Embodied-linguistic conceptual representations during metaphor processing

Pei Q. Liu

MSc

This thesis is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dec 2018

Department of Psychology

Lancaster University

Declaration

The thesis contains original work completed solely by the author under the supervision of Dr Louise Connell and Dr Dermot Lynott, and has not been submitted in the same form for the award of a higher degree at this institution or elsewhere.

This research was supported by a teaching studentship from Lancaster University Psychology Department.

Name: Pei Q Liu

Signature: LiuPei 刘沛

Date: 22/06/2018

Acknowledgements

As I consider these four years, I am filled with gratitude. Although this thesis is a result of many writing sessions at 6am, many error messages from R, and many episodes of the imposter syndrome, I realise that the most crucial things of my PhD are the things that I have not worked for.

My thanks first go to my supervisors, Dr Louise Connell and Dr Dermot Lynott. Thank you for putting your faith in me when I applied for the PhD. During these four years, you have motivated me with your uncompromising standard and your attention to detail. Meanwhile, you have been most generous in your support in different areas of my life. More importantly, I have not met a couple as devoted and compatible as you two. I admire you for being partners in your lives as well as your work. This PhD would not have been so enjoyable without your trust, consideration, and respect for each other. In so many ways, I consider myself fortunate to have you as my mentors and patterns.

I am also grateful to the Department. I do not need to exaggerate to say that the Psychology Department in Lancaster is the warmest department that I have ever known. It is because all the staff and PhD students are committed to create a supportive and inclusive environment. Special thanks to Szilvie, Han, Christian and Francesca for sacrificing their time to help me with my EEG studies; to Rob for being the statistics guru for the whole department; to Barrie and Simon for the ongoing technical support; to James B who has looked out for me from the first day I came to the PhD coffee and cakes reception in C62. There are still numerous people that I have not mentioned by name. Thank you for every kind smile, every encouraging word and every get-together.

Thanks and love to Christiana and John, who have made the office loving and fun. I am so glad we have stuck around from the first day until the end.

My profound gratefulness to my mom. She did her PhD in linguistics when I was a teenager, and I spent a lot of time with her fellow students. She allowed me to participate in their conversations, from which I apprehended the charm and exhiliaration of research. Those years with her and her fellow students planted the seed of academia in my heart, and from then on, she has never failed to encourage and guide me in my pursuit.

To Eric Ke Qiao: You have been with me in every step, always being the first pilot of my experiments and the first audience of my talks. You have comforted and supported me, and have given me more joy than I can ever imagine. Marrying you is the greatest achievement of my PhD.

Most of all, to my Lord and Saviour, to whom I owe it all: may You alone be my exceedingly great reward.

I am your shield and your exceedingly great reward.

– Genesis 15:1

Abstract

Although metaphor processing has fascinated linguists and psychologists alike, the conceptual representations involved have not been fully examined. In the present thesis, I propose metaphor processing should be viewed as an aspect of language processing, involving conceptual representations that are both embodied and linguistic. The thesis includes five self-contained papers, which showed a detailed picture of conceptual representation that was flexible and dynamic.

In the paper contained in Chapter 3, I proposed an operational definition of the effort to generate embodied simulation (i.e., the ease-of-simulation measure, or EoS). As a composite measure, EoS accounted for the speed of successful metaphor processing better than other rating tasks, which suggested that EoS could account for the underlying mechanism of metaphor processing, thus assumed to be embodied simulation. In papers reported in Chapters 4, 5 and 7, I studied influences of embodied simulation and linguistic distributional patterns on metaphor processing. These two components were both found to contribute to metaphor processing, and the interplay between them were were influenced by factors such as the depth of processing required and the time available for responses. Papers reported in Chapter 6 and 7 examined the EEG activations of embodied and linguistic components, in literal language processing and metaphor processing respectively. Both studies revealed that embodied and linguistic components performed various functions, each being activated at several time points. The linguistic component was activated first between 200ms-400ms after the stimulus onset, suggesting that it was involved in lexical and sublexical processing, which also supported the idea that it had a speed advantage compared to the embodied component. The latter was activated around 400ms, being responsible for semantic representations. Moreover, both components were activated again at the later stage of processing, indicating that both components were used and integrated for decision making.

Word count: 57320 words

Statement of Authorship

Here is a breakdown of contribution made by Pei Liu (the student), and Louise Connell and Dermot lynott (the supervisors) to each chapter. The order in which the names appear roughly indicates the proportion of contribution in declining order unless joint contribution is specified.

Chapters 1 and 2 (literature reviews)

- Principle author: Pei Liu
- Comments: Louise Connell and Dermot Lynott

Chapter 3 (submitted paper)

- Conception and creation of study: Louise Connell, Pei Liu, and Dermot Lynott
- Data collection: Pei Liu supervised by Louise Connell
- Writing manuscript: Pei Liu
- Revision: Pei Liu and Louise Connell
- Comments: Dermot Lynott

Chapter 4 (submitted paper)

- Conception and creation of study: Louise Connell, Pei Liu, and Dermot Lynott
- Data collection: Pei Liu supervised by Louise Connell
- Writing manuscript: Pei Liu
- Revision: Pei Liu and Louise Connell
- Comments: Dermot Lynott

Chapter 5 (empirical paper in preparation)

- Conception and creation of study: Pei Liu and Louise Connell (jointly)
- Data collection: Pei Liu supervised by Dermot Lynott
- Writing manuscript: Pei Liu
- Revision: Pei Liu and Louise Connell
- Comments: Dermot Lynott

Chapters 6 and 7 (empirical papers in preparation)

- Conception and creation of study: Pei Liu, Louise Connell, and Dermot Lynott
- Data collection: Pei Liu
- Writing manuscript: Pei Liu
- Revision: Pei Liu
- Comments: Louise Connell and Dermot Lynott

Chapters 8 (general discussion)

- Principle author: Pei Liu
- Comments: Louise Connell and Dermot Lynott

Supervisor's name: Louise Connell Signature:

Supervisor's name: Dermot Lynott Signature:

Date: 22/06/2018

Contents

CHAPTER 1

A New Look on Metaphor Processing

1.1 Introduction

Metaphor is a use of language where a term or a phrase is applied to an object to which the term or phrase cannot be applied literally. Various parts of speech can all be used metaphorically, including nouns (e.g., My surgeon is a *butcher*), verbs (e.g., Susan *flew* down the street on her bike), or adjectives (e.g., *bright* future). Researchers have found that metaphors are pervasive in language, appearing more often than noticed (Coulson & Van Petten, 2002). Furthermore, comprehension of metaphors can be automatic and irrepressible (Glucksberg, Gildea, & Bookin, 1982). In other words, metaphors do not seem to register as an artifice or require extra effort to process. Instead, they are an indispensable part of our daily language. If that is the case, how do people comprehend a metaphor? What gives rise to the metaphoric meaning?

Research on metaphor comprehension has been carried out in various disciplines such as pragmatics, cognitive science and psycholinguistics. In this review, the focus is on the theoretical and empirical work of psycholinguists and cognitive scientists. Through their effort, many factors have been identified as affecting metaphor processing (e.g., conventionality, familiarity or aptness); however, as I will argue later, the definitions of these factors have been inconsistent and ambiguous, which limits their utility in further understanding metaphor processing. Therefore, the goal of my thesis is to depart from this traditional approach and take a new perspective, which treats language processing as a process which relies on grounded conceptual representations. I will further argue that metaphor processing is an aspect of language processing, so it should rely on grounded conceptual representations in the same way. While this thesis will focus on metaphor processing, its findings should have wider implication on our understanding of language processing in general.

In the present thesis, I will use adjectival/predicative metaphors in the empirical component. Before proceeding, I will briefly clarify the issue of terminology. Various terms have been used to refer to the different components in a predicative or nominal metaphor, because of the wide range of research rooted in different disciplines. In a predicative metaphor "A is B", A is the subject and B is the object/predicate judging from the syntactic component they play. They could also be called the topic (or tenor) and the vehicle of a metaphor for linguists and pragmatists. For cognitive scientists, A is referred to as the target and B the source. Properties of the source domain (B) are attributed to the target domain (A). For the sake of consistency, I will use target and source to refer to the components in this chapter and throughout the thesis.

1.1.1 What makes metaphors easy to understand?

Researchers in the cognitive and psycholinguistic field set out to answer the question, what makes some metaphors easier to understand than others. Several factors have emerged throughout the years, such as salience, conventionality and aptness. They were found to affect the speed of processing, and were even suggested to change the mechanism of comprehension.

1.1.1.1 Salience

Salience has been proposed to capture the priority a meaning enjoys in language processing. Giora, in her Graded Salience Hypothesis (GSH), suggested that salience could be a result of "conventionality, frequency, familiarity or prototypicality" (Giora, 1997, 2002). Meanings that enjoy high salience (e.g., *bright student* or *bright daylight*) are activated prior to those with low salience (e.g., *bright solution* or *bright cellar*) regardless of its metaphoricity. Salient meanings were shown to be activated faster and with less difficulty in both behavioural (Giora & Fein, 1999) and electrophysiological studies (Coulson & Van Petten, 2002; De Grauwe, Swain, & Holcomb, 2010; Lai, Curran, & Menn, 2009; Laurent, Denhières, Passerieux, Iakimova, & Hardy-Baylé, 2006; Pynte, Besson, Robichon, & Poli, 1996).

GSH has also been joined with the fine-coarse coding theory, suggesting that the left hemisphere is specialised in rapid interpretation and activation of closely linked meanings while the right hemisphere is prone to perform coarser computations and maintain broader meanings (Jung-Beeman, 2005). The joint hypothesis suggests that salient meanings should activate the left hemisphere more whereas non-salient meanings should activate the right hemisphere more (Mashal, Faust, & Hendler, 2005). When applied to metaphors, it suggests that only unfamiliar or novel metaphors will specifically involve the activation of the right hemisphere. Evidence for the lateralisation of metaphor processing has been drawn from patients with hemispheric damage (Giora, Zaidel, & Soroker, 2000), neural imaging (Mashal et al., 2005; Pobric, Mashal, Faust, & Lavidor, 2008) and the divided visual field paradigm (Faust & Mashal, 2007).

Although GSH and its joint hypothesis with the fine-coarse coding theory have been supported to some extent by empirical evidence, some studies have indicated that the left and right hemispheres are different in more ways than the sensitivity to salience. For example, the two hemispheres are different in their ability to represent the context (Coulson & Severens, 2007), which is important because salience can be altered by context (Giora & Fein, 1999). Thus, the lateralised sensitivity to salient/non-salient meanings may well reflect reliance on context. Furthermore, evidence has also emerged which opposes the joint hypothesis. Some studies have shown that non-salient meanings elicited bilateral activation (Davenport & Coulson, 2011) or even more left hemisphere activation (Mashal, Faust, Hendler, & Jung-Beeman, 2009), instead of right hemisphere activation. The problem may be due to the lack of power and control in the original studies. First, many of these studies were run on a small sample. Giora et al. (2000) study on hemispheric-damaged patients tested 27 right-hemisphere damaged patients, and 31 left-hemisphere damaged patients with only four metaphors; whereas Mashal et al. (2009) tested only 15 participants though with a larger stimulus set of 96 stimuli (including conventional and novel metaphors, as well as literal word pairs and semantically unrelated word pairs). Having such a small size increased the risk of false positive and poor precision in estimating effect sizes.

Later studies were indeed better powered (e.g., Faust & Mashal, 2007), but a further problem was that these studies all have neglected the random variability in the materials. Since GSH broke down the barrier between literal and metaphoric language, many studies included distinct figures of speech and linguistic devices. For instance, Giora and colleagues' (2000) study on hemispheric-damaged patients compared conventional metaphor processing with sarcasm processing, using completely different stimuli. Other studies on metaphor processing also used entirely different sets of words for novel metaphors, conventional metaphors and literal word pairs (Faust & Mashal, 2007; Mashal et al., 2005, 2009). Thus, the effect of salience was confounded with the poorly controlled novelty, figurativeness and discourse. The evidence for the joint hypothesis, or the lack thereof, may be due to factors other than salience.

A further and more crucial problem is salience was proposed to be the underlying essence (i.e., the priority of a meaning) which manifests itself in many factors such as conventionality, prototypicality, etc. As a result, the operational definition of high versus low salience is often ambiguous and lacks control. For instance, according to Mashal and colleagues, metaphors from poetry (e.g., conscience storm) can be regarded as non-salient while metaphors from daily life (e.g., transparent intention) as salient (Mashal & Faust, 2008; Mashal et al., 2005, 2009). Similarly, Giora and colleagues (2000) chose sarcasm (e.g., saying *a lawyer "did a great job"* after a failed court case) and conventional metaphors (e.g., *broken heart*) to define the contrast between high and low salience.

