

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*

tel: +61-7-31389325

fax: +61-7-31384438

^b*School of Philosophy, Australian National University*

^c*School of Computer Science and Electrical Engineering, Queensland University of Technology, Australia*

Abstract

Conceptual combination performs a fundamental role in creating the broad range of compound phrases utilised in everyday language. 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. While the systematicity and productivity of language provide a strong argument in favor of assuming compositionality, this very assumption is still regularly questioned in both cognitive science and philosophy. Additionally, the principle of semantic compositionality is underspecified, which means that notions of both “strong” and “weak” compositionality appear in the literature. 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. In addition, we suggest that the distinction between these is contextually sensitive. Compositionality is equated with a joint

*Corresponding author

Email addresses: p.bruza@qut.edu.au (Peter D. Bruza), kirsty.kitto@qut.edu.au (Kirsty Kitto), brentynramm@gmail.com (Brentyn J. Ramm), laurianne.sitbon@qut.edu.au (Laurianne Sitbon)

probability distribution modeling how the constituent concepts in the combination are interpreted. Marginal selectivity is introduced as a pivotal probabilistic constraint for the application of the Bell/CH and CHSH systems of inequalities. Non-compositionality is equated with a failure of marginal selectivity, or violation of either system of inequalities in the presence of marginal selectivity. This means that the conceptual combination cannot be modeled in a joint probability distribution, the variables of which correspond to how the constituent concepts are being interpreted. The formal analysis methods are demonstrated by applying them to an empirical illustration of twenty-four non-lexicalised conceptual combinations.

Keywords: conceptual combination, semantic compositionality, quantum cognition

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 et al., 2004). Such cases have been used by some authors 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.

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 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, 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 (Frixione and Lieto, 2012; Fodor, 1998). 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 (Frixione and Lieto, 2012; Fodor, 1998). 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 (Hampton, 1997; Fodor, 1998). 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 and Keane, 2000; Mitchell and Lapata, 2010). This is a principle underlying many general theories of language, both natural and artificial. 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 behavior 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., $black(x) \wedge cat(x)$. Such intersective semantics are compositional, as the semantics of BLACK CAT are determined solely in terms of the seman-

tics of the constituent concepts BLACK and CAT. It is tempting to assume that most conceptual combinations can be modeled 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 being a concept 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., color, size, shape etc.) and relational dimensions (e.g., habitat, functions, behaviors 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 and Ward, 2001), and so is sometimes viewed as evidence for non-compositional semantics (Hampton, 1997; Medin and 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 (Wisniewski, 1996; Swinney et al., 2007). 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), because similar concepts share many dimensions (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. Thus an ELEPHANT BOX could be deemed as likely to be made of a strong material such as wood, and hopefully to contain air-holes. [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 the 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. An important point is that this framework can be empirically tested. Thus, we feel that it is possible to shift this debate out of philosophy and into the realms of experimental psychology,¹ and this article is a step in that direction. In what follows we shall

¹In 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](#)).

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 modeling 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 doesn't 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](#)). Opponents to Zadrozny may argue that his position is overly pessimistic because it applies to "strong compositionality", a position that is clearly wrong. Nevertheless, the question remains as to where the "meanings" might come from initially.

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. Over large numbers of subjects, probabilities can be calculated that a certain associate is produced. Fig. 1(a) depicts such a set 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

<i>Associate</i>	<i>Probability</i>
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
...	...

(a)

<i>Associate</i>	<i>Probability</i>
fighter	0.14
gloves	0.14
fight	0.09
dog	0.08
shorts	0.07
punch	0.05
Tyson	0.05
...	...

(b)

Figure 1: (a) Free association probabilities for the word “boxer” (a) and the word “bat” (b).

an ANIMAL sense. Considering the full dataset² 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).

When the preceding is considered in relation to how the conceptual combination BOXER BAT may be interpreted, then four interpretations are possible. For example, when BOXER is interpreted as a sport and BAT as an animal, the corresponding interpretation of the combination maybe 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 as a sport leading a subject to interpret the compound as “a fighter’s implement”.

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 sports person, 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

²Available at <http://web.usf.edu/FreeAssociation/AppendixC/Matrices.A-B> .

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 behavior that can be considered as compositional, then it should be possible to describe this behavior in terms of four dichotomous random variables, $\{\mathbf{A1}, \mathbf{A2}\}$ and $\{\mathbf{B1}, \mathbf{B2}\}$, ranging over two values $\{+1, -1\}$. The numbers 1 (dominant) and 2 (subordinate) correspond to the senses attributed to the respective concepts A and B . However, if a human subject is first shown the word “vampire” and subsequently asked to interpret the compound BOXER BAT, then they may be oriented towards giving an animal interpretation of BAT. This suggests a minimal natural extension where $\mathbf{A1} = +1$ represents a situation where the dominant sense of concept A was first primed and concept A was indeed subsequently interpreted in that sense by the human subject. Conversely, $\mathbf{A1} = -1$ represents the case where the dominant sense of concept A was primed but A was not interpreted in that sense. Similarly, $\mathbf{A2} = +1$ represents a situation where a subordinate sense of concept A was first primed, and concept A was indeed subsequently interpreted in this sense, and $\mathbf{A2} = -1$ represents the case where a subordinate sense of concept A was primed, but A was not interpreted in this subordinate sense. Note that a concept may have more than one subordinate sense. For example, the concept BOXER could be

considered to have subordinate clothing sense, namely “boxer shorts”. Therefore, the previous specification based on a primary and subordinate sense does not preclude that the concept A is interpreted in a third sense, for example, when $\mathbf{A1} = -1$, this can occur when concept is interpreted in the subordinate sense modeled by $\mathbf{A2}$, or a third sense. Similar relationships hold for $\mathbf{B1}$ and $\mathbf{B2}$.

