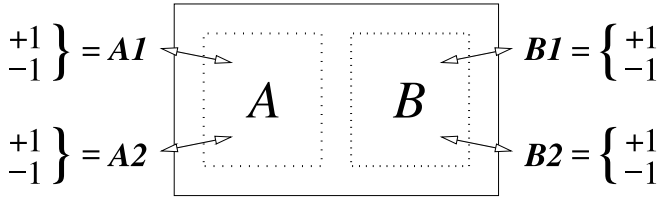
Fig. 2. A potentially compositional system *S*, consisting of two assumed components *A* and *B*. *S* can perhaps be understood in terms of a mutually exclusive choice of experiments performed upon those components, one represented by the random variables **A1**, **A2** (pertaining to an interaction between the experimenter and component *A*), and the other by **B1**, **B2** (pertaining to an interaction between the experimenter and component *B*). Each of these experiments can return a value of +1 or −1.

can occur because the concept primed by **A1** is interpreted in the subordinate sense (as above), or in a third sense not directly primed by the experimental arrangement.

Priming thus allows for the experimental control of the contextual cues influencing conceptual combinations. This is important because conceptual combinations always appear in a context (e.g., a discourse context), which affects how they will ultimately be interpreted. Fig. 2 gives a general representation of the reasoning used in the construction of the above probabilistic scenario. A 'black box' is depicted, with two proposed components, *A* and *B*, inside it. Two different experiments can be carried out upon each of the two presumed components, which will answer a set of 'questions' with binary outcomes, leading to four experimental scenarios. For example, one experimental scenario would be to ask whether subjects return an interpretation of the concept *A* that corresponds to the prime **A1** and similarly for *B* in relation to the prime **B2**. What analysis can be brought to bear upon such a situation?

As with many systems, the outcomes of our experiments will have a statistical distribution over all available outcomes. In what follows, we shall aim to develop a general mathematical apparatus that can be used to discover whether the presumed sub-components can be considered as isolated, influencing one another, or in some sense irreducible. We shall do this through a consideration of the joint probability distribution of the variables **A1**, **A2**, **B1**, and **B2**, $\Pr_{A1,A2,B1,B2}$, which can be used to model the behaviour of the experimental black box. While this analysis will be performed using conceptual combinations, we emphasise that this black box is potentially very general and that the analysis developed here can be applied to far more than the analysis of language.

We start by noting that if priming two concepts *A* and *B* using two senses of each concept, then we can construct 16 joint probabilities, corresponding to all the possible interpretations that a subject might return, across four possible priming conditions corresponding to two senses for each concept:

$$
\begin{aligned}
p_1 &\equiv \Pr(\mathbf{A1}=+1, \mathbf{B1}=+1) & p_2 &\equiv \Pr(\mathbf{A1}=+1, \mathbf{B1}=-1) \\
p_3 &\equiv \Pr(\mathbf{A1}=-1, \mathbf{B1}=+1) & p_4 &\equiv \Pr(\mathbf{A1}=-1, \mathbf{B1}=-1) \\
p_5 &\equiv \Pr(\mathbf{A1}=+1, \mathbf{B2}=+1) & p_6 &\equiv \Pr(\mathbf{A1}=+1, \mathbf{B2}=-1) \\
p_7 &\equiv \Pr(\mathbf{A1}=-1, \mathbf{B2}=+1) & p_8 &\equiv \Pr(\mathbf{A1}=-1, \mathbf{B2}=+1) \\
p_9 &\equiv \Pr(\mathbf{A2}=+1, \mathbf{B1}=+1) & p_{10} &\equiv \Pr(\mathbf{A2}=+1, \mathbf{B1}=-1) \\
p_{11} &\equiv \Pr(\mathbf{A2}=-1, \mathbf{B1}=+1) & p_{12} &\equiv \Pr(\mathbf{A2}=-1, \mathbf{B1}=-1) \\
p_{13} &\equiv \Pr(\mathbf{A2}=+1, \mathbf{B2}=+1) & p_{14} &\equiv \Pr(\mathbf{A2}=+1, \mathbf{B2}=-1) \\
p_{15} &\equiv \Pr(\mathbf{A2}=-1, \mathbf{B2}=+1) & p_{16} &\equiv \Pr(\mathbf{A2}=-1, \mathbf{B2}=-1).
\end{aligned}
\tag{1}
$$

These sixteen probabilities can be set out in an array as follows:

$$
A
\begin{array}{cc}
& \\
\mathbf{A1}\begin{array}{c}+1\\-1\end{array} \\
\mathbf{A2}\begin{array}{c}+1\\-1\end{array}
\end{array}
\left(
\begin{array}{cc|cc}
p_1 & p_2 & p_5 & p_6 \\
p_3 & p_4 & p_7 & p_8 \\
\hline
p_9 & p_{10} & p_{13} & p_{14} \\
p_{11} & p_{12} & p_{15} & p_{16}
\end{array}
\right).
\tag{2}
$$

This matrix lists the different priming conditions in a set of four blocks, which allows us to consider the structure of the probabilities describing the likely interpretation of a given conceptual combination. Observe how the matrix is complete, in that it covers all possible priming conditions across the two concepts for this scenario.

In what follows we will show how this matrix can be used to determine whether a conceptual combination is compositional, or not. We start by considering what might be required in order for a conceptual combination to be deemed compositional.

2.1. Compositional semantics

Were the semantics of the conceptual combination *AB* to be compositional, how would this be reflected in its probabilistic structure? The principle of semantic compositionality would suggest that the joint probability distribution could be recovered from the probability distributions constructed using each individual concept.

Thus, we shall take a given conceptual combination *AB* to be *compositional* if and only if a four way joint distribution $\Pr_{A1,A2,B1,B2}$ exists where $\Pr_{Ai,Bj}$, $i, j \in \{1, 2\}$ are marginal distributions. This opens the door to define non-compositionality via an unusual means, namely the inability to construct a joint probability distribution $\Pr_{A1,A2,B1,B2}$ in this way.

2.2. Non-compositional semantics

To analyse non-compositionality we draw upon results from the field of quantum theory surrounding entangled systems (see e.g. Laloë (2001) for a comprehensive review of the quantum formalism). This step is not as arbitrary as it might at first sight. An entangled system is one for which it is not always possible to construct a four way joint distribution from four empirically collected pairwise joint distributions. Of particular interest for the current argument, Fine's theorem (Fine, 1982) states the *necessary and sufficient conditions* for existence of a joint probability distribution, and hence the notion of compositionality introduced at the end of the previous section.

Fine Theorem 3 (*Fine, 1982*). *If* **A1**, **A2**, **B1**, **B2** *are bivalent random variables with joint distributions* $\Pr_{Ai,Bj}$, $i, j \in \{1, 2\}$, *then necessary and sufficient for a joint distribution* $\Pr_{A1,A2,B1,B2}$ *is that the following system of inequalities is satisfied:*

$$
\begin{aligned}
-1 \leq\ & \Pr(A1, B1) + \Pr(A1, B2) + \Pr(A2, B2) \\
& - \Pr(A2, B1) - \Pr(A1) - \Pr(B2) \leq 0
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
-1 \leq\ & \Pr(A2, B1) + \Pr(A2, B2) + \Pr(A1, B2) \\
& - \Pr(A1, B1) - \Pr(A2) - \Pr(B2) \leq 0
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
-1 \leq\ & \Pr(A1, B2) + \Pr(A1, B1) + \Pr(A2, B1) \\
& - \Pr(A2, B2) - \Pr(A1) - \Pr(B1) \leq 0
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
-1 \leq\ & \Pr(A2, B2) + \Pr(A2, B1) + \Pr(A1, B1) \\
& - \Pr(A1, B2) - \Pr(A2) - \Pr(B1) \leq 0,
\end{aligned}
\tag{6}
$$

where $\Pr(Ai, Bj)$ *is shorthand for* $\Pr(\mathbf{Ai} = +1, \mathbf{Bj} = +1)$, $\Pr(Ai)$ *for* $\Pr(\mathbf{Ai} = +1)$ *and* $\Pr(Bj)$ *represents* $\Pr(\mathbf{Bj} = +1)$, $i, j \in \{1, 2\}$.