These attempts to test of role of salience lacked a clear metric against which degrees of salience can be measured. Therefore in later studies, the operational definition of salience was narrowed down to mainly conventionality or familiarity, which was quantified in a rating test (Arzouan, Goldstein, & Faust, 2007; Columbus et al., 2015; Davenport & Coulson, 2013; L. L. Jones & Estes, 2006). This move was detrimental to GSH because the concept of salience itself, the conglomerate of familiarity, conventionality, prototypicality and many other potentially crucial factors of metaphor processing, had to fade away from the focus of researchers and be replaced by more concrete and measurable factors. However, for the same reason, this move was beneficial to the research, because participants could devise a larger amount of materials that were better controlled in terms of metaphoricity, ambiguity and other confounding factors.

1.1.1.2 Conventionality

Conventionality does not only play an important role in GSH, but is also seen as a critical factor to determine the mechanism of processing, according to the career of metaphor theory (Bowdle & Gentner, 2005). This theory has two crucial components, one accounting for the mechanism of processing (i.e., structure mapping; (Gentner, 1983; Gentner & Wolff, 1997; Wolff & Gentner, 2011)), and the other accounting for the what happens after conventionalisation (i.e., categorisation; (Glucksberg, 1991; Glucksberg & Keysar, 1990)). Structure mapping delineates two steps of metaphor processing. First, people compare the conceptual structures of the source and target by aligning the predicates of the two concepts (as shown in Figure 1.1). After such structural alignment, the predicate unique to

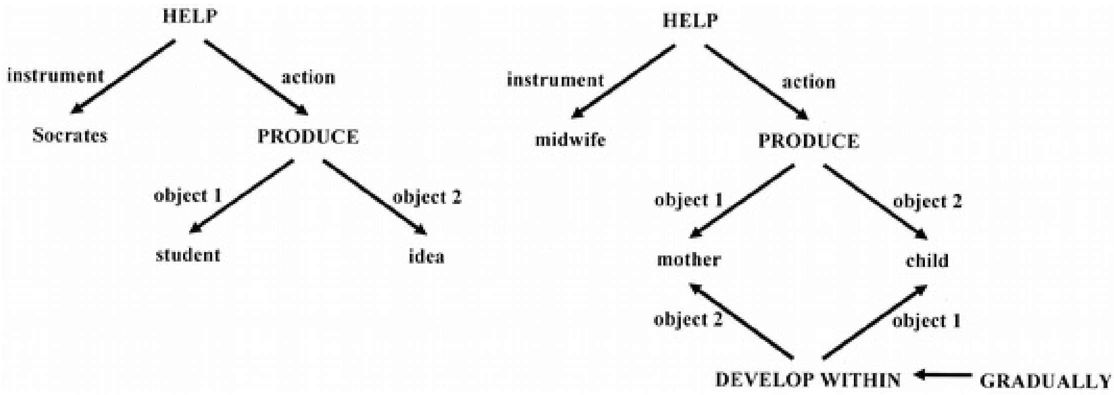


Figure 1.1: Structure of target and source domain in Socrates is a midwife (Bowdle & Gentner, 2005).

the source will be projected to the target, based on the specific dimension determined by the target. For instance, in the metaphor Socrates was a midwife (Figure 1.1), the source (*midwife*) has several properties, such as "helping MOTHER" and "MOTHER producing CHILD", while the target (*Socrates*) has properties such as "teaching STUDENTS" and "STUDENTS producing IDEAS". These properties are aligned based on their conceptual structures, but midwife also produces a property that is unique, that is, "MOTHER has CHILD developed within her", which is not obvious to the target. By projecting this property to the target domain, the relationship between student and idea can be deduced, that is: IDEA has been developing in STUDENT all along. Socrates did not simply teach an idea, but he helped students realise the idea that has been developing in the student.

Structure mapping is pivotal to the career of metaphor theory and related research on metaphor processing. Because of the specific roles of source and target according to the structure-mapping theory, researchers proposed new operationalised definition of conventionality (as well as aptness in the next section). Conventionality is accordingly defined as the association between the source and the property it produces (Bowdle & Gentner, 2005; Cardillo, Schmidt, Kranjec,

& Chatterjee, 2010; Chiappe, Kennedy, & Smykowski, 2003; L. L. Jones & Estes, 2005, 2006; Mashal, 2013; Pierce & Chiappe, 2008; Roncero & Almeida, 2014a, 2014b). When a metaphor such as *"Socrates is a midwife"* is novel, the association between the source (*midwife*) and the property (helping Object 1 produce Object 2) is weak. People cannot activate the property automatically when they encounter the sentence. Rather, they need to compare the conceptual structure of Socrates with the concept of midwife as we did above in order to arrive at this property. However, as the metaphor is encountered repeatedly, the property becomes more readily activated.

This process of repeated activation leads to the second component of the career of metaphor theory, that is metaphor processing based on categorisation. When a metaphor is conventionalised, the property that is readily activated becomes an abstract category of which the source word is a prototypical member. A conventional metaphor, in fact, is processed as what Glucksberg and colleagues called a class-inclusion assertion. For example, after the metaphor *Socrates is a midwife* is conventionalised, people comprehend it as a sentence which means that "Socrates is a member of the metaphoric category in which midwife is a prototypical exemplar (i.e., people or instruments that assist the deliverance of something that exists in the object)". In other words, understanding *Socrates is a midwife* effectively evokes the same mechanism which is used to understand *A robin is a bird.* The metaphoric category for which midwife represents becomes more refined. Imaging studies have supported this suggestion by showing that as a metaphor becomes conventionalised, the activation of irrelevant properties of the literal meanings are suppressed (Cardillo, Watson, Schmidt, Kranjec, &

Chatterjee, 2012; Desai, Binder, Conant, Mano, & Seidenberg, 2011; McGlone & Manfredi, 2001).

Another implication of the career of metaphor theory (especially metaphor processing by categorisation) is that other properties of class-inclusion assertions can be applied to conventional metaphors but no to novel metaphors. One of such properties is non-interchangeability between metaphors and similes. Simile is a figure of speech where one thing is likened to another (e.g., Figurative simile: *Socrates is like a midwife.* Literal simile: *Copper is like tin.*). It is processed solely by comparing the two concepts themselves, not the category of objects to which these concepts belong. In a literal simile such as *Copper is like tin*, it is particularly clear that both copper and tin refer to the substance of these two metals, which is why it is not appropriate to say *Copper is tin* (a class-inclusion assertion). However, it is acceptable to say *Robin is a bird* (a class-inclusion assertion) but unacceptable to say *Robin is like a bird* (a simile), because the latter is comparing the concept of robin with the concept of bird without assigning robin to the category of bird.

This distinction between literal class-inclusion assertions and similes should similarly exist for figurative language (i.e., conventional metaphors and figurative similes). If conventional metaphors are processed as class-inclusion assertions while novel metaphors are processed using comparison and structure mapping, it should be more appropriate to use the metaphoric form to express a conventional metaphor and the simile form to express a novel metaphor (Bowdle & Gentner, 2005; Lai & Curran, 2013). This prediction has been supported by studies which create a process of conventionalisation. For instance, participants were asked to read novel similes such as *An acrobat is like a butterfly* and *A figure skater is like a butterfly*. After such a process of conventionalisation, participants found that a similar figurative expression (e.g., *A ballerina is (like) a butterfly*) more acceptable in the metaphor form than in the simile form. In other words, conventionality indeed appeared to be a critical factor to determine whether a metaphor is processed using categorisation.

Problems with the career of metaphor theory are primarily rooted in its endorsement of the structure-mapping mapping. Since the structure-mapping theory suggests that it is the source term (e.g., *midwife*) that provides the property (e.g., facilitate the deliverance of a baby) which gets projected to the target term (e.g., *Socrates*), the conventionality of a metaphor is ultimately decided by the association between the source term and the property, regardless of the target term (Campbell & Raney, 2015; Cardillo et al., 2010; Katz, Paivio, Marshark, & Clark, 1988; Roncero & Almeida, 2014a). In other words, *Socrates is a midwife* is no more conventional than metaphors such as *Noam Chomsky is a midwife* or *Beyoncé is a midwife* (both of which is clearly not the case). Indeed, Thibodeau & Durgin (2011) have found that conventionality is dependent on the context, especially the immediate context of the target term.

Secondly, structure mapping cannot account for the existence of emergent properties (Gineste, Indurkhya, & Scart, 2000; Glucksberg & Haught, 2006; Haught, 2013). During metaphor comprehension, people often attribute properties to the target concept which do not originate from the source concept, but emerge as a result of the integration of the source domain properties with the target concept (for accounts of emergent properties see (Clement, Mawby, & Giles, 1994) cited in (Bowdle & Gentner, 2005; Gineste et al., 2000; D. Wilson & Carston, 2006)). For example, when reading *Some ideas are diamonds*, people often comes up with interpretations that ideas can be "creative" and "insightful", which are not applicable to the concept of diamonds. If metaphor comprehension is the process whereby the properties of the source domain are projected to the target domain, the resulting representation of the target should only include the properties that originally belong to the source (e.g., ideas are "rare" or "precious"). Therefore, simple structure mapping cannot account for all the mechanism of metaphor processing without further assumptions about how emergent properties arise. Aside from issues related to structure mapping as discussed above, the importance of conventionality is further questioned by the results from research on dual reference by Glucksberg and colleagues, which advocates for the importance of aptness in metaphor processing (see next section).

1.1.1.3 Aptness

In place of conventionality, many have argued that aptness (i.e., how well a metaphor expresses the speaker's intention) is the critical factor that decides whether metaphors are processed by categorisation or by comparison. Numerous studies have found that when aptness is taken into account, the effect of conventionality on metaphor processing (e.g., the preference for a metaphor over a simile) disappeared (Chiappe & Kennedy, 1999; Chiappe et al., 2003; Glucksberg & Haught, 2006; Haught, 2013; L. L. Jones & Estes, 2005, 2006; Pierce & Chiappe, 2008; Roncero & Almeida, 2014b).

Aptness has a high-level, qualitative definition which is how well or appropriately a metaphor can convey the speaker's intention (Blasko & Connine, 1993; Campbell & Raney, 2015; Katz et al., 1988). However, researchers have also operationalised it, in relation to structure mapping, as how well the properties specified by the source can capture important aspects of the target (Chiappe & Kennedy, 1999; Chiappe et al., 2003; L. L. Jones & Estes, 2005, 2006; Pierce & Chiappe, 2008; Roncero & Almeida, 2014b). Instead of stressing the association between the property and the source as conventionality (e.g., midwife as someone who assists the deliverance of something which has been developing in the object), aptness emphasises the how appropriately the property fits the target term (i.e., *Socrates*). Since a property needs to be projected to a specific dimension of the target, the better the property fits the target, the more apt the metaphor is deemed to be. An example often used to illustrate the point is the sentence *A rooster is an alarm clock* is more apt than *A robin is an alarm clock*, because the property of "reliably waking you up in the morning" fits *rooster* more appropriately than *robin*.

At least two issues could be raised concerning the latter treatment towards aptness. First of all, it is questionable whether this operational definition measures the same thing as the theoretical, high-level idea of aptness (i.e., how well a metaphor conveys the speaker's intention). Even when a source term can provide a property suitable for the target (e.g., *Deserts are ovens.* Property: hot), the metaphor is not necessarily rated as apt (aptness rating = 2.85 on a 7-point scale; see Roncero et al., 2014a). This could be because aptness is also correlated with other linguistic characters such as familiarity and conventionality (see next section). It is also not independent from cognitive factors such as working memory capacity (Pierce & Chiappe, 2008). People with a higher working memory capacity are better at interpreting metaphors and using the metaphoric form instead of the simile form. Furthermore, context has been found to be an important factor which influences the aptness rating (Thibodeau & Durgin, 2011; Thibodeau, Sikos, & Durgin, 2017). The operational definition of aptness (i.e., how well the property fits the target of a metaphor) only takes into account still a very narrow context. Although it is an improvement over conventionality to incorporate the target into metaphor processing, it still treats aptness as an intrinsic feature of a metaphor in an isolated phrase/sentence. However, simply from the example above, we can conceive of scenarios where *A robin is an alarm clock* is highly apt (e.g., in the woods. Robins are among the first birds to start the dawn chorus. https://www.rspb.org.uk/birds-and-wildlife/wildlife-guides/bird-a-z/robin/robin-redbreast-song/). In summary, although aptness was found to be a better predictor than conventionality, it could also be confounded by many external factors.