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

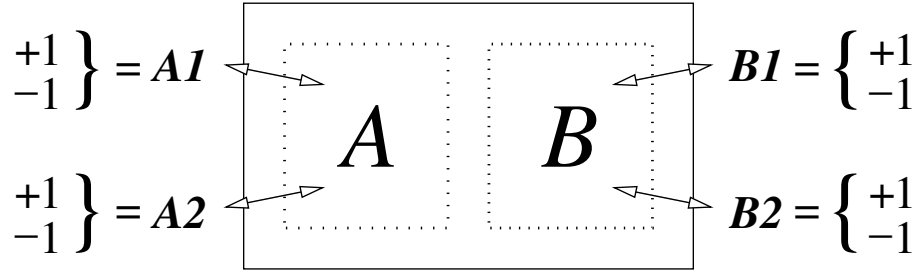


Figure 2: A potentially compositional system S , consisting of two assumed components A and B . S can perhaps be understood in terms of a mutually exclusive choice of experiments performed upon those components, one represented by the random variables $\mathbf{A1}$, $\mathbf{A2}$ (pertaining to an interaction between the experimenter and component A), and the other by $\mathbf{B1}$, $\mathbf{B2}$ (pertaining to an interaction between the experimenter and component B). Each of these experiments can return a value of +1 or -1

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 \mathbf{A} that corresponds to the prime $\mathbf{A1}$ and similarly for \mathbf{B} in relation to the prime $\mathbf{B2}$. 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 $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$ which is used to model the behavior 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 we can construct 16 joint probabilities, corresponding to all the possible interpretations of concepts A and B that a subject might return, across the four priming conditions:

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

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

$$\begin{array}{r}
 \begin{array}{cc}
 & B \\
 & \mathbf{B1} \quad \quad \mathbf{B2} \\
 & +1 \quad -1 \quad \quad +1 \quad -1
 \end{array} \\
 A \begin{array}{cc}
 \mathbf{A1} & +1 \\
 & -1 \\
 \mathbf{A2} & +1 \\
 & -1
 \end{array}
 \left(\begin{array}{cc|cc}
 p_1 & p_2 & p_5 & p_6 \\
 p_3 & p_4 & p_7 & p_8 \\
 \hline
 p_9 & p_{10} & p_{13} & p_{14} \\
 p_{11} & p_{12} & p_{15} & p_{16}
 \end{array} \right) = P_{AB} \quad (2)
 \end{array}$$

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 P_{AB} in equation (2) is complete, in that it covers all possible priming conditions across the respective senses of the concepts.

In what follows we will show that P_{AB} 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. For example, a naive assumption would be that the concepts in the combination can be interpreted independently of one another,

$$\Pr(\mathbf{Ai}, \mathbf{Bj}) = \Pr(\mathbf{Ai}) \Pr(\mathbf{Bj}), i, j \in \{1, 2\}. \quad (3)$$

The syntax of this equation clearly reflects how the probabilistic behaviour of the conceptual combination is represented by the four joint distributions $\Pr(\mathbf{A}_i, \mathbf{B}_j)$.

Less naive formalisations of compositionality are possible. For example, the competition among relations in nominals (CARIN) theory of conceptual combination proposes sixteen possible relations that can be used to link concepts, e.g., *causes, during, for* and *about* (Gagne and Shoben, 1997; Gagne, 2001). One possible assumption assumes that relation independently affects the interpretation of each concept:

$$\Pr(\mathbf{A}_i, \mathbf{B}_j | \mathbf{R}) = \Pr(\mathbf{A}_i | \mathbf{R}) \Pr(\mathbf{B}_j | \mathbf{R}), i, j \in \{1, 2\}. \quad (4)$$

where \mathbf{R} is a random variable over the possible implicit linking relations. CARIN, however, assumes that when people interpret a novel conceptual combination AB , the availability of each of these sixteen relations is determined by the relation frequency distribution for the modifying concept A . The random variable \mathbf{R} is thus assumed to range over these sixteen relations. Therefore, CARIN explains why MOUNTAIN GOAT is easier to interpret than MOUNTAIN MAGAZINE because the *located* relation is more often used with the modifier MOUNTAIN than the *about* relation. The essence of the CARIN approach to interpreting conceptual combinations can therefore be derived as follows using using (4) together with the assumption that the interpretation of concept B is independent of the linking relation,

$$\Pr(\mathbf{A}_i, \mathbf{B}_j) = \sum_{r \in \mathbf{R}} \Pr(\mathbf{A}_i, \mathbf{B}_j | r) \Pr(r) \quad (5)$$

$$= \sum_{r \in \mathbf{R}} \Pr(\mathbf{A}_i | r) \Pr(\mathbf{B}_j | r) \Pr(r) \quad (6)$$

$$= \sum_{r \in \mathbf{R}} \Pr(\mathbf{A}_i, r) \Pr(\mathbf{B}_j) \quad (7)$$

for $i, j \in \{1, 2\}$. In other words, a formalisation of the CARIN model is as follows which shows how the the relation frequency distribution for the modifying concept A is formalised by the joint distribution $\Pr(\mathbf{A}_i, \mathbf{R})$. The goal

of the preceding development of equations is not to formalise existing models, or propose new ones, but to introduce how compositionality may be defined in a probabilistic way. Observe that the equations (3)(4),(7) all define four pairwise joint probability distributions $\Pr(\mathbf{A}_i, \mathbf{B}_j), i, j \in \{1, 2\}$.