Fine referred to this system of inequalities as the Bell/CH inequalities, but they hark back to the separability assumption made in Bell's original theorem (Bell, 1964) and we will refer to this class of inequalities as Bell-type inequalities. Fine's theorem permits us to analyse compositionality from a formal perspective that is open to experimentation. According to this approach, a conceptual combination *AB* is deemed "non-compositional" when the four pair wise joint probability distributions in (2) do not satisfy the Bell-type inequalities provided by Fine's theorem. This scenario implies that a joint distribution $Pr_{A1,A2,B1,B2}$ *cannot* be formed such that the four pairwise joint probability distributions $Pr_{Ai,Bj}$, $i, j \in \{1, 2\}$ are marginal distributions. Conversely, if all inequalities are satisfied then the four way joint probability distribution does exist, and the conceptual combination can be deemed "compositional" in the measurement context.

Physical systems adhere to a constraint variously termed "the causal communication constraint", "parameter independence", "simple locality", "signal locality", or "physical locality" (Maudlin, 1994). This is due to the constrictions of a theory independent from quantum theory; Special Relativity. All physical systems that are spatially separated should behave in such a way that "the probability of a particular measurement outcome on any one part of the system should be independent of which sort of measurement was performed on the other parts" (Cereceda, 2000). Such conditions are termed "marginal selectivity" (Dzhafarov & Kujala, 2012) in cognitive systems, and it is necessary for a system to satisfy them before it can be considered surprising that this system does not satisfy the Bell-type inequalities. Thus, we must have a reason to believe that a system *should* be considered separable before we can be shocked to find that it is not. For example, with respect to the conceptual combination BOXER BAT, marginal selectivity entails the interpretation of BAT does not change when the primes of BOXER are varied from "fighter" to "dog". This is a first indication that the two concepts could perhaps be modelled compositionally.

Marginal selectivity is expressed more formally as follows:

$$\begin{aligned}
Pr(\mathbf{Ai} = +1) &= Pr(\mathbf{Ai} = +1, \mathbf{B1} = +1) \\
&\quad + Pr(\mathbf{Ai} = +1, \mathbf{B1} = -1) \\
&= Pr(\mathbf{Ai} = +1, \mathbf{B2} = +1) \\
&\quad + Pr(\mathbf{Ai} = +1, \mathbf{B2} = -1), \quad i \in \{1, 2\} \quad (7)
\end{aligned}$$

$$\begin{aligned}
Pr(\mathbf{Bj} = +1) &= Pr(\mathbf{A1} = +1, \mathbf{Bj} = +1) \\
&\quad + Pr(\mathbf{A1} = -1, \mathbf{Bj} = +1) \\
&= Pr(\mathbf{A2} = +1, \mathbf{Bj} = +1) \\
&\quad + Pr(\mathbf{A2} = -1, \mathbf{Bj} = +1), \quad j \in \{1, 2\}. \quad (8)
\end{aligned}$$

Note how these four equations express that the interpretation of the concept represented by the marginal probability is stable with respect to how the other concept is primed (e.g., $Pr(\mathbf{Ai} = +1)$ is stable with respect to the different primings of concept *B* as represented by **B1** and **B2**).

Recently, Dzhafarov and Kujala (2012) have established a connection between Fine's theorem and the theory of selective influences in psychology, a result that suggests that Fine's theorem can be usefully applied to cognitive models. In a model with several factors and a set of random variables describing responses, *selective influence* concerns the problem of what factors influence what variables. The interpretation of conceptual combinations within a priming scenario can be treated with a model of selective influence, with primes corresponding to the factors affecting random variable corresponding to the interpretation of concepts. Dzhafarov and Kujala (2012) point out that selective influence implies marginal selectivity. Failure of marginal selectivity means there can be no model of selective influence, meaning there is no joint probability distribution $Pr_{A1,A2,B1,B2}$ where the pairwise distributions $Pr_{A1,B1}$, $Pr_{A1,B2}$, $Pr_{A2,B1}$, $Pr_{A2,B2}$ are marginal distributions.

The proof of Fine's theorem assumes locality, which is the physical equivalent of marginal selectivity for spatially separated physical systems (including the entangled systems of photons that occur in quantum physics). In cognitive science, however, concepts are not as well behaved as photons, so marginal selectivity may or may not hold. This is a crucial point. There has been some confusion about what characteristics a cognitive system should have before it can be modelled using Bell type inequalities. For example, Aerts, Gabora, and Sozzo (2013) present an experiment to establish whether the concepts ANIMAL and ACTS are "entangled" in the expression "The Animal Acts". Placed within the framework presented in this paper, the goal of the experiment was to determine whether the conceptual combination ANIMAL ACTS is compositional, or not. The authors employed the CHSH inequality and achieve a violation, and so claim that the combination is "entangled", i.e., non-compositional. However, a subsequent analysis of the experiment showed that marginal selectivity does not hold (Dzhafarov & Kujala, 2014), which means that the non-compositionality is in a certain sense trivial (the system should never have been modelled using compositional methods to start with).

For applications in cognitive science, marginal selectivity must be tested first, before any Bell-type inequality can be appropriately applied:

1. If marginal selectivity fails, then the conceptual combination is *immediately* judged as "non-compositional".
2. If marginal selectivity holds and any of the Bell-type inequalities are violated, then the conceptual combination is deemed "non-compositional".
3. If marginal selectivity holds and all of the Bell-type inequalities hold, then the conceptual combination is deemed "compositional".

Quantum physics has explored a number of equivalent formulations of the locality condition that is termed marginal selectivity in psychology (including the: Bell; Clauser–Horne (CH); and Clauser–Horne–Shimony–Holt (CHSH) forms (Laloë, 2001)). Of particular interest to the present argument, the CHSH inequality provides a formulation based on correlations between systems *A* and *B*, and so permits some insight to be gained into why conceptual combinations might be noncompositional even if they satisfy marginal selectivity in an experiment.

The CHSH inequality deals with expectation values rather than probabilities, and can be written as (Cereceda, 2000; Laloë, 2001):

$$-2 \leq E(\mathbf{A1}, \mathbf{B1}) + E(\mathbf{A1}, \mathbf{B2}) + E(\mathbf{A2}, \mathbf{B1}) - E(\mathbf{A2}, \mathbf{B2}) \leq 2 \quad (9)$$

where $E(\mathbf{Ai}, \mathbf{Bj})$ $i, j \in \{1, 2\}$ is a correlation function corresponding to the expectation value of a measurement of the experimental scenario depicted in Fig. 2.[3] Expectation values can be computed from the matrix of probabilities (2), e.g., $E(\mathbf{A1}, \mathbf{B1}) = p1 + p4 - (p2+p3)$. Recalling from (1) that $p_1 = Pr(\mathbf{A1} = +1, \mathbf{B1} = +1)$ and $p_4 = Pr(\mathbf{A1} = -1, \mathbf{B1} = -1)$, we recognise that p_1 corresponds to a situation where concepts *A* and *B* have *both* been interpreted in their dominant sense, when in both cases the dominant sense of each concept has been primed. Similarly, p_4 corresponds to both *A* and *B* being interpreted in a subordinate sense when the dominant sense of each concept has been primed. Thus, $p_1 + p_4 = 1$ occurs when the senses of the constituent concepts are perfectly correlated within the given priming condition.