1.1.2 Challenges with psycholinguistic factors

In the previous section, I have introduced several factors that could facilitate metaphor processing (i.e., salience, conventionality and aptness). Apart from the criticisms to each factor individually, they could be further challenged by some mutual problems.

The primary challenge comes from the lack of a clear, unambiguous definition. For instance, familiarity and conventionality, two important factors of salience, are often used interchangeably to refer to the dichotomy, on one end of which is an expression or concept that is frequently encountered, and at the other end completely novel, even though only conventionality, not familiarity, is seen as the critical factor in the career of metaphor. Throughout the literature, familiarity/conventionality can be defined in two different ways, sometimes as familiarity towards the expression (Blasko & Connine, 1993; Bowdle & Gentner, 2005; Cardillo et al., 2012; Faust & Mashal, 2007; Glucksberg, 2003; Lai et al., 2009; Mashal, 2013; Mashal et al., 2005) and other times as familiarity to the concepts to which it refers (Bowdle & Gentner, 2005; Campbell & Raney, 2015; Cardillo et al., 2010; Gentner & Wolff, 1997; Giora, 1997; Jones & Estes, 2006; Peleg, Giora, & Fein, 2001). However, it is conceivable that such operational definitions could measure two different things. An expression could have high familiarity while the concept has low familiarity (e.g., hard Brexit; people may often encounter the term but hazy about what it means), and vice versa (e.g., warm reception; people may encounter the situation often but rarely use the term; for similar arguments, see (Andrews, Vigliocco, & Vinson, 2009; Brysbaert, Warriner, & Kuperman, 2014)).

Similarly, aptness faces the same problem of ambiguous definitions. As we discussed before, aptness could have two definitions, a high-level definition of metaphor suitability and a low-level definition of association between the target term and the property. Studying these factors largely relies on the collecting of subjective ratings in norming studies (Campbell & Raney, 2015; Cardillo et al., 2010; Katz et al., 1988; Roncero & Almeida, 2014a) or as part of a study where the norms are not published (Chiappe et al., 2003; L. L. Jones & Estes, 2006; Lai et al., 2009). Participants were often required to rate metaphors based on the operational definitions given. As such, ambiguity in the definitions of critical concepts could greatly limit the utility of these ratings in the research of metaphor processing.

Another challenge arises from the interdependency between these factors. While conventionality (as well as familiarity) measures the amount of exposure to a metaphor, aptness could conceivably be confounded by the amount of exposure as well. Only apt metaphors, those that can accurately convey the speaker's meaning, get used, and thus encountered, frequently. Conversely, previous usage of a metaphor could make it seem more apt (L. L. Jones & Estes, 2006; Pierce & Chiappe, 2008; Thibodeau & Durgin, 2011). Indeed, these subjective ratings are often highly correlated. Ratings of aptness and familiarity are highly correlated (r = .73-.98: Campbell & Raney, 2015; Katz et al., 1988; Roncero & Almeida, 2014a), as are ratings of aptness and corpus frequency counts of the metaphoric expression (r = .41-.57: Roncero & Almeida, 2014b; Thibodeau & Durgin, 2011). In a novel, large-scale study of existing ratings of nominal metaphors, Thibodeau and colleagues looked at existing norming studies on various scales including conventionality, familiarity, aptness, as well as imagibility, metaphoricity, etc. and found that these factors could be collapsed to two distinct principal components (PC). The first PC was processing fluency, referring to the ease of activating the metaphoric meaning which is largely subject to context and previous exposure. All the factors we have discussed above loaded heavily on this PC. This means that there is a core interdependency between these factors that is not too trivial to disentangle. Subjective ratings on these factors may simply be measuring the same underlying concept from different perspectives. Thus, it is futile to keep all these factors, or compare them in the same study because they are confounded and will cause the problems related multicolinearity, such as imprecise estimation of regression coefficients and overfitting.

In summary, it is problematic to study metaphor processing using these single factors. The ambiguity in their definitions and their core interdependency are detrimental to our understanding of metaphor processing. Therefore, in my thesis, I decide to leave these factors behind to seek a better theoretical construct to account for the online processing of metaphors.

1.2 Metaphor Processing from the Grounded Perspective

As the psycholinguistic research is troubled by many issues, in this thesis I propose a new perspective which examines metaphor processing from the perspective of grounded conceptual representation.

1.2.1 Grounded Views of Conceptual Representations

According to the grounded views, a mental representation draws on two types of information, that is information about linguistic distributional patterns, and information about embodied simulation (Barsalou, Santos, Simmons, & Wilson, 2008; Connell & Lynott, 2014b; Louwerse & Jeuniaux, 2008; Vigliocco, Meteyard, Andrews, & Kousta, 2009). Both types of information activate specific neural circuits in the brain during the acquisition of a new concept and leave traces for conceptual representations. For instance, to acquire the concept of *bright*, people encounter the linguistic label "bright" with its context as well as the sensory (mainly visual) properties of bright objects. The neural activity evoked by the linguistic label happens synchronously with the neural activity evoked by the sensory properties. Thus, the neurons responsive to each type of information are "wired together", which makes it faster to activate the circuits for representation next time when the concept is encountered (Hebb, 1949). During conceptual processing, these traces of neural activation get partially re-enacted, resulting in the simulation of sensory, motor, affective and other bodily experiences.

Linguistic distributional patterns and embodied simulation came from distinct origins and philosophical backgrounds, until recently when they were found to be two interacting and complementing components (see Andrews, Vigliocco, & Vinson, 2009 for a review). In this section, I will review evidence for these two components and their interaction in detail, and in the next section I will explore its application to metaphor processing and some challenges along the way.

Linguistic distributional patterns have been argued to be critical for mental representation of meaning (Landauer & Dumais, 1997). As Firth (1957) famously put it, which is often quoted by researchers, "You shall know a word by the company it keeps." We can acquire the meaning of a word, at least partially, by the context in which it appears. This approach of defining word meaning by the linguistic context is formalised in many ways, such as latent semantic analysis (LSA; Landauer & Dumais, 1997), hyperspace analogue of language (HAL; Lund & Burgess, 1996), and latent Dirichlet allocation (LDA; Blei, Ng, & Jordan, 2003). These methods create a high-dimensional space for each word based on large body of texts in which the word appears (e.g., encyclopaedia, fictions, etc.). Such a space, thus, can represent the meaning of a word using linguistic distributional patterns. These methods are powerful enough to capture many aspects of language processing, such as priming and reading times (Vigliocco et al., 2009), and the geographical positions of cities in China and the Middle East (Louwerse, Hutchinson, & Cai, 2012), or even Tolkien's Middle Earth (Louwerse & Benesh, 2012). Meanwhile, co-occurrence frequency within a small five-word window can also predict important features in the real world. Co-occurrence frequency between sensory adjectives could to some extent distinguish their perceptual modalities (Louwerse & Connell, 2011). This is because objects, places, events or concepts that usually appear in the same situation should often appear close proximity in context.

Although the linguistic distributional pattern is powerful, it cannot be all there is to conceptual representation. If the meaning of a linguistic symbol is only defined by other linguistic symbols around it, we would be trapped in an endless loop. This is the famous grounding problem: at some point, linguistic symbols need to refer to objects and concepts in the real world (Harnad, 1990; Searle, 1980). Embodied simulation is proposed to solve this problem. It suggests that conceptual representation requires the partial activation of motor, sensory, affective and other systems of bodily experiences (Barsalou, 1999; Connell & Lynott, 2016; Glenberg & Gallese, 2012).

Compelling evidence for such an embodied simulation comes from neural imaging studies. Words with sensory-motor properties were found to engage the same or similar distributed networks that respond to corresponding sensory-motor stimuli. For example, nouns with sound-related properties (e.g., *thunder*) selectively activated the auditory association cortex (Bonner & Grossman, 2012; R.

F. Goldberg, Perfetti, & Schneider, 2006; Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008); same with taste-related words (e.g., *salt*; Barrós-Loscertales et al., 2012) and colour words (e.g., *yellow*; R. F. Goldberg et al., 2006); action words referring to different parts of the body (e.g., face, arm or leg) would selectively activate the motor and premotor regions that corresponds with each body part (Boulenger et al., 2008; Desai, Conant, Binder, Park, & Seidenberg, 2013; Hauk, Johnsrude, & Pulvermüller, 2004; Speed, Dam, Hirath, Vigliocco, & Desai, 2017).

Embodied simulation has also been borne out by numerous behavioural effects, which collectively suggest that language comprehension requires modality-specific resources. Connell, Lynott and colleagues collected rating norms on the sensory strength of words, which quantified on a 7-point Likert scale how strongly a concept or property referred to by a word could be experienced by each of the five major sensory modalities (i.e., vision, audition, haptics, gustation and olfaction; Lynott & Connell, 2009, 2013). Not only could these norms account for many semantic effects better than traditional semantic variables such as frequency, concreteness and imagibility, but studies on the norms also demonstrated that conceptual representation during language comprehension co-opted modality-specific perceptual systems. Thus for example, when people performed a lexical decision task (i.e., reading and pressing buttons), because the task required and therefore pre-activated the visual system only, the visual strength of the referent concept facilitate the judgement of the word (i.e., faster to judge *bright* than *loud* or *warm*); but when they performed a word naming task (i.e., reading words out loud), because both visual and auditory attention was engaged, both visual and auditory strength of the referent concept facilitated the judgement of the word

(Connell & Lynott, 2014a).

Modality switching costs also supported the proposal that conceptual representation is embodied in modality-specific systems. In a perceptual task (Spence, Nicholls, & Driver, 2001), there is a delay in response to a stimulus in one perceptual modality (e.g., a beep) when the previous stimulus was in a different modality (e.g., a flash) because people need to redirect their attention from one perceptual modality to another. Similarly, people also slow down in a property verification task if they make a decision concerning a property of one modality (e.g., *HEELS – clicking*) if the previous trial has been of a different modality (e.g., *CANDLE – flickering*) compared to one of the same modality (e.g., *BLENDER – loud*; Connell & Lynott, 2011; Pecher, Zeelenberg, & Barsalou, 2003)). This cost could not be fully explained by the linguistic distributional pattern of the sensory properties (Louwerse & Connell, 2011). Furthermore, such costs occur when the switch happens between a perceptual stimulus in a modality (e.g., a beep) and verifying a perceptual word of a different modality (e.g., *CANDLE – flickering*; Van Dantzig, Pecher, Zeelenberg, & Barsalou, 2008); taxing the perceptual system (e.g., memorising visual shapes) impairs the property verification of the specific modality (e.g., visual properties; Vermeulen, Corneille, & Niedenthal, 2008). Thus, it was suggested that processing sensory or motor language requires specific representation of sensory and motor properties of a concept.

1.2.2 Interplay between linguistic distribution and embodied simulation

Although linguistic distribution and embodied simulation were studied independently, they should not be considered as mutually exclusive. Although as mentioned previously, linguistic distribution can capture important aspects of the real world, it is conceivable that linguistic distribution and embodied simulation do not entirely overlap for at least two reasons. First, some aspects of features of embodied experiences, which are often encountered by bodily perceptions, are less communicated through language (Brysbaert, Mandera, & Keuleers, 2018). The reason could be that these features are default and invariant to the concept. For example, the word *sun* does not co-occur with the word *round* (436 times per million times *sun* occurs according to the Web1T corpus) as often as the word *moon* does (918 times per million times moon occurs), even though the *sun* is always round but the *moon* is round visually only every 27 days. In this sense, the linguistic distributional pattern does not accurately depict the occurrence of sensorimotor properties in the real-world experiences, even though it can capture it (i.e., *round* can co-occur with both *sun* and *moon* because both the sun and the moon can be round). Rather, the linguistic information is capable of highlighting important aspect of the real-world experiences, some of which in fact occur relatively rarely.