A given conceptual combination AB is deemed *compositional* if and only if a four way joint distribution $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$ exists where $\Pr(\mathbf{A}_i, \mathbf{B}_j), i, j \in \{1, 2\}$ are marginal distributions. This opens the door to define non-compositionality via an unusual means, namely the inability to construct a joint probability distribution $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$ in this way.

2.2. Non-compositional semantics

To analyse non-compositionality we draw upon results from the field of quantum theory in relation to entangled systems. This step is not as arbitrary as it might at first seem. An entangled system is one for which it is impossible to construct a four way joint distribution from four empirically collected pairwise joint distributions. Fine's theorem is one of the results from quantum theory which allows entanglement to be formally defined in this way. Importantly, the theorem states the necessary and sufficient conditions for the notion of compositionality introduced at the end of the previous section.

Fine (1982) provides both the *necessary and sufficient conditions* for the existence of the joint probability distribution $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$:

Fine Theorem 3 (Fine, 1982): If $\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2}$ are bivalent random variables with joint distributions $\Pr(\mathbf{A}_i, \mathbf{B}_j), i, j \in \{1, 2\}$, then necessary and sufficient for a joint distribution $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$ is that the following sys-

tem of inequalities is satisfied:

$$-1 \leq \Pr(A1, B1) + \Pr(A1, B2) + \Pr(A2, B2) - \Pr(A2, B1) - \Pr(A1) - \Pr(B2) \leq 0 \quad (8)$$

$$-1 \leq \Pr(A2, B1) + \Pr(A2, B2) + \Pr(A1, B2) - \Pr(A1, B1) - \Pr(A2) - \Pr(B2) \leq 0 \quad (9)$$

$$-1 \leq \Pr(A1, B2) + \Pr(A1, B1) + \Pr(A2, B1) - \Pr(A2, B2) - \Pr(A1) - \Pr(B1) \leq 0 \quad (10)$$

$$-1 \leq \Pr(A2, B2) + \Pr(A2, B1) + \Pr(A1, B1) - \Pr(A1, B2) - \Pr(A2) - \Pr(B1) \leq 0. \quad (11)$$

Fine referred to this system of inequalities as the Bell/CH inequalities, and we will adhere to his labeling. Fine’s theorem permits us to analyse compositionality from the perspective of the above four equations, namely, a conceptual combination AB is deemed “non-compositional” when the four pairwise joint probability distributions depicted in table (2) do not satisfy the Bell/CH inequalities. This implies then a joint distribution $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$ *cannot* be formed such that the four pairwise joint probability distributions $\Pr(\mathbf{Ai}, \mathbf{Bj}), i, j \in \{1, 2\}$ are marginal distributions. Conversely, if all inequalities are satisfied the four way joint probability distribution does exist and the conceptual combination is thus deemed “compositional”.

In quantum physics, entangled systems adhere to a constraint variously termed the “causal communication constraint”, “parameter independence”, “simple locality”, “signal locality”, or “physical locality”. This constraint expresses “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). In the context of cognitive science, this constraint has been termed “marginal selectivity” (Dzhafarov and Kujala, 2012). 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”. Marginal selectivity is expressed

more formally as follows:

$$\Pr(Ai) = \Pr(Ai, B1) + \Pr(Ai, \bar{B}1) = \Pr(Ai, B2) + \Pr(Ai, \bar{B}2), i \in \{1, 2\} \quad (12)$$

$$\Pr(Bj) = \Pr(A1, Bj) + \Pr(\bar{A}1, Bj) = \Pr(A2, Bj) + \Pr(\bar{A}2, Bj), j \in \{1, 2\} \quad (13)$$

Note how these four equations express that the interpretation of the concept represented by the marginal probability, e.g., $\Pr(Ai), i \in \{1, 2\}$ is stable with respect to how the other concept is primed, represented by $B1$ and $B2$.

Recently, [Dzhafarov and Kujala \(2012\)](#) have established a connection between cognitive modeling and Fine’s theorem using the theory of selective influences, a result that adds to the cognitive validity of Fine’s theorem. 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 $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$ where the pairwise distributions $\Pr(\mathbf{A1}, \mathbf{B1})$, $\Pr(\mathbf{A1}, \mathbf{B2})$, $\Pr(\mathbf{A2}, \mathbf{B1})$, $\Pr(\mathbf{A2}, \mathbf{B2})$ are marginal distributions.

The proof of Fine’s theorem assumes marginal selectivity, a constraint that holds for entangled systems of photons 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. For applications in cognitive science, marginal selectivity must first be tested before Fine’s theorem can be 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/CH inequalities are vio-

lated, then the conceptual combination is deemed “non-compositional”.

3. If marginal selectivity holds and all of the Bell/CH inequalities ahold, then the conceptual combination is deemed “compositional”.