For example, assuming that the fruit sense of the concept APPLE was primed, along with the food sense of CHIP. Perfect *correlation*

---

[3] We note that the expectation values used in these equations consider products. Thus Suppes, Acacio de Barros, and Oas (1998) write $E[\mathbf{A_i}\mathbf{B_j}]$. However, as the CHSH inequality is most commonly written in the $E(\mathbf{A_i}, \mathbf{B_j})$ form we have kept this notation throughout this paper to minimise confusion.

of senses in this priming condition means that two conditions hold: (1) when APPLE is interpreted as a fruit CHIP is always interpreted as food ($p1$) and, (2) when APPLE is *not* interpreted as fruit, CHIP is *not* interpreted as food ($p4$). The combination of these two conditions implies that $p_1 + p_4 = 1$ and $p_2 + p_3 = 0$. Conversely, $p_2 + p_3 = 1$ occurs when the senses are perfectly *anti-correlated*. If we assume the fruit sense of APPLE is primed and CHIP is primed in its electronic circuit sense then perfect anti-correlation of senses means two new conditions hold: (3) When APPLE is *not* interpreted as a fruit, CHIP is always interpreted as a circuit ($p3$), and (4) when APPLE is interpreted as fruit, CHIP is *not* interpreted in its circuit sense ($p2$).

The expectation value $E(\mathbf{Ai}, \mathbf{Bj})$ computes the degree to which the senses of the constituent concepts are (anti-)correlating. The arrangement of probabilities in figure (2) is not significant. There are thus four possible ways to arrange the quadrants, with each arrangement leading to a variant of the CHSH inequality:

$$-2 \leq E(\mathbf{A1}, \mathbf{B1}) - E(\mathbf{A1}, \mathbf{B2}) + E(\mathbf{A2}, \mathbf{B1}) + E(\mathbf{A2}, \mathbf{B2}) \leq 2 \quad (10)$$

$$-2 \leq E(\mathbf{A1}, \mathbf{B1}) + E(\mathbf{A1}, \mathbf{B2}) - E(\mathbf{A2}, \mathbf{B1}) + E(\mathbf{A2}, \mathbf{B2}) \leq 2 \quad (11)$$

$$-2 \leq -E(\mathbf{A1}, \mathbf{B1}) + E(\mathbf{A1}, \mathbf{B2}) + E(\mathbf{A2}, \mathbf{B1}) + E(\mathbf{A2}, \mathbf{B2}) \leq 2. \quad (12)$$

Therefore, there are four CHSH inequalities in total (9)–(12), each differing in where the minus sign is placed. The heart of each inequality is a computation involving correlations which will be referred to as the CHSH value. When the CHSH value of any of the inequalities lies outside of the range $[-2, 2]$, meaning its absolute value is greater than 2, then there is no joint probability distribution $\Pr_{\mathbf{A1,A2,B1,B2}}$ such that the four empirically collected pairwise distributions $\Pr_{\mathbf{A1,B1}}$, $\Pr_{\mathbf{A1,B2}}$, $\Pr_{\mathbf{A2,B1}}$, $\Pr_{\mathbf{A2,B2}}$ are marginal distributions. In such a case, the associated conceptual combination is deemed "non-compositional".

Conversely, when the CHSH value lies within $[-2, 2]$ for all four inequalities, there *is* a joint probability distribution $\Pr_{\mathbf{A1,A2,B1,B2}}$ where the four empirically collected pairwise distributions: $\Pr_{\mathbf{A1,B1}}$, $\Pr_{\mathbf{A1,B2}}$, $\Pr_{\mathbf{A2,B1}}$, $\Pr_{\mathbf{A2,B2}}$ are marginal distributions. In this case, the conceptual combination is deemed "compositional".

When marginal selectivity holds, the CHSH inequalities and Bell/CH inequalities are algebraically equivalent. However, as the CHSH inequalities are based on correlations, they offer a means to explain non-compositionality in terms of correlations between senses. As was the case with Fine/CH inequalities, marginal selectivity must first be tested before the four CHSH inequalities can be applied. Thus, in QT we have found a probabilistic formalism that allows for the assumption of compositionality to be tested. If a system of probabilistic relationships can be constructed for a cognitive scenario (or any other scenario that matches the structure depicted in Fig. 2) then we can test whether it should be deemed compositional. Non-compositionality would then be determined by the inability to construct a joint probability distributions across the four variables modelling how the primary and a subordinate sense of the concepts A and B are interpreted. We now illustrate how these probabilistic methods for analysing compositionality can be deployed in an experimental setting.

3. Empirical illustration

3.1. Subjects

Sixty-five subjects were recruited from the undergraduate psychology pool at Griffith University and received credit for their participation. Only native English speakers were selected in order to remove the possibility that the interpretation of conceptual combinations would be confounded by language issues.

3.2. Design and materials

We utilised four different priming regimes in order to generate the four different experimental scenarios suggested by Fig. 2. In these experiments, subjects were first primed and then presented with a non-lexicalised conceptual combination which they were asked to interpret, also designating the senses that were used in that interpretation (see Fig. 3). A probabilistic analysis was then performed upon the data so obtained. Subjects were presented with twenty-four 'true' conceptual combinations (see below for an explanation), and so participated in twenty-four test trials. Table 2 lists the set of conceptual combinations used, as well as the corresponding primes.

Primes were selected from the USF free association norms (Nelson et al., 2004) and the University of Alberta norms of homographs (Twilley, Dixon, Taylor, & Clark, 1994). The majority of primes were selected from the USF norms. The procedure for selecting primes from these norms was to view a potential prime as a cue which produces the required concept as an associate with a high probability. As an example, "money" was chosen from the USF norms to prime the financial sense of BANK as "bank" is produced as a free associate of the cue "money" with high probability. Similarly, "river" was chosen to prime the natural sense of BANK. Occasionally when a particular sense was not present in the USF norms, we drew upon the University of Alberta norms. Importantly, the USF norms were used to avoid cues such as "account" which are associated with both BANK and LOG, thereby minimising the possibility of priming more than one concept at a time. Specific conceptual combinations were chosen with the expectation that the ambiguity of constituents would allow a number of alternative interpretations, where each interpretation arose from a different attribution of meaning to the underlying sense of the ambiguous concepts (Costello & Keane, 1997).

A single factor design was used, which analysed responses to non-lexicalised conceptual combinations under priming conditions that varied between subjects. A subject was assigned to one of four priming conditions for each presented conceptual combination. For example, the four priming conditions for BANK LOG are (1) "money" and "journal" ($\mathbf{A1} - \mathbf{B1}$), (2) "money" and "tree" ($\mathbf{A1} - \mathbf{B2}$), (3) "river" and "journal" ($\mathbf{A2} - \mathbf{B1}$), or (4) "river" and "tree" ($\mathbf{A2} - \mathbf{B2}$). This assignment of primes was based upon a between groups Latin square design, such that for the 24 combinations, each participant completed each priming condition 6 times.

3.3. Procedure

Fig. 3 shows a schematic illustration of the procedure followed during a test trial. Participants completed 3 practice trials, 24 test trials and 24 filler trials. All trials were composed of six phases, consisting of three initial time-pressured tasks followed by three non-timed tasks. The time limitation of the first three phases was introduced with the expectation that this would maximise the effectiveness of the priming. The experiment took around 20–30 min to complete, and participants pushed the ENTER key to begin each trial.

Phases 1–2:

Two consecutive double lexical decision tasks were carried out, where participants were asked to decide as quickly as possible whether two letter strings, a prime and the concept to be presented as a part of the compound given in Phase 3, were legitimate words, or if one of the strings was a non-word. Each lexical decision consisted of the two letter strings presented in the centre of screen, one below the other in order to discourage participants from interpreting the two words as a phrase. Participants responded to the decision tasks by pushing a button on the keyboard, labelled 'word' or a button labelled 'non-word' (left arrow and right arrow
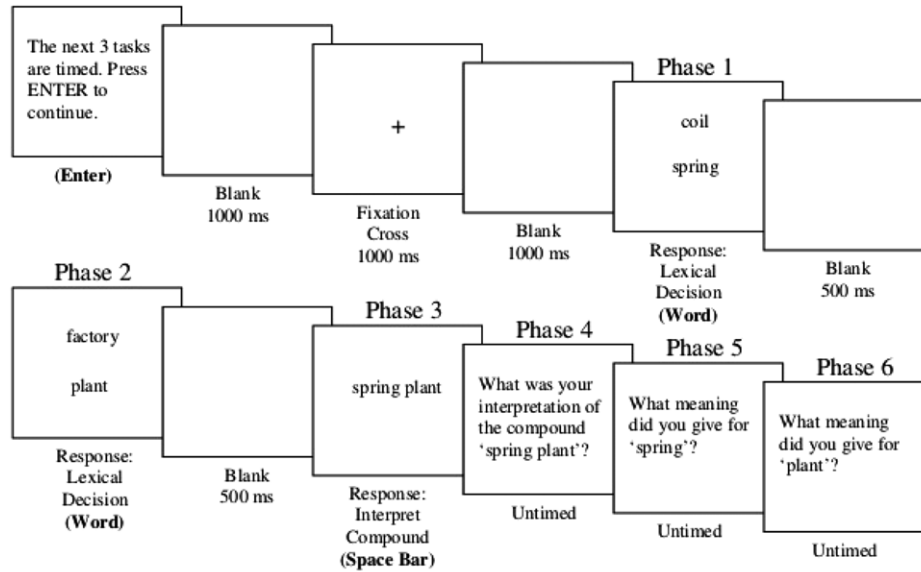
Fig. 3. Example experimental structure for a trial. Non-word trials followed a similar structure, with primes in Phase 1 and/or Phase 2 replaced with non-words. The sequence of squares moving from left to right shows the experimental flow, with each square a representation of the screen shown to a participant. Note: the figure does not show the exact text given to participants, and stimuli are not to scale.

keys respectively). For instance, if given the strings "coil" and "spring", then participants were expected to decide that both strings were words and so push the 'word' key, whereas if given "grod" and "church" then participants were expected to decide that they had been shown a non-word and to push the 'non-word' key. For all of the test trials participants received two phases of word–word strings. The response ratio for the two priming phases was: 50% word → word (test trial), 25% non-word → non-word (filler trial), 12.5% word → non-word (filler trial), 12.5% non-word → word (filler trial). In phases where a non-word was present, it appeared equally often in the top or the bottom portion of the screen.

The double-lexical decision task was used to associate the priming word and test word together without participants interpreting them as a compound (Gagne, 2001). This procedure isolates the experimental priming to each concept in the combination. For example, the lexical decision task applied to "coil" and "spring" was designed to prime the coil sense of the concept SPRING in the conceptual combination SPRING PLANT. The order of the two double lexical decision tasks was counter-balanced, so that half were presented in the same order as the compound words (e.g., "coil" and "spring" were first presented, then "factory" and "plant") and half were presented in the reverse order (e.g., first "factory" and "plant" were presented for lexical decision, followed by "coil" and "spring").

Phase 3:

A conceptual combination was presented in the centre of the screen (e.g., "spring plant"). Participants were asked to push the space bar as soon as they thought of an interpretation for the compound. Filler compounds were included for the filler (i.e., non-word) trials so as not to disrupt the participant's rhythm in making two lexical decisions followed by an interpretation.

Phase 4:

Participants were asked to type in a description of their interpretation.

Phases 5–6:

Two disambiguation tasks were carried out, where participants chose what sense they gave to each word from a list (e.g., plant = A. 'a living thing'; B. 'a factory'; C. 'other').

3.4. Results

Experimental subcomponents utilising non-words were discarded during the analysis. In total, 91.5% of the interpretations provided by the subjects fell within one of the four primed senses of the studied conceptual combinations.

As stated previously, in order to apply Bell-type inequalities for compositional analysis, marginal selectivity must first be tested. Table 1 depicts an analysis of marginal selectivity where the values in the columns depict the difference of marginal probabilities across the conditions of the associated variable, as well as the confidence intervals. For example, diff(**A1**) is the difference between the one-marginal $\Pr(\mathbf{A1} = +1, \mathbf{B1} = +1) + \Pr(\mathbf{A1} = +1, \mathbf{B1} = -1)$ and the second one-marginal $\Pr(\mathbf{A1} = +1, \mathbf{B2} = +1) + \Pr(\mathbf{A1} = +1, \mathbf{B2} = -1)$. Marginal selectivity holds when these differences are zero across all four variables. The breadth of the confidence intervals and the fact that many are not centred around zero does not allow us to conclude with confidence that any of these conceptual combinations satisfy marginal selectivity, although BATTERY CHARGE, BILL SCALE and TOAST GAG could possibly be adhering to this condition as the differences in marginal probabilities are low, and their confidence intervals are overlap 0 for all values of diff. However, the sample size is small (see Table 2) and so we cannot be confident that this condition is satisfied. However, for the purposes of illustration, we will assume in the analysis to follow that these three combinations do satisfy marginal selectivity.

The result of the compositional analysis is depicted in Table 2. We have tentatively flagged combinations as "non-compositional" if they appear likely to fail marginal selectivity given the current dataset. Of the combinations that are assumed to satisfy marginal selectivity, BILL SCALE ($|CHSH| = 1.63$) and TOAST GAG ($|CHSH| = 1.63$) are deemed "compositional" as their CHSH values are less than or equal to 2. BATTERY CHARGE shows a slight violation of the CHSH inequalities ($|CHSH| = 2.01$), so could be deemed "non-compositional", but due to the lack of confidence in whether marginal selectivity is holding, we cannot make any firm conclusions. More experiments will be required.

Table 1
Analysis of marginal selectivity. Values represent differences of marginal probabilities with associated 95% confidence interval (obtained using a 2-sample test for equality of proportions with continuity correction). BATTERY CHARGE, BILL SCALE and TOAST GAG are starred (*) as they are assumed to satisfy marginal selectivity.

Combination	diff (**A1**)	diff (**A2**)	diff (**B1**)	diff (**B2**)
Boxer bat	0.175 [−0.19,0.30]	0.140 [−0.32,0.50]	0.338 [−0.04,0.71]	0.158 [0.25,0.57]
Bank log	0.055 [−0.22,0.32]	0.092 [−0.27,0.42]	0.338 [0,0.67]	0.257 [−0.08,0.60]
Apple chip	0.250 [−0.02,0.52]	0.114 [−0.28,−0.51]	0.294 [−0.01,0.61]	0.217 [−0.17,0.61]
Stock tick	0.163 [−0.24,0.57]	0.085 [−0.30,0.48]	0.488 [0.11,0.85]	0.386 [0.02,0.74]
Seal pack	0.083 [−0.30,0.47]	0.213 [−0.10,0.37]	0.162 [−0.21,0.53]	0.221 [−0.18,0.63]
Spring plant	0.294 [0.01,0.57]	0.133 [−0.13,0.41]	0 [0,0]	0.173 [−0.12,0.46]
Poker spade	0.136 [−0.21,0.48]	0.035 [−0.28,0.35]	0 [−0.29,0.29]	0.113 [−0.26,0.48]
Slug duck	0.096 [−0.31,0.52]	0.153 [−0.21,0.52]	0.133 [−0.22,0.48]	0.026 [−0.34,0.38]
Club bar	0.133 [−0.10,0.37]	0 [−0.23,0.23]	0.125 [−0.10,0.35]	0.138 [−0.16,0.44]
Web bug	0.210 [−0.18,0.60]	0.067 [−0.30,0.44]	0.296 [−0.10,0.69]	0.153 [−0.21,0.52]
Table file	0.058 [−0.21,0.32]	0.235 [−0.27,0.32]	0.114 [−0.25,0.40]	0.113 [−0.26,0.48]
Match bowl	0.137 [−0.26,0.54]	0.250 [−0.12,0.62]	0.075 [−0.23,0.38]	0.022 [−0.33,0.37]
Net cap	0.035 [−0.29,0.36]	0.092 [−0.31,0.49]	0.059 [−0.30,0.51]	0.175 [−0.34,0.43]
Stag yarn	0.375 [0.02,0.73]	0.219 [−0.12,0.56]	0.104 [−0.26,0.43]	0.045 [−0.30,0.39]
Mole pen	0.125 [−0.16,0.41]	0.021 [−0.33,0.37]	0.063 [−0.34,0.46]	0.3 [−0.08,0.68]
Battery charge*	0.067 [−0.21,0.35]	0.048 [−0.28,0.37]	0.117 [−0.22,0.45]	0.120 [−0.23,0.43]
Count watch	0.195 [−0.14,0.53]	0.063 [−0.25,0.38]	0.011 [−0.26,0.28]	0.063 [−0.29,0.41]
Bill scale*	0.081 [−0.30,0.46]	0.113 [−0.26,0.48]	0.054 [−0.26,0.37]	0.051 [−0.24,0.34]
Rock strike	0.188 [−0.07,0.44]	0.117 [−0.22,0.45]	0.313 [0.02,0.60]	0.013 [−0.28,0.30]
Port vessel	0.106 [−0.29,0.50]	0.085 [−0.31,0.48]	0.113 [−0.26,0.48]	0.118 [−0.19,0.43]
Crane hatch	0.141 [−0.15,0.44]	0.296 [−0.10,0.69]	0.149 [−0.18,0.48]	0.233 [−0.16,0.63]
Toast gag*	0.0625 [−0.12,0.24]	0.008 [−0.32,0.34]	0.018 [−0.35,0.38]	0.015 [−0.34,0.38]
Star suit	0.308 [−0.02,0.64]	0.163 [−0.25,0.48]	0.054 [−0.26,0.37]	0.058 [−0.20,−0.32]
Fan post	0.35 [−0.04,0.74]	0.125 [−0.28,0.53]	0.025 [−0.34,0.39]	0.188 [−0.20,0.57]