The second reason why embodied and linguistic information may not completely overlap is that linguistic distributional pattern may encode unique information that is not directly experienced with sensory and motor systems. Here the information mainly refers to our knowledge of abstraction and conceptual relationship. Take the concept of FOOD as an example. Andrews et al. (2009) showed that the embodied information from speaker-generated feature norms encodes mainly sensory and motor information such as bodily sensations (e.g., *hunger*, *thirst*, *taste*) and actions (e.g., *swallow*, *ingest*, *enjoy*); whereas the linguistic information from large corpus mainly encodes taxonomically or functionally related, abstract concepts such as *drinks*, *lunch*, *diet*. Because of the different types of information encoded in these components, the combined model can explain language processing better than either alone (see also Johns & Jones, 2012; Louwerse & Connell, 2011). Therefore, in order to fully represent a concept, both components need to be utilised. In this thesis, I will study metaphor processing with both of these components in view. Specifically, my research will have the specific question in view regarding the function of the linguistic component. However, before proceeding to this specific question, two more questions concerning grounded representations need to be addressed.

1.2.3 Questions concerning grounded representations

1.2.3.1 Is embodied simulation epiphenomenal?

One challenge for embodied simulation suggests that the activation of modality-specific neural regions is merely epiphenomenal (Leshinskaya & Caramazza, 2016; Mahon, 2015; Mahon & Caramazza, 2008). In other words, concepts and sensorimotor processing belong to separate levels. Although there might be no denying that conceptual processing could involve neural regions specific to processing sensory and motor stimuli, there is another level of conceptual representation that is "amodal" and "symbolic" (not to be confused with the linguistic distributional patterns that deal with linguistic symbols). This conceptual level interacts with the perceptual and motor level, but it could stand alone for conceptual representation without activating the sensorimotor level. Thus, the recruitment of modality-specific regions during conceptual representation could be explained by the spreading of neural activation from the amodal conceptual cores to the sensorimotor regions. In short, the argument against the embodied view goes like this: according to the embodied view, conceptual representation equates sensorimotor processing. Therefore, any evidence for the activation of other neural regions during conceptual processing counts as evidence for higher-level, stand-alone, amodal representation.

This argument, first and foremost, is a straw-man. Only the most radical reductionists will propose to do away with conceptual representation altogether and equate conceptual representations sensorimotor processing. For most proponents of the embodied view, conceptual representation can be hierarchically organised, with convergence zones processing information from multiple modalities (Simmons & Barsalou, 2003) and linguistic distributional patterns as symbolic placeholders (Zwaan, 2016).

Another argument against the amodalist "attack" is that the latter part of the argument gives the amodal view an unfair advantage. It is a logical fallacy, a "black hole in conceptual space" as coined by (Barsalou, 2016), to suggest that anything that is not reducible to sensorimotor processing is by default amodal. Much evidence was found that conceptual representation activates multimodal neural regions and adjacent regions to the sensory and motor areas (e.g., Bonner & Grossman, 2012; Fernandino et al., 2016; Kiefer et al., 2008). Such findings could be evidence for the hierarchical organisation of embodied concepts. If conceptual representation is indeed symbolic or amodal, its interface with the sensorimotor areas should be arbitrarily organised (i.e., independent to where the modality-specific regions are). This is analogous to defining language as an arbitrary symbolic system, in that the linguistic symbols (e.g., *horse*) do not bear an analogous relationship with its referent (the four-legged farm animal with hooves and mane). It might be equally a logical fallacy to argue that anything short of such arbitrariness should be regarded embodied representation. The point is that it is the responsibility of the amodalists to define what amodal, symbolic concepts are, instead of the responsibility of the embodiists to rescue evidence for embodied simulation. Furthermore, the fact that it is difficult to draw a clear line between embodied and amodal concepts should be an argument in itself for abandoning the distinction between the separate levels of concepts and sensorimotor processing.

1.2.3.2 To what extent are linguistic symbols grounded?

As both components exist in conceptual representation, it is of interest to understand to what extent linguistic symbols are grounded. In simpler terms, if there exists a continuum of "groundedness", at one end of which conceptual representation is entirely symbolic and disembodied, while at the other end all linguistic symbols are fully grounded in sensorimotor processing, where does conceptual representation realistically stand, given that both embodied and linguistic components exist although some have proposed theories at the two extremes (e.g., a fully symbolic and abstract view by Kintsch & Dijk (1978), or radical embodied science by Chemero (2009))? Most researchers support a combinatorial view. For example, Harnad (1990) proposed that there is a group of “elementary symbols” which are grounded in sensorimotor experiences, while higher order symbols are related to sensorimotor experiences via these elementary symbols. If this proposal is true, people should be able to learn new symbols via the linguistic context and retain traces of sensorimotor experiences for embodied simulation during future conceptual representation.

To test this hypothesis, Dudschig, Kaup and colleagues asked participants to learn novel objects or words whose meanings were associated with the vertical space. When participants were learning the objects and the associated word labels, the objects were consistently presented in the upper half or the lower half of the visual space (Öttl, Dudschig, & Kaup, 2017); whereas when participants were learning the novel words only, the words (without any objects) were presented in pairs with either “up” or “down” or in a sentential context (Günther, Dudschig, & Kaup, 2017). They found that participants were able to judge explicitly whether the words’ meanings were associated with the upper or lower spaces after the learning phase, but only after they had learned the words through the visual space could the participants activate the sensorimotor experiences automatically later during an implicit judgement task (i.e., they were faster to associate up words with upward movements or visual space, and down words with downward movements or visual space). These findings made it explicit how important the mode of learning is to conceptual representations. Linguistic symbols can only be grounded in embodied simulation when they are learned through sensorimotor experiences but not via the association between linguistic symbols.

The implication of these findings is great. It seems to suggest a more limited role of embodiment, than researchers formerly think. Although research on embodied simulation has largely focused on concrete concepts, many researchers would agree that processing both concrete and abstract concepts relies on embodied simulation (Barsalou, 1999; Zwaan, 2014), not to mention that concrete concepts are defined as concepts whose referents have sensorimotor features (Paivio, 1986). These novel word acquisition studies, in contrast, suggest that embodied simulation is not a necessary condition of conceptual representation even for many concrete concepts, thus pushing our position on the groundedness continuum more towards the disembodied end. If a student in a land-locked country has only learned about shark from a textbook, without every seeing a picture or watching a TV show about sharks, they will not simulate the concept of a shark in the sea (its colour, its speed, etc.) via their knowledge about the category membership of shark and their available sensorimotor experiences with a gold fish in a tank.

Two counterarguments could be raised against such a conservative view of embodiment. First, it is unclear whether the short-term learning of novel words is comparable to the long-term acquisition of concepts in real life. In a learning phase as brief as in those studies, participants were not required and may not have enough time to engage in any deep processing of the sensorimotor features of the words "up" and "down" or the simulated representation of the linguistic context (more about the importance of deep processing in next section). This could mean that the traces of sensorimotor experiences had never been activated in the learning phase. However, when children build up their vocabulary, they are often required to engage in deep processing repeatedly in various reading and writing tasks. Therefore, people are more likely to engage in embodied simulation in real life than in those studies. Second, in real life, few concepts are acquired through linguistic association alone, totally devoid of bodily experiences. Few, if any, of us learned our first language by memorising a dictionary. Even an abstract concept is often accompanied by a large amount of useful sensorimotor information, and children are able to learn abstract concepts by generalisation over similar sensorimotor experiences (Colunga & Smith, 2003). For instance, we have learned the concept of *democracy* by second- or first-hand experiences of going to the polling booth (i.e., seeing other people vote or voting ourselves); so is the case with other words such as *mathematics*, *frustration*, etc. In a word, conceptual representation, from its acquisition to its processing, relies on the interaction between embodied and linguistic components. Therefore, it is high time to study this interaction, which will be the focus of my thesis.

There are many open questions concerning the content and structure of conceptual representation. In this thesis on metaphor processing, I will focus on two characteristics of the combined, simulation-linguistic conceptual representation. First, linguistic distribution and embodied simulation are activated at different speeds; and second, they play different roles based on the task demands.

1.2.3.3 Linguistic shortcut hypothesis

Linguistic distributional information gets activated faster than embodied simulation. Barsalou and colleagues in their language and situated simulation (LASS) theory suggested that the linguistic system becomes engaged first in a language processing task (Barsalou et al., 2008; Connell & Lynott, 2014b; Louwerse & Jeuniaux, 2008). Once a word is perceived (e.g., *food*), the neural activity spreads to adjacent and connected areas in the brain which encodes similar stimuli, that is, in the case of language processing, associated words based on linguistic distributional patterns (e.g., *drinks*, *lunch*, etc.). As this linguistic information is similar to the perceived word in memory, the spread of activation is rapid. Once the word is recognised, it will also activate simulation of embodied experiences as well (e.g., hunger, swallow, etc.). Such activation information encoded in the sensory and motor systems, is more distant, but it encodes more information. Thus, the linguistic component will reach the peak of its activity earlier than the simulation component. Evidence from behavioural studies and neuroimaging studies supports this hypothesis. For example, modality switching costs show that people respond to a sensory property (e.g., *bright*) slower if the previous property they encounter was of a different modality (e.g., *loud*). This effect could be accounted for by both embodied simulation (i.e., switch between sensory modalities) and linguistic distributional frequency (i.e., switch between different co-occurrence clusters), yet the latter explained fast response better while the former explained slow response better (Louwerse & Connell, 2011). In another study, when people were asked to generate a list of properties for a concept, properties bearing a linguistic association would be generated before properties originating from embodied simulation (Santos, Chaigneau, Simmons, & Barsalou, 2011). This temporal difference was also borne out by neuroimaging evidence which shows that early conceptual processing activates cortical regions specifically for language processing such as the left iFG, left superior temporal gyrus and the medial inferior frontal cortex; whereas late conceptual processing activates multiple cortical regions associated with processing perceptual stimuli such as the parietal cortex (Louwerse & Hutchinson, 2012; Santos et al., 2011; Simmons, Hamann, Harenski, Hu, & Barsalou, 2008).

Although it has been argued that the linguistic component reaches the peak of its activation rapidly, it does not imply that embodied simulation is slow or regarded ineffectual during early stage of conceptual processing. In fact, the motor cortex, for example, can be activated very quickly (within 200ms from word onset) when a word like kick is encountered (Hauk & Pulvermüller, 2004; Pulvermüller, Shtyrov, & Ilmoniemi, 2005). However, the important point here is that the role and importance of embodied and linguistic components could vary during the process of conceptual representation. Because the linguistic component is activated faster and is less precise, it has the potential to form good-enough representations (Ferreira, Bailey, & Ferraro, 2002), which can be taken advantage of when the context and task requirements allow (Connell & Lynott, 2014b; Louwerse & Hutchinson, 2012). In line with this argument, Connell & Lynott (2013) proposed the cognitive triage mechanism to account for the interaction of the two components in language processing. They suggested that linguistic distributional information could be a pointer for further processing before embodied simulation is fully engaged. If the linguistic component indicates that future processing is likely to fail (i.e., the words rarely co-occur in the same context and so their combined meaning might not be simulated successfully, e.g., *cactus beetle*), then it could abandon the processing before any more cognitive effort is expended by the embodied component. On the other hand, if the linguistic component indicates that future processing is likely to succeed (i.e., the words often co-occur in the same context and so their combined meaning can probably be simulated successfully, e.g., *military plan*), then it could either inform to accept the word combination immediately (i.e., based on the linguistic shortcut alone) or allow the embodied component to continue developing a detailed simulation of meaning.

Since whether or not people take the linguistic shortcut depends on the requirements of conceptual representation tasks, it is worthwhile to study which condition exactly encourages the use of the linguistic shortcut. One possible candidate is relatively shallow processing – people should be more likely to rely on the linguistic distributional information when it does not require a detailed account of the representation to complete the task. Indeed, many studies which show effects of simulation often required deep and deliberate processing such as producing interpretation for a sentence or judging specific aspect (e.g., location in space, relationship between the semantic representation of two concepts (Borghi, Caramelli, & Setti, 2005; Lebois, Wilson-Mendenhall, & Barsalou, 2015; see review Louwerse & Jeuniaux, 2010). Meanwhile, the linguistic component is more prominent when participants are engaged in relatively shallow processing. For example, Connell & Lynott (2013) applied the same set of noun-noun compounds to two tasks: a sensibility judgement task which required only a simple yes/no response, or an interpretation generation task which required participants to type down the interpretation to the compounds. Although the sensibility judgement task is often considered as a "deep" semantic task, it does not explicitly require an elaborate, verbal account of the conceptual representation, thus relatively shallower than the interpretation generation task. By introducing such a contrast, it was found that the linguistic component only had an effect on the performance of the shallow sensibility judgement task. In summary, shallow processing could be a condition that encourages the use of the linguistic shortcut, which will be a hypothesis tested throughout this thesis.