In quantum physics, the so called CHSH inequality has also been used to analyse entangled systems. The advantage of the CHSH inequality is that its formulation based on correlations permits some insight to be gained into why conceptual combinations are (non-)compositional. The CHSH inequality is as follows (Cereceda, 2000):

$$-2 \leq E(\mathbf{A1}, \mathbf{B1}) + E(\mathbf{A1}, \mathbf{B2}) + E(\mathbf{A2}, \mathbf{B1}) - E(\mathbf{A2}, \mathbf{B2}) \leq 2 \quad (14)$$

The expectations can easily be computed from the matrix of probabilities (2). For example, $E(\mathbf{A1}, \mathbf{B1}) = p_1 + p_4 - (p_2 + p_3)$. Recalling from (1) that $p_1 = \Pr(\mathbf{A1} = +1, \mathbf{B1} = +1)$ and $p_4 = \Pr(\mathbf{A1} = -1, \mathbf{B1} = -1)$, we recognize 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 APPLE was primed and food sense of CHIP was primed, perfect correlation of senses in this priming condition means two conditions hold: (1) when APPLE is interpreted as a fruit CHIP is always interpreted as food (p_1) and (2) when APPLE is *not* interpreted as fruit, CHIP is *not* interpreted as food (p_4). The combination of these two conditions imply 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. For example, assume the fruit sense of APPLE is primed and CHIP is primed in its electronic circuit sense. Perfect anti-correlation of senses means two conditions hold: (1) when APPLE is *not* interpreted as a fruit, CHIP is always interpreted as a circuit (p_3) and (2) when APPLE is interpreted as fruit, CHIP is *not* interpreted in its circuit sense (p_2).

With this as underlying intuition, the expectation value $E(\mathbf{A}_i, \mathbf{B}_j)$ captures how well 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, each arrangement leading to a variant of the CHSH inequality:

$$-2 \leq E(\mathbf{A1}, \mathbf{B1}) - E(\mathbf{A1}, \mathbf{B2}) + E(\mathbf{A2}, \mathbf{B1}) + E(\mathbf{A2}, \mathbf{B2}) \leq 2 \quad (15)$$

$$-2 \leq E(\mathbf{A1}, \mathbf{B1}) + E(\mathbf{A1}, \mathbf{B2}) - E(\mathbf{A2}, \mathbf{B1}) + E(\mathbf{A2}, \mathbf{B2}) \leq 2 \quad (16)$$

$$-2 \leq -E(\mathbf{A1}, \mathbf{B1}) + E(\mathbf{A1}, \mathbf{B2}) + E(\mathbf{A2}, \mathbf{B1}) + E(\mathbf{A2}, \mathbf{B2}) \leq 2 \quad (17)$$

$$(18)$$

Therefore, there are four CHSH inequalities, each differing where the minus sign is placed. The heart of each inequality is a computation involving expectations 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, there is no joint probability distribution $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$ such that the four empirically collected pairwise distributions $\Pr(\mathbf{A1}, \mathbf{B1})$, $\Pr(\mathbf{A1}, \mathbf{B2})$, $\Pr(\mathbf{A2}, \mathbf{B1})$, $\Pr(\mathbf{A2}, \mathbf{B2})$ 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 $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$ where the four empirically collected pairwise distributions: $\Pr(\mathbf{A1}, \mathbf{B1})$, $\Pr(\mathbf{A1}, \mathbf{B2})$, $\Pr(\mathbf{A2}, \mathbf{B1})$, $\Pr(\mathbf{A2}, \mathbf{B2})$ are marginal distributions. In this case, the conceptual combination is deemed “compositional”.

As was the case with Bell/CH inequalities, marginal selectivity must first be tested before the four CHSH inequalities can be applied. The formal connection between the CHSH inequalities and Fine’s theorem is as follows: By assuming marginal selectivity together with the CHSH inequalities, the Bell/CH inequalities can be derived.

In summary, both the Bell/CH inequalities and the CHSH inequalities allow

non-compositionality to be determined by the inability to construct a joint probability distributions across the four variables modeling 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. 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 et al., 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 concept as an associate. 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’. 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 was associated with both BANK and LOG, thereby minimising the possibility of priming more than one concept at a time.

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” (**A1 – B1**), (2) “money” and “tree” (**A1 – B2**), (3) “river” and “journal” (**A2 – B1**), or (4) “river” and “tree” (**A2 – 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. Combinations were chosen with the expectation that the ambiguity of its 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 and Keane, 1997).

3.3. Procedure

Participants completed 3 practice trials, 24 test trials and 24 filler trials, and Fig. 3 shows a schematic illustration of the procedure followed during a test trial. 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 utilised in order to maximise the effectiveness of the priming. The experiment took around 20–30 minutes 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

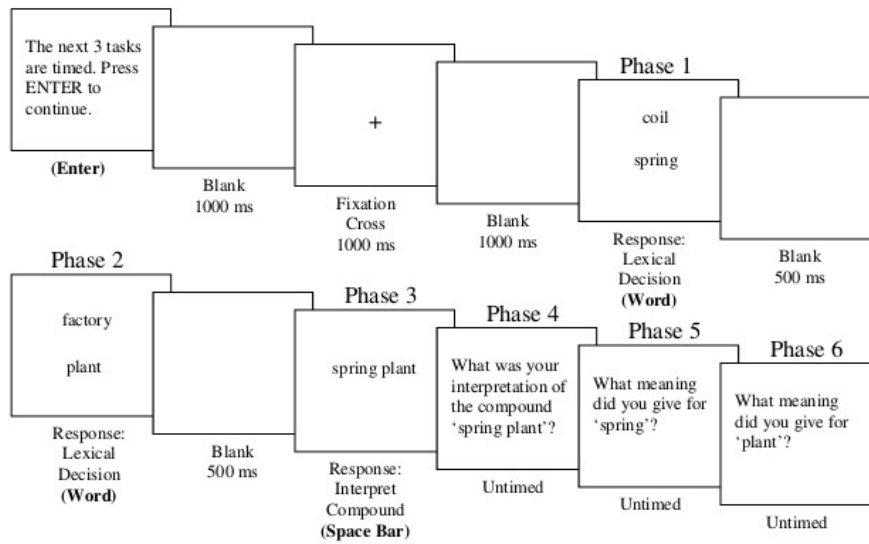


Figure 3: Example experimental structure for a trial. Non-word trials followed a similar structure, with primes in Phase 1 and/or Phase 2 replaced with non-words. The sequence of squares moving from left to right show the experimental flow, with each square a representation of the screen shown to a participant. Note: the figure does not show the exact text given to participants, and stimuli are not to scale.