Table 2
Results of the compositionality analysis: 'Y/N' indicates whether the conceptual combination is compositional, or not *N* the number of subjects. Conceptual combinations adhering to marginal selectivity have CHSH value shown in brackets.

Combination	Concept A		Concept B		Results	
	Prime 1 (**A1**)	Prime 2 (**A2**)	Prime 3 (**B1**)	Prime 4 (**B2**)	*Compositional*	*N*
Boxer bat	*dog*	*fighter*	*ball*	*vampire*	N	64
Bank log	*money*	*river*	*journal*	*tree*	N	64
Apple chip	*banana*	*computer*	*potato*	*circuit*	N	65
Stock tick	*shares*	*cow*	*mark*	*flea*	N	64
Seal pack	*walrus*	*envelop*	*leader*	*suitcase*	N	64
Spring plant	*summer*	*coil*	*leaf*	*factory*	N	64
Poker spade	*card*	*fire*	*ace*	*shovel*	N	65
Slug duck	*snail*	*punch*	*quack*	*dodge*	N	63
Club bar	*member*	*golf*	*pub*	*handle*	N	64
Web bug	*spider*	*internet*	*beetle*	*computer*	N	63
Table file	*chair*	*chart*	*nail*	*folder*	N	63
Match bowl	*flame*	*contest*	*disk*	*throw*	N	64
Net cap	*gain*	*volleyball*	*limit*	*hat*	N	65
Stag yarn	*party*	*deer*	*story*	*wool*	N	61
Mole pen	*dig*	*face*	*pig*	*ink*	N	63
Battery charge	*car*	*assault*	*volt*	*prosecute*	? [2.01]	63
Count watch	*number*	*dracula*	*time*	*look*	N	65
Bill scale	*phone*	*pelican*	*weight*	*fish*	Y? [1.63]	64
Rock strike	*stone*	*music*	*hit*	*union*	N	63
Port vessel	*harbour*	*wine*	*ship*	*bottle*	N	65
Crane hatch	*lift*	*bird*	*door*	*egg*	N	63
Toast gag	*jam*	*speech*	*choke*	*joke*	Y? [1.23]	63
Star suit	*moon*	*movie*	*vest*	*law*	N	62
Fan post	*football*	*cool*	*mail*	*light*	N	63

3.5. Discussion

In this discussion we provide further details with the aim of shedding light on *how* the joint probability distribution is structured and what this might mean when a violation of compositionality occurs. We shall utilise two examples: TOAST GAG and APPLE CHIP.

TOAST GAG

Matrix (13) depicts the empirical results for TOAST GAG. Here, we see no particular ordering or patterns. In particular, when we compare the form of the equation required for a violation of Eqs. (9)–(12) and the actual values in matrix (13) we can see that the probability mass does not centre sufficiently around the diagonals in such a way that it can produce the correlations

between the senses necessary to violate the CHSH inequality as $|CHSH| = 1.23 \leq 2$. The conceptual combination TOAST GAG is therefore deemed to be "compositional" as a joint probability distribution $\Pr_{A1,A2,B1,B2}$ can be constructed, which models how it is interpreted within the given priming conditions.

$$
\begin{array}{c}
\\
\text{TOAST}
\end{array}
\begin{array}{cc}
 & \\
\mathbf{A1}(jam) & \begin{array}{c}+1\\-1\end{array}\\
\\
\mathbf{A2}(speech) & \begin{array}{c}+1\\-1\end{array}
\end{array}
\begin{array}{c}
\text{GAG}\\
\begin{array}{cccc}
\mathbf{B2}(choke) & & \mathbf{B1}(joke) & \\
+1 & -1 & +1 & -1
\end{array}\\
\left(\begin{array}{cc|cc}
0.50 & 0.4375 & 0.625 & 0.375\\
0.0625 & 0 & 0 & 0\\
\hline
0.29 & 0 & 0.07 & 0.21\\
0.29 & 0.42 & 0.57 & 0.14
\end{array}\right)
\end{array}. \quad (13)
$$

APPLE CHIP

In contrast, APPLE CHIP leads to a joint distribution that has a more interesting structure:

$$
\begin{array}{c}
\\
\text{APPLE}
\end{array}
\begin{array}{cc}
 & \\
\mathbf{A1}(banana) & \begin{array}{c}+1\\-1\end{array}\\
\\
\mathbf{A2}(computer) & \begin{array}{c}+1\\-1\end{array}
\end{array}
\begin{array}{c}
\text{CHIP}\\
\begin{array}{cccc}
\mathbf{B1}(potato) & & \mathbf{B2}(circuit) & \\
+1 & -1 & +1 & -1
\end{array}\\
\left(\begin{array}{cc|cc}
0.94 & 0.06 & 0 & 0.75\\
0 & 0 & 0.25 & 0\\
\hline
0 & 0.35 & 0.47 & 0\\
0.65 & 0 & 0 & 0.53
\end{array}\right)
\end{array}. \quad (14)
$$

It is clear from the values in Table 1 that APPLE CHIP fails marginal selectivity. Therefore the joint probability distribution $\Pr_{A1,A2,B1,B2}$ cannot be constructed from the four empirically collected pairwise joint probability distributions such that these four pairwise distributions depicted in matrix (14) can be recovered by marginalising this four way joint distribution. This conceptual combination is therefore deemed "non-compositional".

However, we claim that the status of this conceptual combination as non-compositional is likely to be more interesting than a simple failure of marginal selectivity. APPLE CHIP shows a strong pattern of correlation between the senses across the four priming conditions because the probabilities are concentrated on the diagonals or reverse diagonals. Thus, whenever a subject interprets APPLE as a fruit they tend to interpret CHIP in its FOOD sense. Conversely, if APPLE is interpreted as a 'computer' then a CHIP is interpreted as an 'electronic device'. This structure was quite common in the conceptual combinations that were studied. A second key factor is that a non-zero value has been returned by the ensemble of subjects for one off-diagonal case $p_2 = \Pr(A1, B1) = 0.06$ (see Matrix (2)). Even though the food sense of CHIP has been primed, atypical interpretations of the compound are produced, for example, "apple's growth is controlled by an internal chip". Costello and Keane (2000) identify three categories of non-compositionality in novel conceptual combinations, and atypical instances are at the basis of one of these categories. Some other non-compositional conceptual combinations showed similar atypical interpretations. For example, BANK LOG also exhibits a strong correlation between the senses: When BANK is interpreted as a financial institution, LOG tends to be interpreted as a "record". Conversely, when BANK is interpreted in it's "river" sense, LOG is interpreted as a "piece of wood". However, there were atypical cases where the senses cross over which produces an off-diagonal probability e.g., "a record of a bank of a river".