1.3 Grounded metaphor processing

1.3.1 Grounded language processing

Grounded conceptual representation has been studied for more than two decades, with a special focus on literal, concrete language. This aspect of language processing is easy to study because we can safely assume that processing a sensory word (e.g., *bright*) mainly requires the reactivation and "re-living" of sensory experiences denoted by the word (i.e., emitting much light; Zwaan, 2004). Researchers have found both behavioural and imaging evidence in favour of this view. Meanwhile, some researchers have tried to bridge the gap between general language processing and metaphor processing. Most of them have tried to use a specific theory, the conceptual metaphor theory, as a way to explore simulation during metaphor processing. However, such an exploration has not been successful empirically. In the next section, I will discuss the conceptual metaphor theory and related empirical research in detail, and I will argue why this path will eventually lead to a dead end.

Embodied simulation during literal language processing has been tested both behaviourally and with imaging techniques. People can represent the perceptual features of a concept described in a sentence/word (e.g., an eagle with spread wings after reading *The ranger saw the eagle in the sky*) and preferring a matching image faster to a mismatch image (Engelen, Bouwmeester, Bruin, & Zwaan, 2011; Lynott & Connell, 2009; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Zwaan & Pecher, 2012; Zwaan & Yaxley, 2003). Other behavioural studies have also found that people can represent the spatial feature of concepts in an iconic manner (e.g., preferring the word *attic* above the word *basement* to the reversed arrangement; Zwaan & Yaxley, 2003) and the direction of an action described in a verb (e.g., being faster at moving the arm towards the body after reading *open the drawer*; Glenberg & Kaschak, 2002). However, the iconicity effect could also be attributed to the linguistic distributional pattern (i.e., attic appears more often in context before basement than the reversed; Louwerse & Jeuniaux, 2008); and the action-sentence compatibility effect was found to be generally weak (Papesh, 2015), and related to the planning of motor response, instead of semantic processing or conceptual representation (Borreggine & Kaschak, 2006; T. Brouillet, Heurley, Martin, & Brouillet, 2010; T. M. C. Miller, Schmidt, Blankenburg, & Pulvermüller, 2018).

Greater support for grounded language comprehension is found through imaging studies. It was consistently supported that processing words or sentences whose meanings are related to specific sensory modality (e.g., *thunder*), action (e.g., *kick*) or affection (e.g., story about getting compliment) activated specific neural regions associated with processing such stimuli (R. F. Goldberg et al., 2006; Hauk & Pulvermüller, 2004; Raposo, Moss, Stamatakis, & Tyler, 2009). Moreover, the activations of modality specific regions were found beyond the processing of single words and sentences, and existed during comprehension of stories and discourse (Chow et al., 2014; Kurby & Zacks, 2013). Such activations were crucially influenced by the context, such that a coherent representation of the context produced the strongest activations in the modality specific regions; and modality specific activations of a single word can be overridden by the context. For example, the word *kick* activates the motor and premotor cortex to different degrees when it was embedded in different context. It elicited greater activation when read in isolation than in a literal phrase such as *kick the ball* (Raposo et al., 2009).

1.3.2 Conceptual metaphor theory

Studies on grounded metaphor processing have spawned from the studies above on grounded language comprehension. However, a particular interest has been paid on the modality-specific activations of the source domain properties, and here I will argue that this approach in fact limits our understanding of metaphor processing. The rationale behind this research is based on the theory of conceptual metaphor (Lakoff & Johnson, 1999). A conceptual metaphor establishes an analogous relationship between an abstract concept with the sensory-motor experiences. For example, KNOWING IS SEEING is a conceptual metaphor which draws an analogy between the conceptual behaviour of KNOWING with the physical behaviour of SEEING. These analogous relationships are argued to derive from a strong correlation between the abstract concept and sensory-motor experiences. SEEING a physical object entails that the object (at least its existence) is KNOWN. The conceptual metaphor theory as such has been used to account for the grounding of abstract concepts into sensori-motor experiences (Boroditsky, 2000; Hurtienne & Meschke, 2016; Lakoff & Johnsen, 2003). Because KNOWING

IS SEEING, when we say “I see your point”, motor experiences of seeing a physical object is activated, but the object to be seen and thus known is the intentioned meaning of “you”.

The conceptual metaphor theory was used to account for the simulation process activated during metaphor comprehension. As a conceptual metaphor establishes an analogous relationship between two concepts, people could use the conceptual structure, the sensory and motor properties, of one concept (i.e., the source domain) to understand the conceptual structure of another concept (i.e., the target domain), which is often more abstract (Gibbs, 2006; Gibbs, Costa Lima, & Francozo, 2004). For example, a linguistic metaphor such as “a bright student” is to draw an analogy between the sensorimotor experience of bright with the abstract concept of intellect. Thus, to understand “a bright student” is to ground the intellect of a student to the sensorimotor experience with something bright. According to our past experiences, if something is bright (e.g., the sun), we will be able to see things clearly because of it. As KNOWING IS SEEING according to the conceptual metaphor, a “bright” student who is able to see things clearly is a student whose intellect enables them to learn and know things quickly. In other words, people use linguistic metaphors in language in accordance with the conceptual metaphor in our mind, such that the related sensorimotor experiences can be simulated for us to comprehend the abstract concept in the linguistic metaphor.

Evidence for this kind of embodiment needs to show that metaphor comprehension activates the sensorimotor experiences of the source domain. To achieve this aim, the action compatibility effect (ACE) is studied (Glenberg & Kaschak, 2002). It was found that reading sentences with metaphorical actions (e.g., *rise to* *power*) facilitates the physical action in the same direction (e.g., vertical upward movement) but impedes actions in the opposite direction (Santana & Vega, 2011). This suggests that the motor properties of rising are activated when people process metaphoric actions even though the meaning of the metaphor does not include a physical vertical motion.

Although research on conceptual metaphor has offered many ingenious attempts to study the embodied simulation in metaphor processing, this approach is limited because it has constricted the definition of embodied simulation to the re-enactment of properties from the source domain. Although the behavioural effect supports the activation of such source domain properties, there is no direct evidence showing the activation of neural regions specific to the processing of perceptual and motor stimuli during metaphor processing. The behavioural effects described above, which has been interpreted as a result of embodied simulation, can be accounted for by established principles of perception, language and memory (Casasanto & Gijssels, 2015). When people read about a metaphoric action (e.g., *rise to power*), their response could be facilitated because the word rise can prime the upward response space. Therefore, even if there is the simulation of sensorimotor experiences of some sort, the evidence does not show that such a process is essential to metaphor comprehension.

Indeed, there has been little neuro-imaging evidence showing activations in exclusively sensorimotor regions when people engage in metaphor processing. Many fMRI studies have examined action verbs hoping to find activation in specific motor/pre-motor regions associated with the body parts that perform the action. However, metaphors with action verbs (e.g., *grasping the idea*; *My mind is run-*

ning) do not activate the motor and premotor cortices to the same extent as literal use of the same verbs (Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006; Raposo et al., 2009; Romero Lauro, Mattavelli, Papagno, & Tettamanti, 2013). Not only so, metaphors do not activate the primary cortices any more than their literal counterpart. For example, if *a rough day* had recruited the somatosensory cortex more than *a bad day*, or *a bright student* had recruited the primary visual cortex more than *a clever student*, we might have been able to argue that metaphor processing involves the embodied simulation of sensorimotor experiences. However, such evidence was not found. Comparison using fMRI between perceptual metaphors and their literal counterparts (e.g., *She had a rough day* vs *She had a bad day*) did not find greater activation in the primary sensorimotor cortices (Citron & Goldberg, 2014; Desai et al., 2013; Romero Lauro et al., 2013; Yang & Shu, 2016). Although they have found activations in other regions associated with processing sensorimotor stimuli, these regions are also activated in other cognitive and perceptual processes not related to the specific sensory and motor modalities (Casasanto & Gijssels, 2015).

These null effects from neuro-imaging studies actually should not be surprising to researchers on metaphor processing, because metaphors by definition should be detached from its source domain properties. After all, the definition of metaphor is the use of language where the metaphoric term does not refer to its literal meaning, whether this literal meaning pertains to the sensory, motor or other bodily experiences or abstract senses. Many studies have found that the literal meaning of the source domain is quickly suppressed by the context, even when it is activated at the initial stage of processing (Glucksberg & Haught, 2006; Mac-

Gregor, Bouwsema, & Klepousniotou, 2015; Mashal, 2013; Raposo et al., 2009). The word bright may activate the primary visual cortex at the beginning of the processing, pertaining to its literal meaning (i.e., emitting much light), but people should be able to suppress this activation in the context of bright student, favouring the metaphoric meaning (i.e., quick-witted and intelligent). In a metaphor that is conventional and apt, source domain properties should not have sustained activity which competes with the metaphoric meaning (Desai et al., 2013).

One possibility is that the activation of source-domain properties is brief. However, fMRI is not able to pick up such fast activations and suppression due to its low temporal resolution. Only one study, to my knowledge, has used MEG to study the timecourse of neural activation during metaphor processing (Boulenger, Shtyrov, & Pulvermüller, 2012). It indeed found that source domain activation only happened at the early time window (150-250ms). Although it lends support to the idea that metaphor processing activates some sensorimotor features of the source domain, it also suggests that such source domain activation is suppressed in metaphor comprehension, instead of activated to facilitate comprehension as the conceptual metaphor theory suggests.

1.4 The present thesis: metaphor processing involving simulation-linguistic conceptual representations

While pointing out the lack of evidence for embodied conceptual metaphors, it is not my intention to suggest that metaphor comprehension is not grounded in any past bodily experiences. It is unprofitable for researchers to be limited by the current definition of simulation based on the conceptual metaphor theory. Instead, we should align our view of metaphor processing with the current understanding of grounded language processing, that is to define the simulation involved in metaphor processing simply as the partial re-enactment of sensory, motor, introspective, affective and other experiences related to the concept. In other words, when we use the phrase “a bright student”, we do not need to picture a student with a light-bulb flashing above their head; but we should still reactivate our physical, emotional and psychological experiences with a quick-witted student in order to achieve comprehension.

To propose a new definition of simulation in metaphor processing is indeed the first challenge I will face in my thesis. Since simulation is complex and subconscious (Connell & Lynott, 2016), this definition needs to capture the effort for simulation as a whole, allowing us to remain agnostic about the specific experiences reactivated. Throughout my thesis, this definition will be scrutinised and validated in both behavioural studies and EEG studies. Furthermore, I will study how embodied simulation can be coupled with linguistic distributional patterns. Specifically, I will investigate the conditions under which each component is involved, and their interplay to create conceptual representations.

1.4.1 Thesis Outline

Thus, in my thesis, I will first propose an index of ease of simulation and generate a body of metaphoric sentences that vary independently on the measure of ease of simulation and the linguistic distributional frequency (Chapter 3). On the one hand, the ease of simulation measure will capture metaphor comprehension as a whole, measuring the effort required to arrive at a successful representation. On the other hand, linguistic distributional frequency will be defined in a minimal fashion, capturing the distributional patterns of metaphors' constituent words in close proximity, within a narrow five-word window (e.g., Louwerse & Connell, 2011). This treatment of linguistic distributional frequency will advocate strongly for the efficacy of the linguistic information in following chapters.

Then I will investigate whether metaphor processing relies on both embodied simulation and linguistic distributional information, as the grounded approach suggests in the following chapters (Chapters 4-7). If both components play a role in metaphor processing, then the ease of simulation and linguistic distributional frequency should each contribute to the performance of metaphor processing with a unique portion. Furthermore in Chapter 4, I will test the hypothesis that the linguistic component can act as a shortcut when the task requires only shallow processing. Thus, I will contrast shallow and deep processing with two sentence processing tasks as Connell & Lynott (2013): that is, a shallower sensibility judgement task and a deeper interpretation generation task.

Studies in the following chapters (Chapter 5 and 7) will develop upon the same paradigm as the study in Chapter 4, in order to further examine the interplay between ease of simulation and linguistic distributional frequency. Chapter 5 will continue to test the linguistic shortcut hypothesis and investigate the role linguistic distributional frequency plays under time constraint. If people use the distributional pattern of words as a shortcut to make judgments, the effect should be larger when the time resource is limited. Accordingly, the effect of embodied simulation will be smaller when the time resource is limited because to engage in embodied simulation is time-consuming.