decision consisted of the the two letter strings presented in the center 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, labeled ‘word’ or a button labeled ‘non-word’ (left arrow and right arrow 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 were: 50% word \rightarrow word (test trial), 25% non-word \rightarrow non-word (filler trial), 12.5% word \rightarrow non-word (filler trial), 12.5% non-word \rightarrow 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 center 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 utilizing 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 Fine’s theorem or the CHSH 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. For example, $\text{diff}(\mathbf{A1})$ is the difference between the one-marginal $\Pr(A1, B1) + \Pr(A1, \bar{B}1)$ and the other second marginal $\Pr(A1, B2) + \Pr(A1, \bar{B}2)$. Marginal selectivity holds when these differences are zero across all for variables. The statistical test reveals five conceptual combinations that fail marginal selectivity. However, as there are usually sixteen data points in each of the four pair-wise distributions (see equation (2)), the statistical test is underpowered. As a consequence, there is uncertainty about which combinations are failing marginal selectivity. For the purposes of illustration, the four conceptual combinations BATTERY CHARGE, BILL SCALE, TABLE FILE and TOAST GAG are assumed to satisfy marginal selectivity, as these are the four combinations closest to satisfying marginal selectivity based on the sum of their Chi square values (see table 1).

The result of the compositional analysis is depicted in Table 2. Despite the findings of the analysis of marginal selectivity, we have adopted a conservative approach and flagged combinations as “non-compositional” based solely on the assumption that they are likely to fail marginal selectivity. Of the combina-

tions that are assumed to satisfy marginal selectivity, BILL SCALE ($-\text{CHSH}-=1.63$), TOAST GAG ($-\text{CHSH}-=1.63$) and TABLE FILE ($-\text{CHSH}-=0.33$) are deemed “compositional” as their CHSH values are less than or equal to 2. BATTERY CHARGE shows a slight violation of the CHSH inequalities ($-\text{CHSH}-=2.01$), but due to the lack of statistical power just mentioned, we cannot conclude whether this combination does represent an actual case where non-compositionality manifests in the presence of marginal selectivity being satisfied. Only a subsequent study with sufficient statistical power can resolve whether such conceptual combinations exist.

Combination	diff (A1)	diff (A2)	diff (B1)	diff (B2)
boxer bat	0.175 (0.46)	0.140 (0.24)	0.338 (2.6)	0.158 (0.27)
bank log	0.055 (0)	0.092 (0.02)	0.338 (3.30*)	0.257 (1.73)
apple chip*	0.250 (2.77*)	0.114 (0.09)	0.294 (2.88*)	0.217 (0.78)
stock tick*	0.163 (0.30)	0.085 (0.02)	0.488 (5.59*)	0.386 (3.77*)
seal pack	0.083 (0.01)	0.213 (0.77)	0.162 (0.38)	0.221 (0.78)
spring plant*	0.294 (3.49*)	0.133 (0.61)	0	0.173 (0.81)
poker spade	0.136 (0.25)	0.035 (0)	0 (0)	0.113 (0.09)
slug duck	0.096 (0.03)	0.153 (0.32)	0.133 (0.21)	0.026 (0)
club bar	0.133 (0.68)	0 (0)	0.125 (0.60)	0.138 (0.37)
web bug	0.210 (0.74)	0.067 (0)	0.296 (1.70)	0.153 (0.32)
table file	0.058 (0)	0.235 (0)	0.114 (0.09)	0.113 (0.09)
match bowl	0.137 (0.18)	0.250 (1.31)	0.075 (0.01)	0.022 (0)
net cap	0.035 (0)	0.092 (0.03)	0.059 (0)	0.175 (0.46)
stag yarn*	0.375 (3.64*)	0.219 (1.14)	0.104 (0.05)	0.045 (0)
mole pen	0.125 (0.29)	0.021 (0)	0.063 (0)	0.3 (1.87)
battery charge	0.067 (0)	0.048 (0)	0.117 (0)	0.120 (0.08)
count watch	0.195 (0.89)	0.063 (0)	0.011 (0)	0.063 (0)
bill scale	0.081 (0.02)	0.113 (0.09)	0.054 (0)	0.051 (0)
rock strike	0.188 (1.47)	0.117 (0.13)	0.313 (3.79)	0.013 (0)
port vessel	0.106 (0.07)	0.085 (0.02)	0.113 (0.09)	0.118 (0.20)
crane hatch	0.141 (0.45)	0.296 (1.70)	0.149 (0.39)	0.233 (0.93)
toast gag	0.0625 (0)	0.008 (0)	0.018 (0)	0.015 (0)
star suit	0.308 (2.63)	0.163 (0.28)	0.054 (0)	0.058 (0)
fan post	0.35 (2.59)	0.125 (0.13)	0.025 (0)	0.188 (0.55)

Table 1: Analysis of marginal selectivity. Starred conceptual combinations fail marginal selectivity (Chi square test of proportions: critical value=2.71 ($\alpha = 0.1$)). Bolded conceptual combinations look to satisfy marginal selectivity

3.5. Discussion

In this discussion further details are provided in order to shed light on *how* the joint probability distribution is structured when a violation occurs which serves to illustrate a number of key features about non-compositionality. In what follows, we shall utilize two examples: TOAST GAG and APPLE CHIP.