We hypothesise that one way for a conceptual combination to satisfy marginal selectivity and yet be deemed non-compositional involves a particular structure, which is demonstrated by the example depicted in matrix (15). Here we see the probability mass is largely concentrated along diagonals, because typical

interpretations can often occur when senses are (anti-)correlating. For example, when APPLE is interpreted as fruit, CHIP is interpreted as food. The small off-diagonal probabilities reflect the atypical interpretations of Costello and Keane (2000), which may signify non-compositionality. Assuming each of quadrants in Matrix (15) is based on 100 data points, then diff(**A1**) = 0.02 and the 95% confidence interval is $[-0.07, 0.11]$. Similarly, diff(**A2**) = 0.01$[-0.06, 0.8]$, diff(**B1**) = 0.01$[-0.10, 0.12]$ and diff(**B2**) = 0.01$[-0.07, 0.09]$. These figures demonstrate what the statistics should look like when marginal selectivity is holding, i.e., the differences in marginal probabilities are very small and the confidence intervals are tightly centred around zero. In addition, the probabilities in matrix (15) yield an absolute CHSH value of 2.06. The atypical interpretations, highlighted by the bolded probabilities, are what force the CHSH value to exceed the threshold of two, and thus into non-compositionality.

$$
\begin{array}{c}
\\
\text{A}
\end{array}
\begin{array}{cc}
 & \\
\mathbf{A1}\ (\text{prime a1}) & \begin{array}{c}+1\\-1\end{array}\\
\\
\mathbf{A2}\ (\text{prime a2}) & \begin{array}{c}+1\\-1\end{array}
\end{array}
\begin{array}{c}
\text{B}\\
\begin{array}{cccc}
\mathbf{B1}\ (\text{prime b1}) & & \mathbf{B2}\ (\text{prime b2}) & \\
+1 & -1 & +1 & -1
\end{array}\\
\left(\begin{array}{cc|cc}
0.85 & \mathbf{0.05} & 0 & 0.92\\
0 & 0.10 & 0.08 & 0\\
\hline
0 & 0.06 & 0.07 & 0\\
0.86 & \mathbf{0.08} & 0 & 0.93
\end{array}\right)
\end{array}. \quad (15)
$$

4. Broader reflections on compositionality and non-compositionality

A major contribution of this paper is its demonstration of a methodology by which we might start to explore the debate about compositionality empirically. One question that has not yet been satisfactorily answered to date concerns the underlying cause of languages' apparently non-compositional behaviour: does language *just* violate Marginal Selectivity, or is it possible that some conceptual combinations might satisfy Marginal Selectivity and yet violate a Bell type inequality? This paper has pointed to a systematic way in which this question might be answered. While the experiments discussed in Section 3 are not conclusive about Marginal Selectivity, they do point towards some compounds that could perhaps be shown to satisfy this property with a large enough sample size. If such an experimental scenario was then shown to violate a Bell-type inequality then we would have learned much about the cognitive processes underlying language comprehension. More experimentation is required.

The importance of definitively answering such a question can be illustrated with reference to Costello and Keane (2000), who classify non-compositional conceptual combinations into three categories depending upon how their apparent non-compositionality arises. Firstly, some combinations are deemed non-compositional because of *emergent* properties, which generally arise from a meaning which is based on a subset of atypical instances. The aforementioned PET FISH example is placed in this category. A second set of conceptual combinations are classified non-compositional due to the manner in which the senses of the combining words are extended beyond their standard usage, to refer to instances outside the categories usually named by those words. Finally, some conceptual combinations are classified as non-compositional because they make use of cognitive processes such as metaphor, analogy or metonymy in their interpretation. Costello and Keane (2000) use the conceptual combination SHOVEL BIRD to illustrate all three categories:

1. A "shovel bird" could be a bird with a flat beak for digging up food.

2. A "shovel bird" could be a bird that comes to eat worms when you dig in the garden.
3. A "shovel bird" could be a plane that scoops up water from lakes to dump on fires.
4. A "shovel bird" could be a company logo stamped on the handle of a shovel.
5. A "shovel bird" could be someone allowed out of jail (free as a bird) as long as he works on a road crew.

They argue that (1) and (2) are examples of the first category because a bird with a flat beak is atypical, whereas (3) illustrates the second category because it extends the sense of both SHOVEL and BIRD beyond their normal usage. Finally (4) and (5) are put forward as examples of third category due to their metaphoric nature. Costello and Keane (2000) detail how their constraint-based theory of conceptual combination specifically relates to each of these categories. The framework presented in this paper, however, models the non-compositionality of SHOVEL BIRD irrespective of the category of non-compositionality involved. For example, SHOVEL has the sense of being a tool, or being shaped like a shovel. The concept BIRD has three senses in the preceding example: relating to an animal, a plane, and a prisoner. Thus, the concept BIRD could be modelled as consisting of both a dominant ANIMAL (**A1**) and a subordinate PLANE (**A2**) sense. In this more general scenario, the broad class of Bell-type inequalities (including the CHSH, CH and Fine variants discussed in this paper) could be applied to test for the non-compositionality of each possible interpretation resulting from a combination of SHOVEL with BIRD.

In addition, there is no requirement in the presented analytical framework that the concepts be explicit homographs, it can also be applied in a number of other scenarios. For example, a weaker form of ambiguity, *polysemy* can also be explored with this framework. A WordNet analysis of the noun–noun combinations used in the compositional models explored by Mitchell and Lapata (2010) reveals that the vast majority have more than one synset and hence more than one shade of meaning, and thus that they are polysemous (as was the case for the concept SHOVEL above). Similarly, ambiguity could also derive from the *relations* that link two conceptual combinations and thus our framework could allow for a clarification as to which word is acting as a head and which a modifier in conceptual combinations. For example, the CARIN model assumes that relations apply to the modifier, so in ADOLESCENT DOCTOR, an ambiguous concept that is discussed by Gagne (2001), an ambiguity arises between the competing relations in "doctor FOR adolescents" and "doctor IS adolescent". Both of these possibilities for the concept ADOLESCENT could be accessed through priming, and then probabilistically represented with their corresponding variables **A1** or **A2** (Gagne (2001) provides an experimental procedure for priming relations). DOCTOR is also ambiguous because it is polysemous, e.g., a medical doctor, or someone holding a Ph.D. Both of these possibilities could be modelled by the variables **B1** and **B2**. The analytical framework presented here could be applied to both of these scenarios, and thus the study of (non-)compositionality in conceptual combinations which have already been considered in the psychological literature. The joint distribution criterion (JDC) proposed by Dzhafarov and Kujala (2012) provides an extension of the framework presented here that could be put to sensible use in testing these more general scenarios. This condition is decided by solving a linear programming problem of the form $\mathbf{MQ} = \mathbf{P}, \mathbf{Q} \geq 0$. In the context of this article, the vector $\mathbf{P}$ would comprise the sixteen probabilities depicted in (2) and $\mathbf{Q}$ would represent the global joint distribution $\Pr_{A1,A2,B1,B2}$. Dzhafarov and Kujala (2012) prove that if marginal selectivity does not hold, then there is no solution for $\mathbf{Q}$. If marginal selectivity holds and no distribution $\mathbf{Q}$ can be found, then the associated conceptual combination can be deemed non-compositional. The linear programming approach is more general than the CHSH and Bell/CH inequalities in that it applies to any number of random variables with any number of possible values, and is a potentially rich area for consideration in future work.

In summary, this paper has proposed a framework for empirically testing the dividing line between compositionality and non-compositionality, not an adjudication upon the ongoing debate about compositionality in conceptual representation. One test that we provide is based on the violation of the Bell class of inequalities, but this can only be considered surprising in a scenario where the other test (Marginal Selectivity) is also satisfied. This is because, similarly to the locality condition in Bell-type inequalities, Marginal Selectivity can be understood as the underlying basis upon which a system could initially be assumed to be separable in the first place. A major contribution of this paper is a method capable of determining which of the two underlies non-compositional behaviour. However, the determination of compositionality that this analysis provides must take into account the priming conditions of the test, which empirically simulate the context (e.g., the discourse context) of the interpretation. As discussed by Kitto (2014), there is no result without a supplied context (in this case the priming), and it is important that we capture this information.