Then in Chapters 6 and 7, I will search for the neurophysiological manifestations of the embodied and linguistic components using electroencephalography (EEG). Chapter 6 will identify these two components in literal language processing, and Chapter 7 will match them with metaphor processing. Using EEG will offer us greater insights into the online processing during metaphor comprehension with high temporal resolution. However, the conventional practice among the EEG literature is problematic and prone to Type I error, for reasons discussed in the next chapter (Chapter 2). Therefore, apart from studying metaphor processing, another mission of my research in these chapters will be to develop a more rigorous way to analyse EEG data for psycholinguistic research.

Finally in the conclusions chapter (Chapter 8), I will summarise the theoretical and empirical advances in this thesis, specifically in relations to the linguistic shortcut hypothesis. I will discuss, first, how well current theories of grounded conceptual representations can be applied to the research on metaphor processing. Second, I will also discuss the implications of my thesis on grounded language processing. Since starting point of my thesis is to cast metaphor processing under the big framework of language processing. My findings will increase our understanding of conceptual representations in general and point out a direction for future studies on grounded language processing.

CHAPTER 2

Using EEG to Study Metaphor Processing

2.1 Introduction

In my thesis, as discussed in the previous chapter, I will study the conceptual representation during metaphor with the specific question regarding the role of linguistic distributional patterns in mind. To test the linguistic shortcut hypothesis, it is important to validate a crucial condition for the linguistic component to become a shortcut: that is the linguistic component reaches the peak of its activation before the embodied component. In other words, it is necessary for the linguistic component to have a temporal advantage over the embodied component (though both components are activated simultaneously) for the linguistic component to become a valid guide for the processing of embodied simulation. Therefore, a crucial component of my research is the use of electroencephalography (EEG).

In order to study the grounded representation with EEG, it is important to identify and differentiate the effect of the embodied and linguistic components on the EEG waveforms. In previous research, the study of EEG waveforms primarily relies on the identification of event-related potential (ERP) components, which are segments of EEG waveforms circumscribed to a period of time and an area on the scalp. Several ERP components have been found to be associated with semantic processing and even embodied conceptual representations (e.g., N400 and late positivity complex). However, as well be discussed in this review, the study of ERP components suffers from great problems, such as high researcher degrees of freedom and the use of ANOVA. Therefore, it is a mission of my research to seek a more rigorous way to study EEG, by limiting researcher degrees of freedom and using better statistical tools (i.e., linear mixed-effect model plus Bayes factor). In this chapter, I will first review available findings of the ERP markers of conceptual representation and argue why the current way to study EEG lacks the needed rigour. Finally, I will propose a better way to analyse EEG data, which will be used in studies of Chapters 6 and 7 in my thesis.

2.2 EEG in psycholinguistics

EEG is the electrical neural activity measured at the scalp over time. It is recorded at various sites on the scalp, as the summation of all the electrical fields generated by the neural activities in the brain (Coulson, 2007). What is of interest to psycholinguistic research is the event-related potential (ERP), an average of several EEG waveforms during or after similar events, such as the presentation of a word. ERP provides different parameters that could be informative to the researchers, such as the amplitude, the latency and the scalp topography (Figure 2.1). These parameters are used to make functional inferences about the processing in the brain (Otten & Rugg, 2005). For example, if participants are presented with concrete words (e.g., *weapon*) and abstract words (e.g., *fashion*) matched on other variables while their ERP is recorded, the differences in ERPs can be interpreted as the difference in concreteness (Barber, Otten, Kousta, & Vigliocco, 2013). If the amplitude to concrete words is larger than abstract words, we may infer that this change in amplitude reflects retrieval of sensorimotor features. If concrete and abstract words produces different topographies, we can also infer that the conceptual representation of these words relies on different neural population. It is now even possible to infer with some accuracy the neural activities from the scalp topography whereby knowing which area(s) in the brain is responsible for the processing (Koles, 1998; Slotnick, 2005), though it is not common practice among psycholinguists. Rather, the greatest advantage of EEG is its high temporal resolution. It could capture the online, immediate electrical activity on the scalp to the precision of several milliseconds.

2.2.1 The N400 component

The most common way to characterise ERP waveforms is to capture them in ERP "components", which are segments of a waveform within a period of time "with a circumscribed scalp distribution and a circumscribed relationship to experimental variables" (Otten & Rugg, 2005). Many ERP components have been identified since EEG was first applied to psychology research, among which most relevant to my thesis is the N400 component. N400 is a segment of relative negative going waveform around 400ms after the onset of a stimulus, commonly found in the centro-parietal sites (Kutas & Federmeier, 2011).

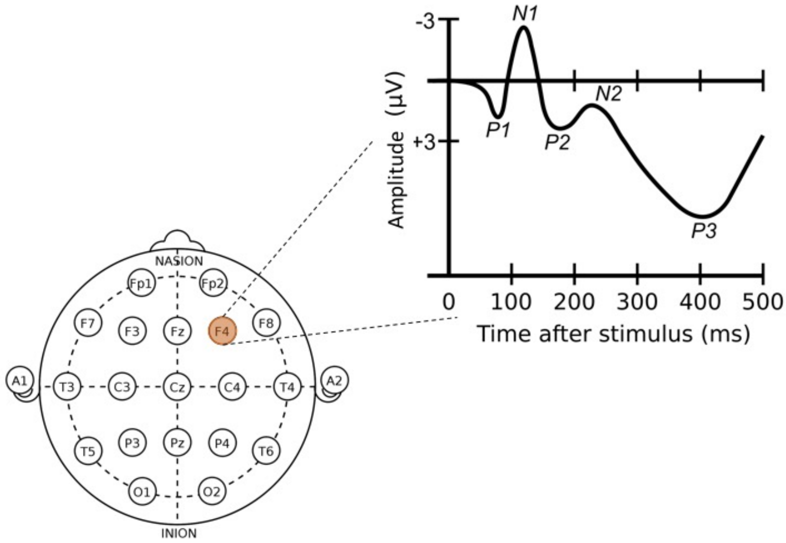


Figure 2.1: A schematic drawing of ERP from the right frontal electrode (F4). The crucial parameters of ERP are amplitude: the magnitude of the waveform in microVolts from 0 (i.e., the vertical axis of the waveform on the right); latency: the time between the stimulus onset (0 milliseconds) and the peak or start of a component, e.g. N1 (i.e., the horizontal axis); and scalp topography: amplitude of the whole scalp at a certain point of time. Pictures from Wikipedia.com.

N400 component was initially found to be a marker of semantic anomaly, such as *He took a sip from the transmitter* comparing to *He took a sip from the fountain* (Kutas & Hillyard, 1980). Later it was found that in sentence processing, the best predictor of the N400 effect was the cloze probability, that is the probability of using a certain word in a cloze test (Kutas & Hillyard, 1984). For example, in a sentence *He takes sugar and lemon with his* . . . , it is more probable to finish the sentence with *tea* than *coffee.* Even though both cases are semantically correct, the amplitude of N400 to tea was found more negative than coffee. Later N400 was also found to be sensitive to other lexical factors such as abstractness (Huang, Lee, & Federmeier, 2010) and even non-linguistic stimuli such as unexpected shapes (Ganis, Kutas, & Sereno, 1996), faces (Olivares, Iglesias, & Antonieta Bobes, 1999) and odours (Olofsson et al., 2014). Because of the wide range of stimuli that can elicit an N400 effect, Kutas & Federmeier (2011) argued that N400 reflects the effort to integrate new stimuli into a multimodal, dynamic representation of concepts, which is shaped by both recent events and long-term memory (see also Baggio & Hagoort, 2011; Rabovsky & McRae, 2014).

N400 component has also been used to study metaphor processing. Metaphors (e.g., *Every second of our time was attacked*) were found to elicit an N400 effect compared to literal sentences (e.g., *Every soldier in the frontline was attacked*), which nevertheless was not as large as the effect of an anomalous sentence (e.g., *Every drop of rain was attacked*; Arzouan et al., 2007; Forgacs, Bardolph, Amsel, DeLong, & Kutas, 2015; Lai et al., 2009). The more familiar or easily interpretable was a metaphor, the smaller was the effect (i.e., *Every point of my argument was attacked* elicited a smaller N400 effect than Every second of our time was attacked). Furthermore, the N400 effect of metaphors can be attenuated by the context (Coulson & Van Petten, 2002; Lai & Curran, 2013). For example, people read prime-target pairs of sentences with the same final word (e.g., *bumpy*) in a literal condition versus a metaphor condition. In the metaphor condition, people read the prime sentence (e.g., *I can see the path of his life*) followed by the target *Life can sometimes be bumpy.* In contrast in the literal condition, people read the prime *I can see the path of this road* followed by *Roads can sometimes be bumpy.* The N400 effect between metaphors and literal sentences to the final word (*bumpy*) would diminish. Both the effect of conventionality and the effect of context show that the harder it is to integrate new stimuli into the conceptual representation, the larger the N400 effect is, consistent with Kutas & Federmeier (2011).

The N400 effect is, moreover, used to study embodied simulation. As Kutas & Federmeier (2011) suggested, the conceptual representation draws from perception, memory and language and is instantly shaped by incoming events. Therefore, modality switching costs, which show that verification of sensory properties recruit modality-specific resources, should also elicit the N400 effect. Modality switching costs refer to the delay in response time to verify a property of one perceptual modality (*bright*) if the previous trial is of a different modality (*loud*; Pecher et al., 2003). Several studies have looked for the ERP manifestations of modality switching costs (Bernabeu, Willems, & Louwerse, 2017; Collins, Pecher, Zeelenberg, & Coulson, 2011; Hald, Hocking, Vernon, Marshall, & Garnham, 2013; Hald, Marshall, Janssen, & Garnham, 2011). However, what they found was not a clear picture. Not only was a significant effect found in the time window for the N400 component, but also a so-called early N400-like effect around 300ms (Bernabeu et al., 2017; Hald et al., 2011), the N1-P2 complex around 200ms (Bernabeu et al., 2017; Hald et al., 2013, 2011), as well as the late positivity component (LPC) after 600ms (Bernabeu et al., 2017; Collins et al., 2011; Hald et al., 2011).

Finding a significant effect at a time window outside the original hypothesis is of course not inherently bad or wrong. In fact, it is reasonable since EEG is a much finer measure than response time. What is manifested as a delay in response could be a result of many differences in semantic processing (Hauk, 2016). However, the finding of these components reveals a fundamental problem with the conventional method of analysis which suffers from high researcher degrees of freedom, plus an overlook of the nature of an ERP waveform. In the next section, I will discuss these problem in detail using the research on modality switching costs as an example. However, it should not be taken as a problem that exists only in these few studies, but reflecting the general practice across the EEG literature. To understand the problem, it is necessary to critically evaluate our current understanding of ERP components.

2.3 Critical evaluation of ERP components

The problem with characterizing EEG with ERP components are two-fold. First, there is a discrepancy between what we generally call ERP components and what is supposed to represent: latent components. Second, ERP records continuous and clustered data (e.g., many data points from one participant), whereas ERP components are discrete and aggregated (i.e., averaging over many trials for one participant). These problems keep us from appropriately analysing and interpreting ERP results.

The first problem with ERP components concerns a matter of proper understanding. There is a gap between what we regard as ERP components and what we expect it to do. What we call an ERP components generally refers to is a segment of the observed ERP waveform. The reason to define ERP components is because we expect an ERP component to reflect the activity of a specific process in the brain. In the brain at a certain window of time, a cluster of the neurons (either in a specific area of the brain or in a network of different areas) is engaged in a specific process, while others are engaged in other functions either related to or unrelated to the processing of the incoming stimuli. For example, during a property verification task, a certain neural cluster is engaged in semantic processing, specifically with the recruitment of modality-specific resources. The activity of this neural cluster will generate some electrical activities, which is called a latent ERP component. The latent component is directly associated with the neural cluster, and thus functionally significant.

However, the latent component is not the ERP component we know, because the ERP component obtained from the observed waveform measures the electrical activities generated by neurons in the whole brain. In other words, the observed component is a sum of all the latent components generated by different functionally significant neural clusters in the brain. In the ideal world, there would have been methods to reliably decompose the observed ERP waveform into latent components, but the reality is each observed waveform can be decomposed to several, if not infinite, latent components, each generated by one functionally devoted neural cluster (Figure 2.2; Luck, 2005).