	Concept A		Concept B		Results	
Combination	Prime 1 (A1)	Prime 2 (A2)	Prime 3 (B1)	Prime 4 (B2)	<i>Compositional</i>	<i>N</i>
boxer bat	<i>dog</i>	<i>fighter</i>	<i>ball</i>	<i>vampire</i>	N	64
bank log	<i>money</i>	<i>river</i>	<i>journal</i>	<i>tree</i>	N	64
apple chip	<i>banana</i>	<i>computer</i>	<i>potato</i>	<i>circuit</i>	N	65
stock tick	<i>shares</i>	<i>cow</i>	<i>mark</i>	<i>fla</i>	N	64
seal pack	<i>walrus</i>	<i>envelop</i>	<i>leader</i>	<i>suitcase</i>	N	64
spring plant	<i>summer</i>	<i>coil</i>	<i>leaf</i>	<i>factory</i>	N	64
poker spade	<i>card</i>	<i>fire</i>	<i>ace</i>	<i>shovel</i>	N	65
slug duck	<i>snail</i>	<i>punch</i>	<i>quack</i>	<i>dodge</i>	N	63
club bar	<i>member</i>	<i>golf</i>	<i>pub</i>	<i>handle</i>	N	64
web bug	<i>spider</i>	<i>internet</i>	<i>beetle</i>	<i>computer</i>	N	63
table file	<i>chair</i>	<i>chart</i>	<i>nail</i>	<i>folder</i>	Y [0.33]	63
match bowl	<i>flame</i>	<i>contest</i>	<i>disk</i>	<i>throw</i>	N	64
net cap	<i>gain</i>	<i>volleyball</i>	<i>limit</i>	<i>hat</i>	N	65
stag yarn	<i>party</i>	<i>deer</i>	<i>story</i>	<i>wool</i>	N	61
mole pen	<i>dig</i>	<i>face</i>	<i>pig</i>	<i>ink</i>	N	63
battery charge	<i>car</i>	<i>assault</i>	<i>volt</i>	<i>prosecute</i>	? [2.01]	63
count watch	<i>number</i>	<i>dracula</i>	<i>time</i>	<i>look</i>	Y	65
bill scale	<i>phone</i>	<i>pelican</i>	<i>weight</i>	<i>fish</i>	Y [1.63]	64
rock strike	<i>stone</i>	<i>music</i>	<i>hit</i>	<i>union</i>	N	63
port vessel	<i>harbour</i>	<i>wine</i>	<i>ship</i>	<i>bottle</i>	N	65
crane hatch	<i>lift</i>	<i>bird</i>	<i>door</i>	<i>egg</i>	N	63
toast gag	<i>jam</i>	<i>speech</i>	<i>choke</i>	<i>joke</i>	Y [1.23]	63
star suit	<i>moon</i>	<i>movie</i>	<i>vest</i>	<i>law</i>	N	62
fan post	<i>football</i>	<i>cool</i>	<i>mail</i>	<i>light</i>	N	63

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 classified as compositional are bolded with associated CHSH values in brackets

TOAST GAG

Table (19) depicts the 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 (18) and the the actual values in table (19) we can see that the probability mass does not center sufficiently around the diagonals in such a way that it can produce the correlations between the senses necessary to violate the CHSH inequality as $|\text{CHSH} = 1.23| \leq 2$. The conceptual combination TOAST GAG is therefore deemed to be “compositional” as a joint probability distribution $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$ can be constructed, which models

how it is interpreted within the given priming conditions.

$$\begin{array}{c}
 \text{TOAST} \\
 \mathbf{A1}(jam) \\
 \mathbf{A2}(speech)
 \end{array}
 \begin{array}{cc}
 +1 \\
 -1 \\
 +1 \\
 -1
 \end{array}
 \begin{array}{c}
 \mathbf{B2}(choke) \\
 \mathbf{B1}(joke)
 \end{array}
 \begin{array}{cccc}
 & +1 & -1 & +1 & -1 \\
 \left(\begin{array}{cc|cc}
 0.50 & 0.4375 & 0.625 & 0.375 \\
 0.0625 & 0 & 0 & 0 \\
 \hline
 0.29 & 0 & 0.07 & 0.21 \\
 0.29 & 0.42 & 0.57 & 0.14
 \end{array} \right)
 \end{array}
 \quad (19)$$

APPLE CHIP

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

$$\begin{array}{c}
 \text{APPLE} \\
 \mathbf{A1}(banana) \\
 \mathbf{A2}(computer)
 \end{array}
 \begin{array}{cc}
 +1 \\
 -1 \\
 +1 \\
 -1
 \end{array}
 \begin{array}{c}
 \mathbf{B1}(potato) \\
 \mathbf{B2}(circuit)
 \end{array}
 \begin{array}{cccc}
 & +1 & -1 & +1 & -1 \\
 \left(\begin{array}{cc|cc}
 0.94 & 0.06 & 0 & 0.75 \\
 0 & 0 & 0.25 & 0 \\
 \hline
 0 & 0.35 & 0.47 & 0 \\
 0.65 & 0 & 0 & 0.53
 \end{array} \right)
 \end{array}
 \quad (20)$$

It is clear from the values that APPLE CHIP fails marginal selectivity. Therefore the joint probability distribution $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$ cannot be constructed from the four empirically collected pairwise joint probability distributions such that these four pairwise distributions depicted in table (20) can be recovered by marginalising this four way joint distribution. This conceptual combination is therefore deemed “non-compositional”.