It appears that historically George Boole considered the problem of the constraints involved when trying to construct a global distribution of three variables from pairwise joint distributions (Pitowsky, 1994), however, Vorob'ev discovered results constraining this approach, providing a set of results more general to that of Fine's theorem. Vorob'ev was ignored (Khrennikov, 2010), apparently because his results pointed to the potential limits of standard probability theory, which was gaining in popularity as it was developed by Kolmogorov. Thus, it was quantum physics that became famous for demonstrating the impossibility of modelling entangled systems in a single probability space. In our opinion, this is but a quirk of the past, and Dzhafarov and Kujala (2012) have independently shown how such results can appear in cognitive psychology. The history just sketched, together with the fact that the Bell-type inequalities are based solely on conventional probability theory, opens the possibility to non-controversially apply them outside of quantum physics (Aerts, Aerts, Broeckaert, & Gabora, 2000; Aerts et al., 2014; Bruza, Kitto, Nelson, & McEvoy, 2009; Khrennikov, 2010).

5. Conclusions

This article departed from the assumption that conceptual combinations may not exclusively exhibit compositional semantics. The very idea of a non-compositional semantics has been resisted in the literature spanning cognitive science, philosophy and linguistics, probably because the "principle of compositionality" has had such a significant track record of success over a long period. It is, however, precisely the assumption that semantics must *necessarily be* of a compositional form that has been regularly questioned in a wide range of literature. Despite this state of confusion, few analytical approaches have been proposed that are capable of demarcating the difference between the two forms of behaviour. We have shown that it is possible to analyse the manner in which the semantics of a given conceptual combination might be considered as compositional, or non-compositional. Indeed, it is perhaps timely to remind the reader that we do not argue against compositional semantics *per se*. Rather, we have tried in this article to shed light on the line at which it breaks down: We believe that both compositional and non-compositional models will be necessary in order to provide a full account of the semantics of language.

We modelled the semantics of concepts in terms of the different senses in which a concept may be understood, where a given

sense corresponds to the interpretation attributed to a particular ambiguous concept. These senses have a reliable intersubjective cognitive underpinning, as they were grounded in terms of human word association norm data, which was used to predict the probability that a subject would attribute a particular sense to an ambiguous concept. Utilising formal frameworks developed for analysing composite systems in quantum theory, we presented two methods that allow the semantics of conceptual combinations to be classified as "compositional" or "non-compositional". This classification differs from previous research in two ways. Firstly, compositionality is not graded, e.g., "weak" versus "strong" compositionality. Secondly, the declaration of compositionality, or non-compositionality, is not an absolute classification, but context sensitive. An empirical study of twenty-four novel conceptual combinations illustrates how the classifications can be applied. Important corollaries are:

- Conceptual combinations violating marginal selectivity cannot be modelled in a single probability space across the four variables modelling the respective interpretations of the constituent concepts. Such conceptual combinations are immediately "non-compositional".

- When marginal selectivity does hold, and the Bell-type inequalities are not violated, then the semantics of the conceptual combination cannot be modelled in a four way joint probability distribution, the variables of which correspond to how the constituent concepts are being interpreted in their respective dominant and subordinate senses. Such conceptual combinations are "compositional".

- When marginal selectivity does hold, and any of the Bell class of inequalities *are* violated, then the semantics of the conceptual combination cannot be modelled in a four way joint probability distribution. Such conceptual combinations are "non-compositional".

This result could have a marked impact in modelling cognitive phenomena more generally, as these phenomena are frequently assumed to be compositional, and no thought is given as to whether the phenomenon can be modelled within a given probability space that the modeller constructs in terms of random variables. It is simply assumed that it can. Experiments from quantum physics show that for entangled systems no such model exists.

Finally, this article adds further weight to the claim that quantum theory is a fruitful source of new theoretical insights and tools for modelling conceptual semantics, as it has already done in other areas of cognition (Aerts, 2009; Aerts et al., 2013; Blutner, Pothos, & Bruza, 2013; Bruza, Busemeyer, & Gabora, 2009; Busemeyer & Bruza, 2012; Busemeyer, Pothos, Franco, & Trueblood, 2011; Khrennikov, 2010; Nelson et al., 2013; Pothos, Busemeyer, & Trueblood, 2013).

Acknowledgments

	Consistent	Inconsistent
Overall	6.88	4.72
Same Order	3.20	2.32
Reverse Order	3.67	2.40

Fig. A.4. Mean number of interpretations (consistent or inconsistent) with the primes by prime order (overall, same prime order, reverse prime order).

Appendix. Possible confounding factors

A number of factors beyond marginal selectivity must be considered when constructing experiments such as the ones introduced here. Are the primes working as intended? How familiar are the compound conceptual combinations? Could response time be taken as an indicator that the experimental design is inappropriate? Factors such as these could influence the frequency of resulting interpretations at a statistical level, and so must be carefully controlled. In this appendix we show that a number of possible confounding factors have been taken into account in this work, demonstrating that the priming used in these experiments can be considered effective, despite the complexity of the protocol.

A.1. Frequency of interpretations

The frequency of interpretations was analysed using Wilcoxon signed-rank tests. The results are summarised in Fig. A.4.

As expected, overall participants gave significantly more interpretations that were consistent with the primes (mean = 6.88), than inconsistent with the primes (mean = 4.72), $z = 4.06$, $p < .0001$. This provides evidence that the primes were affecting the interpretations given in the correct direction. To analyse whether the order in which the primes were shown had an effect on number of interpretations, we divided the consistent and inconsistent interpretations into whether the priming words were in the same order or reverse order to that of the compound. No significant differences were found. Furthermore, the priming effect was still present within the priming order conditions. That is, when prime order was the same, participants gave significantly more consistent interpretations (mean = 3.20) than inconsistent interpretations (mean = 2.32), $z = 2.77$, $p = .006$. Likewise, when prime order was reversed, participants again gave significantly more consistent interpretations (mean = 3.67) than inconsistent interpretations (mean = 2.40), $z = 3.34$, $p = .001$. Overall, these results provide strong evidence that the priming was effective, and that it is independent of priming order.

A.2. Response time

The speed of producing an interpretation was analysed according to whether it was consistent or inconsistent with regards to the priming words, and whether this was affected by prime order. It was expected that if the priming was effective then interpretations that were inconsistent with the primes would be produced slower than interpretations that were consistent with the primes. As seen in Fig. A.5, the mean response times were in the correct direction. Since a number of participants did not give responses for all of the categories, the number of participants in the analysis was 51. The analysis showed no main effect of Interpretation ($p = 0.297$), Prime Order ($p = 0.718$), nor an Interpretation x Prime Order interaction ($p = 0.994$). One likely reason for the non-significant effects is the large variance in response times (range = 369–10,035 ms), thus making it difficult for the mean differences to reach significance. For this reason we feel that the frequency scores are more reliable measures, and importantly these showed significant effects of priming.

a

	Consistent	Inconsistent
Overall	3095.51	3288.99
Same Order	3066.47	3273.25
Reverse Order	3083.10	3299.40

b

	Consistent	Inconsistent
Overall	3138.76	3352.58
Same Order	3155.78	3329.85
Reverse Order	3098.23	3274.58

Fig. A.5. Mean response time for producing interpretations (consistent or inconsistent) with the primes by prime order (overall, same prime order, reverse prime order) (a) mean response times (ms) before analysis ($N = 65$) (b) mean response times (ms) used in ANOVA ($N = 51$).