As shown in Figure 2.2 (Panels A-C), an observed waveform (Panel A) can be decomposed to three latent components in two ways (Panels B and C). Both Panels D and Panel F display an increase in the amplitude of Peak 3 (from solid to dotted line), but the increase actually results from different changes of the latent component. Panel D displays the effect of decreasing the amplitude of latent component C2' (in Panel C), whereas Panel F displays the effect of increasing the amplitude of latent component C3 (in Panel B). Therefore, the peaks/troughs on an observed waveform cannot inform us which neural cluster is engaged in different activities. From the cognitive perspective, what is observed from the ERP waveform as semantic processing could be a conglomerate of many sub-processes. To add to the complexity, Hauk (2016) suggested that these sub-

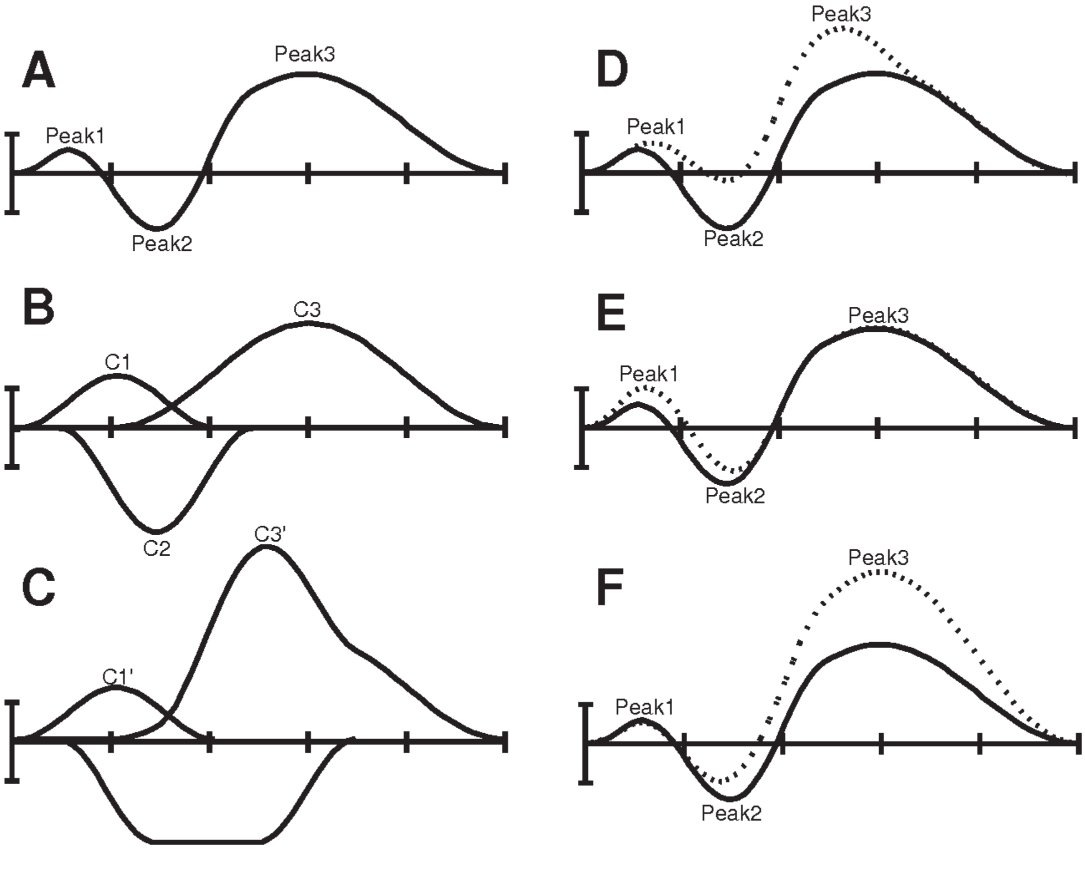


Figure 2.2: ERP waveform in Panel A can be decomposed to different latent components Panels B and C. Changes in the observed waveform (Panels D-F) could result from changes in different latent components. Figure from Luck (2005)

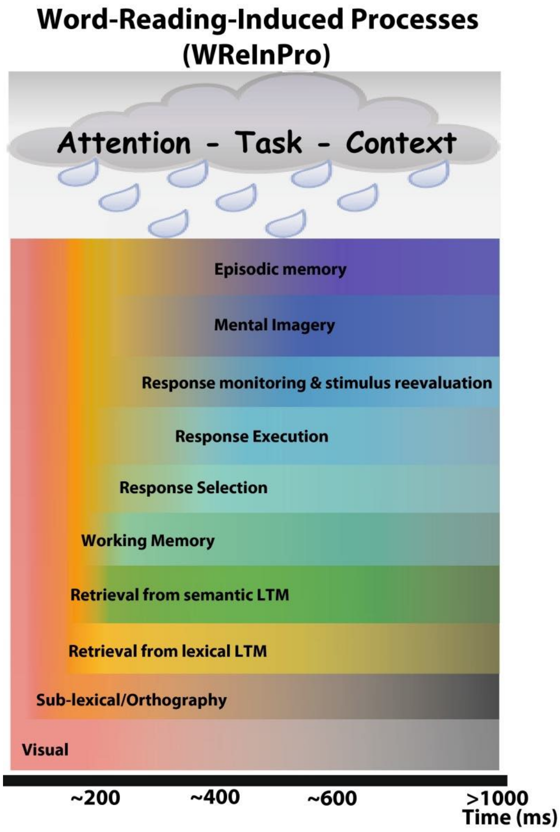


Figure 2.3: Schematic illustration of components in semantic processing in Hauk(2016)

processes, though carried out by different neural clusters, overlap and communicate with each other (Figure 2.3). The N400 and LPC effects of modality switching costs are subject to distortion from changes in all the sub-processes active during the N400 time window. Therefore, Luck (2005) has reasonably warned us that using raw waveform to study the effect of functional processes is "extremely dangerous". However, in spite of such a danger, there are strategies to design better studies overcome the issue of latent components, which will be discussed later when I outline the solutions available to my research.

2.3.1 Data structure of ERP.

The second problem of ERP components is that the structure of EEG data is largely overlooked. This problem in practice results in high researcher degrees of freedom, thereby adding the risk of an inflated probability of Type I error.

2.3.1.1 Continuity

The first characteristic of EEG data is that it is continuous. The data is recorded over the epoch at a high frequency (e.g., 1000Hz) and at different sites across the scalp (e.g., 128 channels). Therefore, the data is correlated across time and space. A data point at a certain time should be similar to the data point before and after it; a waveform of a certain channel should be similar to those of the adjacent channels. Such an intercorrelation poses a problem when researchers try to cut the continuous waveform into discrete components that is averaged across a time window and several channels during analysis.

Because of the continuous data structure, researchers face many choices during the analysis such as which time window and scalp sites to choose. When making these choices, researchers run the risk of *p*-hacking behaviour because there is the freedom to explore alternative methods of analysis and end up reporting only the analysis that "works" (J. P. Simmons, Nelson, & Simonsohn, 2011). This is the reason why we should try our best to remove flexibility for researchers from the analysis (Nelson, Simmons, & Simonsohn, 2018; Wagenmakers, Wetzels, Borsboom, & Maas, 2011). If response time is analysed, it is unacceptable to selectively analyse only those trials whose response times between two conditions are consistently larger than 100ms; or if a questionnaire contains 100 questions, it is unacceptable to only report 10 that has given a result in the direction expected.

We should similarly be aware of such risks when we conduct ERP studies, if not more so because of the great flexibility afforded by the sheer amount of data. For example, simply in order to choose a time/time window, there are three strategies available. Researchers could perform a peak-based analysis by selecting one time point when the waveform is at its peak/trough, a mean-based analysis which takes the mean amplitude within a time window, prescribed before analysis, upon visual inspection of the data, or inherited from previous studies. Otherwise, if there are no clear hypotheses for components, researchers could also analyse the whole waveform by slicing it into equal time intervals such as 50ms or 100ms slices. Similarly, to choose a scalp region, one can pick a cluster of electrodes upon visual inspection, or segregating the scalp into two halves (left/right, anterior or posterior), four quarters (left/right x anterior/posterior) or nine areas (left/mid/right x front/centre/back). The sheer number of choices available should be a cause of alert because of the flexibility it affords. Moreover, one of these choices, that is visual inspection, is particularly risky, because researchers are allowed to manually pick the time window and electrodes that are more likely to yield a big difference between conditions, whereas the data that do not seem to yield an effect will simply be discarded. As these options are all equally allowed, although it is not the researchers' intention to engage in dishonest behaviour, the risk of a Type I error ends up being higher than the critical value of 5%.

The inflation of the probability of Type I error leads to an over-confidence in the interpretation of the results. For example, in the studies on modality switching costs, different time windows were chosen to test the early effect of modality switching costs. While Bernabeu et al. (2017), Hald et al. (2011) and Hald et al. (2013) examined the segment of ERP waveform between 190ms and 300ms or 160ms and 215ms based on visual inspection and found significant effects, Collins et al. (2011) chose a prescribed time window between 100ms and 200ms before the analysis and did not find the effect. Both choices are accepted by the research community based on the fact that these studies are all published in high-impact, peer-reviewed journals. All of these segments are labelled as the N1-P2 effect. However, these windows in fact barely overlapped. It is hard to accept that they reflect the same underlying processing.

2.3.1.2 Clustered data

The second characteristic of EEG data, especially in psycholinguistic research, is that it is clustered. In other words, the data points are nested under participants and stimuli. The responses, or ERP in this case, of each participant should be more similar within this participant than across different participants. Similarly, the ERP of each stimuli should be more similar than the ERP across different stimuli. This hierarchical data structure is often neglected by conventional research which uses ANOVA that averages across participants or stimuli. Thus, the language-as-fixed-effect fallacy is committed (H. H. Clark, 1973). In other words, when you find an effect between two conditions by comparing the averaged waveform of each participant, you cannot rule out that the effect is a result of several peculiar stimuli. Ignoring random variability like this can lead to inflated Type I error rate (Amsel, 2011; H. H. Clark, 1973).

Traditionally, the language-as-fixed-effect fallacy has been treated by matching the stimuli on a series of psycholinguistic factors that could potentially affect language processing such as frequency, length for items. Similarly, it is also possible to control the random variability of participants by matching them by age, gender, working memory capacity, etc. However, this option is increasingly impractical given the vast number of factors, known and unknown, that need to be matched which grows as our understanding of language processing accumulates (Cutler, 1981).

In summary, characterising ERP using components is dangerous because the components obtained from observed waveform does not necessarily correspond to the latent components. Thus, the observed ERP components do not accurately reflect the critical process we hope to study and are easily distorted by changes of those surrounding processes. Furthermore, severing continuous waveform into discrete components and analysing them in ANOVA run a high risk of inflating Type I error. However, these problems do not mean that studying ERP is futile by any means. Instead, we need to be very cautious when designing an ERP study and it is imperative for this thesis to seek for a more rigorous way to characterise ERP data.

2.4 A better way for ERP

A more rigorous way to treat ERP data needs to solve three problems existing in the conventional method, which are,

1. how to examine the underlying processes in the brain when latent compo-

nents are not accessible;

2. how to limit researcher degrees of freedom;

3. how to account for multilevel, clustered data. In this section, I would argue that the way to take care of all three problems is to conduct component-free analysis, and to adopt more advanced statistical methods such as linear mixed-effect models and Bayes factors.

2.4.1 Problem with latent component

As discussed previously, observed ERP waveforms do not directly reflect underlying processes. However, it is still possible to interpret ERP data in terms of cognitive processes as long we design better study. Luck (2005) offered six strategies to avoid ambiguity in interpreting ERP components, which are (1) focusing on a specific component, (2) using well-studied experimental manipulations, (3) focusing on large components, (4) isolate components with difference waves, (5) focusing on components that are easily isolated, and (6) using component-independent designs. Among these strategies, (3), (4) and (6) are most relevant to my thesis. Strategy (3) does not merit much discussion because in this thesis we will pay attention mostly to N400 and LPC, both of which are large components. Here in this section, I would like to show how isolating components with difference waves (Strategy (4)) and using component-independent designs (Strategy (5)) are able to improve the analysis, with modality switching costs as an example.

Difference waves can isolate components of interests and remove the confound of other latent components (Vogel, Luck, & Shapiro, 1998). For example, to study modality switching costs, we are interested in the effect of a switch of perceptual modality (e.g., loud bright) compared with a non-switch (e.g., red bright), and we expect that a recruiting modality-specific resources for semantic representation should affect the amplitude of the N400 component. However, it is difficult to know whether the change in the amplitude of the observed waveform is a result of a change in the N400 component or some other components (e.g., P3). To isolate the N400 component, we could construct difference waves by subtracting the waveform of the prime item (e.g., *loud*) from the target trial (e.g., *bright*) for both the switching and non-switching conditions. The resulting waveform from the switching condition will reflect simply the effort to disengage from one perceptual modality (e.g., audition) and re-engage another perceptual modality (e.g., vision), while other processes of property verification are cancelled out; whereas the resulting waveform from the non-switch condition will not reflect such an effort. Therefore, the difference between the difference waves of switching and non-switching conditions will simply reflect the effort to engage in modality-specific resources.