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 $p_2 = \Pr(\mathbf{A1} = +1, \mathbf{B1} = -1) = 0.06$ (see Table (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 combinations similarly showed 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 be deemed non-compositional when marginal selectivity is satisfied, is for the probability mass to be largely concentrated along diagonals together with off-diagonal elements with small probabilities. These small probabilities reflect the atypical interpretations. (Costello and Keane, 2000). It is interesting to note that BATTERY CHARGE, which is borderline between “compositional” and “non-compositional”, does exhibit this structure. Based on this structure we generated a hypothetical example (N=400), with one hundred data points in each quadrant (See equation (21)). The Chi square values for the differences in marginal selectivity are (0.06,0,0,0) for $\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2}$ respectively. Hence there are minor differences in marginal probabilities, where the differences are not significant at the 90% level. The probabilities in equation (21) yield an absolute CHSH value of 2.06, therefore this hypothetical conceptual combination can

be classified as “non-compositional”. The atypical interpretations, highlighted by the bolded probabilities, are what force the CHSH value to exceed the threshold of two. It is this threshold that marks the border between compositionality and non-compositionality. It is not surprising that the absolute CHSH value is only slightly above 2, as by their very nature, atypical interpretations are infrequent. .

$$\begin{array}{c}
 \text{A} \\
 \text{A1}(primea1) \\
 \text{A2}(primea2)
 \end{array}
 \begin{array}{c}
 +1 \\
 -1 \\
 +1 \\
 -1
 \end{array}
 \begin{array}{c}
 \text{B} \\
 \text{B1}(primeb1) \\
 \text{B2}(primeb2)
 \end{array}
 \begin{array}{c}
 +1 \quad -1 \\
 +1 \quad -1
 \end{array}
 \left(\begin{array}{cc|cc}
 0.85 & \mathbf{0.05} & 0 & 0.92 \\
 0 & 0.10 & 0.08 & 0 \\
 \hline
 0 & 0.06 & 0.07 & 0 \\
 0.86 & \mathbf{0.08} & 0 & 0.93
 \end{array} \right) \quad (21)$$

Frequency of Interpretations

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

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

Figure 4: Mean Number of Interpretations (Consistent or Inconsistent) with the Primes by Prime Order (Overall, Same Prime Order, Reverse Prime Order)

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.

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 table 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 = 369ms to 10035 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.

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

	Consistent	Inconsistent		Consistent	Inconsistent
Overall	3095.51	3288.99	Overall	3138.76	3352.58
Same Order	3066.47	3273.25	Same Order	3155.78	3329.85
Reverse Order	3083.10	3299.40	Reverse Order	3098.23	3274.58

(a)

(b)

Figure 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)

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 and Halford, 2012; Wisniewski and 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.

4. Broader reflections on compositionality and non-compositionality

[Costello and Keane \(2000\)](#) 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 modeled as consisting of both a dominant ANIMAL (**A1**) and a subordinate PLANE (**A2**) sense, and then Bell/CH inequalities or the CHSH inequalities applied to test for the non-compositionality of each combination resulting from a combination of SHOVEL with BIRD.

In addition, there is no requirement in the presented analytical framework that the concepts be homographs. We require only that there be ambiguity caused by multiple possible interpretations of a concept, and this readily presents. 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 these may even be related (as was the case for the polysemous concept SHOVEL). Ambiguity could also derive from relations. For example, the CARIN model assumes that relations apply to the modifier, so in ADOLESCENT DOCTOR (taken from (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 correspond-

ing 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 PhD. Both of these possibilities could be modeled by the variables **B1** and **B2**. This example shows that the analytical framework presented here could be applied to the study of (non-)compositionality in conceptual combinations which have already been considered in the literature.

As the framework is general, and can be empirically tested, we argue that it has wide applicability for the analysis of conceptual combinations. 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; there is no result without a supplied context (in this case the priming). This is also the case in quantum physics; a system may be deemed compositional in one measurement context, and not in another. It is the theory for empirically testing a dividing line between compositionality and non-compositionality that this article contributes, not an adjudication of the ongoing debate on compositionality in conceptual representation. One test is based on the violation of the CHSH inequalities and the other is based on the Bell/CH inequalities. Both of these tests are examples of the more general joint distribution criterion (JDC) proposed by (Dzhafarov and Kujala, 2012). The JDC is decided by solving a linear programming problem of the form $\mathbf{M}\mathbf{Q} = \mathbf{P}, \mathbf{Q} \geq 0$. In the context of this article, the vector \mathbf{P} would comprise the sixteen probabilities depicted in (2) and \mathbf{Q} represents the global joint distribution $\Pr(\mathbf{A1}, \mathbf{A2}, \mathbf{B1}, \mathbf{B2})$. 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 failure of the linear programming solution is analogous to a violation of the Bell/CH or CHSH inequalities.

Both the Bell/CH inequalities and the CHSH inequalities require that marginal selectivity is satisfied. There has been some confusion about the role of marginal

selectivity when applying these formal mechanisms to cognitive phenomena (Dzhafarov and Kujala, 2014b). For example, Aerts et al. (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, meaning the combination is “entangled”, i.e., non-compositional. However, a subsequent analysis of the experiment showed that marginal selectivity does not hold (Dzhafarov and Kujala, 2014b). The associated formal analysis shows that it is inapplicable to employ the CHSH inequality, or the Bell/CH inequalities, when marginal selectivity does not hold. Therefore, marginal selectivity should be checked before the CHSH inequalities or the Bell/CH inequalities are applied, as has been followed in the empirical illustration presented in this paper. This is crucially different to the situation in quantum physics where marginal selectivity seemingly always holds.