A.3. Compound familiarity

One concern is that the evidence for non-compositionality found in this study may be a function of familiarity. In particular, highly familiar compounds would be expected to require less combinatorial processing as the combined meaning may simply be retrieved from long term memory. We consider this possibility unlikely due to the experimental procedure followed. The fact that both words are ambiguous allows the priming procedure to shift participants into considering new combined meanings. For instance, while most participants (86%) interpreted SPRING PLANT as "a plant that grows in spring", when primed with 'coil' and 'leaf', 3% of participants gave the interpretation "a springy plant". Thus these participants have arguably been influenced by priming towards generating a new meaning, even though a highly common meaning already exists. In fact, as previously mentioned for spring plant and other compounds the findings of non-compositionality seem to depend upon participants producing novel meanings for the compounds. This finding goes against the hypothesis that non-compositionality is driven entirely by the retrieval of pre-stored meanings. To test whether familiarity is associated with non-compositionality, we obtained hit rates for each compound by typing each into Google with quotes. This measure of familiarity has been used in previous studies, e.g., (Ramm & Halford, 2012; Wisniewski & Murphy, 2005). It was found that the novelty of compounds based upon hit rates ranged from 144 (STAG YARN) to 9,460,000 (BATTERY CHARGE). To reduce the large variance obtained in the hit rates we transformed the scores into logs of ten. If familiarity is driving the non-compositionality results it would be expected that CHSH scores would be positively correlated with Google hit rates. To test this we calculated a Pearson R correlation. This showed a weak positive correlation between the two variables, though this was non-significant, $r = 0.21, p = .337$. Thus we did not find evidence for the hypothesis that the non-compositionality of compounds in this study is driven by familiarity. However, as there were only 24 compounds under study, we acknowledge that there may not have been enough power to derive a significant correlation.

More generally, the primes are an experimentally pragmatic means to manipulate the manner in which context affects the interpretation applied to conceptual combinations, and so they need only influence the interpretation, not determine it. The violations that do occur arise only with respect to the reported priming conditions, and may not occur in a different experimental context.

References

Aerts, D. (2009). Quantum structure in cognition. *Journal of Mathematical Psychology*, 53(5), 314–348.

Aerts, D., Aerts, S., Broeckaert, J., & Gabora, L. (2000). The violation of Bell inequalities in the macroworld. *Foundations of Physics*, 30, 1378–1414.

Aerts, D., Broeakert, J., Czachor, M., Kuna, M., Sinervo, B., & Sozzo, S. (2014). Quantum structure in competing lizard communities. *Ecological Modeling*, 281, 38–51.

Aerts, D., Gabora, L., & Sozzo, S. (2013). Concepts and their dynamics: A quantum-theoretic modeling of human thought. *Topics in Cognitive Science*, 5(4), 737–772.

Bell, J. S. (1964). On the Einstein–Podolsky–Rosen–Bohm paradox. *Physics*, 1(3), 195–200.

Blutner, R., Pothos, E. M., & Bruza, P. (2013). A quantum probability perspective on borderline vagueness. *Topics in Cognitive Science*, 5(4), 711–736.

Bruza, P., Busemeyer, J., & Gabora, L. (2009). Introduction to the special issue on quantum cognition. *Journal of Mathematical Psychology*, 53.

Bruza, P., Kitto, K., Nelson, D., & McEvoy, C. (2009). Is there something quantum-like about the human mental lexicon? *Journal of Mathematical Psychology*, 53, 362–377.

Busemeyer, J., & Bruza, P. (2012). *Quantum cognition and decision*. Cambridge University Press.

Busemeyer, J., Pothos, E., Franco, R., & Trueblood, J. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological Review*, 118(2), 193–218.

Cereceda, J. (2000). Quantum mechanical probabilities and general probabilistic constraints for Einstein–Podolsky–Rosen–Bohm experiments. *Foundations of Physics Letters*, 13(5), 427–442.

Costello, F., & Keane, M. (1997). Polysemy in conceptual combination: Testing the constraint theory of combination. In *Nineteenth annual conference of the cognitive science society*. Erlbaum.

Costello, F., & Keane, M. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, 24(2), 299–349.

Dzhafarov, R., & Kujala, J. (2012). Selectivity in probabilistic causality: Where psychology runs into quantum physics. *Journal of Mathematical Psychology*, 56(1), 54–63.

Dzhafarov, E., & Kujala, J. (2014). On selective influences, marginal selectivity, and Bell/CHSH inequalities. *Topics in Cognitive Science*, 6(1), 121–128.

Fine, A. (1982). Joint distributions, quantum correlations and commuting observables. *Journal of Mathematical Physics*, 23(7), 1306–1310.

Fodor, J. (1998). *Concepts, where cognitive science went wrong. Oxford cognitive science series*. Oxford University Press.

Frixione, M., & Lieto, A. (2012). Representing concepts in formal ontologies: Compositionality vs. typicality effects. *International Journal of Logic and Logic Philosophy*, 21(4), 391–414.

Gagne, C. (2001). Relation and lexical priming during the interpretation of noun–noun combinations. *Journal of Experimental Psychology*, 27(1), 236–254.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. MIT Press.

Hampton, J. (1997). Conceptual combination. In K. Lamberts, & D. Shank (Eds.), *Knowledge, concepts, and categories* (pp. 133–160). MIT Press.

Isham, C. (1995). *Lectures on quantum theory*. Imperial College Press.

Khrennikov, A. (2010). *Ubiquitous quantum structure: From psychology to finance*. Springer.

Kitto, K. (2014). A Contextualised general systems theory. *Systems*, 2(4), 541–565.

Laloë, F. (2001). Do we really understand quantum mechanics? Strange correlations, paradoxes, and theorems. *American Journal of Physics*, 69(6), 655–701.

Maudlin, T. (1994). *Quantum non-locality and relativity: metaphysical intimations of modern physics*. Blackwell Publishers Limited.

Medin, D., & Shoben, E. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, 20, 58–190.

Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, 34, 1388–1429.

Murphy, G. (1988). Comprehending complex concepts. *Cognitive Science*, 12, 529–562.

Murphy, G. (2002). *The big book of concepts*. MIT Press.

Nelson, D., Kitto, K., Galea, D., McEvoy, C., & Bruza, P. (2013). How activation, entanglement, and searching a semantic network contribute to event memory. *Memory and Cognition*, 41, 717–819.

Nelson, D., McEvoy, C., & Schreiber, T. (2004). The University of South Florida, word association, rhyme and word fragment norms. *Behavior Research Methods, Instruments & Computers*, 36, 408–420.

Partee, B. (1995). Lexical semantics and compositionality. *An invitation to cognitive science: Language*, 1, 311–360.

Pelletier, J. (1994). The principle of semantic compositionality. *Topoi*, 13, 11–24.

Pitowsky, I. (1994). George Boole's 'conditions of possible experience' and the quantum puzzle. *The British Journal for the Philosophy of Science*, 45(1), 95–125.

Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. *Psychological Review*, 120, 679–696.

Ramm, B. (2000). *Pet rocks, tame robots and desert fish: Developmental differences in the understanding of combined concepts* (Unpublished Honours thesis). University of Adelaide.

Ramm, B., & Halford, G. (2012). Novelty and processing demands in conceptual combination. *Australian Journal of Psychology*, 64(4), 199–208.

Suppes, P., Acacio de Barros, J., & Oas, G. (1998). A collection of probabilsitic hidden variables theorems and counterexamples. In R. Pratesi, & L. Ronchi (Eds.), *Waves, information and foundations of physics, conference proceedings: Vol. 60.* (pp. 267–291). Italian Physical Society.

Swinney, D., Love, T., Walenski, M., & Smith, E. (2007). Conceptual combination during sentence comprehension: Evidence for compositional processes. *Psychological Science*, 18(5), 397–400.

Twilley, L., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition*, *22*(1), 111–126.

Weiskopf, D. (2007). Compound nominals, context and compositionality. *Synthese*, *156*, 161–204.

Wilkenfeld, M., & Ward, T. (2001). Similarity and emergence in conceptual combination. *Journal of Memory and Language*, *45*, 21–38.

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language*, *35*(3), 435–453.

Wisniewski, E. J., & Murphy, G. (2005). Frequency of relation type as a determinant of conceptual combination: A reanalysis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*(1), 169–174.

Zadrozny, W. (1992). On compositional semantics. In *Proceedings of the international conference on computational linguistics (COLING-92)*. (pp. 260–266).