A parsimonious approach to modelling a conceptual combinations entails that a *single* model can describe how it is being interpreted, namely a global joint probability distribution can be constructed from the four empirically collected pairwise joint distributions such that the empirically collected distributions can be recovered from the global distribution. Acacio De Barros (2012) labels this fact “contextuality” because the inability to construct the joint distribution over the four variables is equivalent to the inability to assign values to the four variables that is consistent with all the experimentally observed marginal distributions. This notion of contextuality provides some insight into non-compositionality as presented in this article. Intuitively, if the way the conceptual combination is being interpreted varies sufficiently across the different priming conditions it will not be possible to provide a global model of the interpretations which is consistent with how the the interpretations are behaving with respect to the marginal distributions. We contend that in such cases the combination is non-compositional and moreover provides an empiri-

cally testable dividing line between compositionality and non-compositionality. This view also holds that non-compositionality is a context-sensitive notion. [Dzhafarov and Kujala \(2014a\)](#) have extended this notion of contextuality to allow its determination within the presence of marginal selectivity not being satisfied. This constitutes an important theoretical development as it is possible that marginal selectivity is often not satisfied for conceptual combinations, and perhaps also for cognitive phenomena more broadly.

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, it is the Russian mathematician Vorob'ev who discovered results equivalent to that of Fine's theorem. As he was a contemporary of Kolmogorov, who axiomatized probability theory, Vorob'ev was apparently ignored ([Khrennikov, 2010](#)). Thus, it was quantum physics that became famous for demonstrating the impossibility of modeling 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 appear in cognitive psychology. The history just sketched, together with the fact that both the CHSH inequalities and Fine's theorem are based solely on conventional probability theory, opens the possibility to non-controversially apply them outside of quantum physics ([Aerts et al., 2000](#); [Khrennikov, 2010](#); [Aerts and Sozzo, 2011](#)).

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 lit-

erature. Despite this state of confusion, few analytical approaches have been proposed that are capable of demarcating the difference between the two forms of behavior. 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 analyses will be necessary in order to provide a full account of the semantics of language.

The semantics of concepts were modeled 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” vs. “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 modeled 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/CH inequalities or the CHSH inequalities are not violated, then the semantics of the conceptual

combination cannot be modeled 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/CH inequalities or the CHSH inequalities *are* violated, then the semantics of the conceptual combination cannot be modeled in a four way joint probability distribution. Such conceptual combinations are “non-compositional”.

This result could have a marked impact in modeling 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 modeled within a given probability space that the modeler 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 shows quantum theory is a fruitful source of new theoretical insights and tools for modeling conceptual semantics as it has already provided for other areas of cognition (Bruza et al., 2009; Aerts, 2009; Khrennikov, 2010; Busemeyer et al., 2011; Busemeyer and Bruza, 2012).

Acknowledgements

This project was supported in part by the Australian Research Council Discovery grants DP0773341 and DP1094974, and by the U.K. Engineering and Physical Sciences Research Council, grant number: EP/F014708/2. Welcome support was also provided by the Marie Curie International Research Staff Exchange Scheme: Project 247590, “QONTEXT - Quantum Contextual Information Access and Retrieval”). We thank Ehtibar Dzhafarov and Jerome Busemeyer for informative discussions. Thanks also to Dr. Mark Chappell (Griffith University) for his assistance in running the experiments.

References

- Acacio De Barros, J., 2012. Joint probabilities and quantum cognition. In: Khrennikov, A., Atmanspacher, H., Migdal, A., Polyakov, S. (Eds.), *Proceedings of Quantum Theory: Reconsiderations of Foundations*. Vol. 6. AIP Conference Proceedings, pp. 98–107.
- 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., Gabora, L., Sozzo, S., 2013. Concepts and their dynamics: A quantum-theoretic modeling of human thought. *Topics in Cognitive Science* 5 (4), 737–772.
- Aerts, D., Sozzo, S., 2011. Quantum structure in cognition: Why and how concepts are entangled, arXiv:1104.1322v1.
- Bruza, P., Busemeyer, J., Gabora, L., 2009. Introduction to the special issue on quantum cognition. *Journal of Mathematical Psychology* 53.
- 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, E., Kujala, J., 2014a. Generalizing Bell-type and Leggett-Garg-type inequalities to systems with signaling. arXiv:1407.2886.
- Dzhafarov, E., Kujala, J., 2014b. On selective influences, marginal selectivity, and Bell/CHSH inequalities. *Topics in Cognitive Science* 6 (1), 121–128.
- Dzhafarov, R., Kujala, J., 2012. Selectivity in probabilistic causality: Where psychology runs into quantum physics. *Journal of Mathematical Psychology* 56 (1), 54–63.
- 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.
- Gagne, C., Shoben, E., 1997. Influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology* 23, 71–87.
- Gärdenfors, P., 2000. *Conceptual Spaces: The Geometry of Thought*. MIT Press.
- Hampton, J., 1997. Conceptual combination. In: Lamberts, K., Shank, D. (Eds.), *Knowledge, concepts, and categories*. MIT Press, pp. 133–160.
- Isham, C., 1995. *Lectures on quantum theory*. Imperial College Press.

- Khrennikov, A., 2010. Ubiquitous quantum structure: From Psychology to Finance. Springer.
- Laloë, F., 2001. Do we really understand quantum mechanics? Strange correlations, paradoxes, and theorems. *American Journal of Physics* 69 (6), 655–701.
- 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., 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.
- 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.
- 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.
- 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.