Using component-independent designs is a better method to avoid the problem with latent component (Hillyard, Hink, Schwent, & Picton, 1973; J. Miller & Hackley, 1992; Thorpe, Fize, & Marlot, 1996). It is applicable to the study of modality switching costs because it does not matter which latent component is responsible for the observed change in the waveform. What is of concern is the time course of embodied simulation throughout the whole process of property verification. As one of the aims of the thesis is to test the linguistic shortcut hypothesis (i.e., the linguistic component reaches its peak before the embodied component, and is used to guide the activation of the embodied component), we are able to draw a theoretically significant conclusion if we find that the linguistic component can better predict the observed ERP waveform at an earlier time window than the embodied component.

In this thesis, I will adopt the latter strategy as the more economical choice, because it removes the unnecessary time and effort to create difference waves, especially as salvaging ERP components is not the aim of my research. It is also the more rigorous choice because it limits the researcher degrees of freedom, as will be discussed in the next section.

2.4.2 Problems with researcher degrees of freedom

The large flexibility in the analysis of ERP leads to problems with replicability. Conducting a component-independent analysis can effectively limit such flexibility. By putting away the need to identify discrete components, there is no room to alter the length of a time window or to choose various sites on the scalp. In my ERP studies, I will simply analyse the data one electrode at a time in uniformed time windows of 50ms. The length of the time window is decided simply to reduce data to a manageable amount. It is theoretically possible to take finer slices (e.g., 10ms) or perform point-by-point analysis at each millisecond. However, I believe the 50ms slices will be able to serve the purpose of this thesis.

An additional way to avoid some of the problems of *p*-hacking is to use Bayesian analysis rather than null-hypothesis significant testing (NHST), where statistical significance ($p < .05$ results) is no longer a concern and inferential statistics instead quantify the evidence in the data for or against a particular hypothesis (Wagenmakers, 2007). In NHST, statistical significance is measured by

p-value, which is the probability of obtaining the data if no actual effect exists in the population (i.e., if H_0 is true). Problems of p-value have been demonstrated extensively (Jarosz & Wiley, 2014; Wagenmakers, 2007), the most fundamental being that NHST is comparing the data to a distribution that is never observed (where H_0 is known to be true), due to which p-values cannot inform us about the size of effect or the confidence we can hold for the alternative hypothesis, even if it was shown to be significant. On the flip side, if NHST yielded non-significant result, the p-value does not allow us to conclude that the alternative hypothesis is false.

A practical encumbrance of p-values is the need for post hoc correction after multiple comparisons. It is typical for ERP studies to conduct multiple comparisons (e.g., running the same ANOVA repeatedly on different subsets of data like different time windows and different groups of electrodes. This would massively increase Type I error if no post hoc correction is conducted. However, if Bonferroni or other correction is conducted, it will render the study over-conservative, thus increasing the chance of Type II error. In the present thesis, 90 electrodes will be analysed individually, with 20 time slices in each trial. That results in 1800 NHSTs for each critical variable. A correction of multiple comparison will require a critical level of 2.78 x 10ˆ-5 for each test for a family-wise critical level of .05 (and an uncorrected test will almost definitely lead to false positive results). This stringent criterion could conceivably render it meaningless any p-values we can obtain from a statistical package. Bayes Factors avoid this problem completely because their quantification of evidence is unaffected by multiple comparisons, and so you can run as many Bayesian model comparisons as you like without increasing type 1

Table 2.1: Interpretation of BF as evidence for H 1

BF_{01}	BF_{10}	Support for H_1
1-.33	1-3	Anecdotal
.33-.10	3-10	Substantial
.10-.03	10-30	Strong
.03-.01	30-100	Very strong
<.01	>100	Decisive

error.

Bayes Factors denotes the ratio of probability of the data under both null and alternative hypothesis. The inverse of BF_{01} which quantifies how much more likely the data is under H_1 than H_0. That is, if the inverse of BF_{01} (i.e., BF_{10}) is 5, it means the current data is five times more likely under the alternative hypothesis than the null hypothesis. This can already inform us about how confident we can be with the alternative hypothesis. We can be more confident with the alternative hypothesis if $BF_{10} = 100$ than if $BF_{10} = 10$, while if $BF_{10} = 1$, the data was equally likely under either hypothesis. Inversely, if BF_{10} is less than 1, we would be more confident about the null hypothesis than the alternative hypothesis. Therefore, BF offers the possibility to conclude that the alternative hypothesis is a bad model for the data. While the continuous scale of BF is useful enough, we could conveniently describe the results, using (Jeffreys, 1998) guideline to label BF (Table 2.1)

There is one more possible measure to limit researcher degrees of freedom, thanks to recent endeavours of pre-registration and open science (Nelson et al., 2018; van 't Veer & Giner-Sorolla, 2016; Wagenmakers et al., 2011). It is recommended that all decisions concerning the design and analysis (e.g., number of participants, hypotheses, primary and secondary analyses, etc.) should be made and pre-registered before conducting a study. In order to create a more rigorous way to conduct ERP studies, pre-registration of methods to select time windows and scalp sites, as well as the hypotheses, should be carried out. This practice will also help differentiating hypothesis-driven analysis and exploratory analysis. Pre-registration is able to accommodate the need for exploring alternative and multiple methods of analysis, as long as they are clearly documented. Unfortunately, the ERP studies in my thesis are not able to be pre-registered because of the timing of which these studies were conducted during my PhD. However, the decisions during the analysis are made with limiting flexibility as the highest priority, and I will clearly distinguish hypothesis-driven analysis from exploratory analysis.

2.4.3 Problem with data structure

The structure of ERP data is clustered and inter-correlated. The best way to account for such data structure is mixed-effect models (Bagiella & Sloan, 2000). In a mixed-effect model, fixed effects of treatment (e.g., modality switch vs non-switch) from the random variability of participant, item, time and electrode. Recently, studies have emerged that treat EEG data with linear mixed-effect models (LME; Amsel, 2011) and generalised additive models (GAM; Hendrix, Bolger, & Baayen, 2017). These models are superior to ANOVA not only because they can better account for the data structure, but also because they can take continuous predictors. ANOVA requires factorial manipulation of independent variables, such as conventional versus novel metaphors, visual versus auditory properties. Many of these variables can be better operationalised on a continuous scale. For example, in my study of metaphor processing, the linguistic distributional pattern is better measured by co-occurrence frequency than a dichotomy of high association versus low association. Forcing continuous variables into categorical variables lowers the statistical power, thus making it less likely to detect an effect (Baayen, 2010). Therefore, mixed-effect models will allow us to analyse ERP data with greater power and rigour.

Ideally, GAM is a superior way to analyse ERP data because it can also allow for non-linear effect of different dimensions, including time and scalp site. However, LME has the advantage of producing results that are similar to those of ANOVA and linear models. Therefore, the results are more comprehensible and easily accessible for the readers. Furthermore, LME will be used to analyse the behavioural results of this thesis, which makes the practical transition to the ERP studies easier and less consuming (considering the time constraint of the PhD). Therefore, I have opted to use LME as a stepping stone, with that hope that future research can take the full data structure of ERP into account.

CHAPTER 3

Ease-of-Simulation Norms

In order to examine the conceptual representations of metaphor processing, I need well operationalised definitions of the two components of metaphor processing, that is the distributional pattern of linguistic symbols and the simulation of sensory, motor, affective, introspective and other bodily experiences. While the former could be straightforwardly operationalised by distributional frequencies (i.e., logarithmically transformed sum of n-gram frequencies throughout the thesis), the latter poses a challenge. This chapter presents a novel norming measure of the effort of creating embodied simulations, by examining the end product of successful metaphor processing. The resulting interval metric will be used as one of the key predictors of metaphor comprehension in the following chapters.

Draft was submitted to the journal Behavioral Research Methods in January 2018 and was produced in collaboration with Dr Louise Connell and Dr Dermot Lynott as co-authors.

Paper 1: Ease-of-Simulation Norms for 452 Adjective Metaphors

Page numbers: 65 - 96

Ease-of-Simulation Norms for 452 Adjective Metaphors

Pei Q. Liu[1] & Louise Connell[1] & Dermot Lynott[1]

[1] Department of Psychology, Lancaster University

Author Note

Department of Psychology, Lancaster University, UK.

Correspondence concerning this article should be addressed to Pei Q. Liu, Department of Psychology, Fylde College, Lancaster University, Bailrigg, LA1 4HX, UK. E-mail: p.liu1@lancaster.ac.uk

Abstract

Successful language comprehension results in a multimodal simulation of meaning. Hence, the ease of simulation of a given metaphor is likely to be a critical predictor of performance in metaphor processing. Since simulations are unconscious, and may lose information when brought to conscious awareness, the ease of simulation cannot be directly rated (e.g., as ease of understanding or of generating mental imagery). Therefore, we propose an alternative method of measuring ease of simulation indirectly by assessing the end product of successful metaphor processing using multiple rating scales that, together, offer a proxy for ease of simulation: sensibility, usability, and imaginability of metaphors. We normed 452 sentences comprising adjective metaphors (e.g., Students can be bright; Minutes can be lukewarm) by asking participants to rate the sentences on sensibility, usability, and imaginability scales. We then used principal components analysis to combine these three individual rating scales into a single measure that captured their common variance: ease-of-simulation (EoS). To validate the norms, we examined the extent to which this new EoS measure can explain response times in a metaphor processing task, and found that it better predicted sensibility judgement times than any of the individual ratings, including the ostensibly more relevant sensibility ratings. These findings support the idea that EoS provides a valuable index of processing difficulty in metaphor comprehension. We hope these norms will benefit research on metaphor processing and the role of simulation in language comprehension.

Keywords: norms, ease of simulation, metaphor processing, language comprehension

Word count: 4553 words

Ease-of-Simulation Norms for 452 Adjective Metaphors

Successful language comprehension, regardless of being literal or metaphoric, results in multimodal simulation of the concepts described in the sentence. That is, sensorimotor, affective, introspective and other experiences are reactivated in order to achieve comprehension (Allport, 1985; Barsalou, 1999; Glenberg & Gallese, 2012; Niedenthal, 2007). Evidence for such simulation includes shared activation between brain areas involved in perceptual processing (e.g. sound stimuli activating auditory cortex) and comprehension of words whose meaning relates to perception (e.g., sound-related words like *thunder*; Bonner & Grossman, 2012; Boulenger et al., 2008; Hauk, Johnsrude, & Pulvermüller, 2004), as well as many other behavioural effects which supports the simulation-based account of language processing (Aravena et al., 2010; Buccino et al., 2005; Connell, 2007; Glenberg et al., 2008; Speed & Majid, 2017). Moreover, even abstract concepts – which traditionally have been assumed to lack perceptual information – have been shown to be grounded in people's sensorimotor and affective experiences (Connell, Lynott, & Dreyer, 2012; Glenberg et al., 2008; Vigliocco et al., 2014).

However, most research on simulation has focused on literal language processing, while metaphor processing has generally been neglected. Since simulation happens for both literal and figurative language processing, ease of simulation should be a critical factor in performance (e.g., speed and accuracy) during metaphor processing. The aim of this paper is to address the lack of attention given to embodied (i.e., grounded) views of meaning in metaphor processing by defining and operationalising a measure of ease of simulation as an index of comprehension difficulty, and then to validate this measure using results from a metaphor sensibility judgement task.

Defining Ease of Simulation

In literal language processing, the extent to which simulation happens is straightforward to operationalise because there is an apparent match between the meaning of the language and the sensori-motor and affective characteristics of the referent. If you would like to see whether comprehension of *The sun can be bright* is grounded in sensorimotor simulation, you could examine whether it activates the neural regions related to processing visual stimuli, you could ask people to read the phrase while performing visual distraction task, or you could ask simply how strongly their experience of the sentence involves the visual modality. However, simulation in metaphor processing is more complex because the grounded characteristics of a metaphor are not directly corresponding to the characteristics of its constituent words. In other words, the conceptual representation of *Students can be bright* is not directly associated with the visual feature of emitting much light. Instead, bright here refers to intelligent and quick-witted. The association between the literal and metaphoric meanings is not always easily apparent and there is no consensus regarding the mechanism through which the metaphoric meaning arises. *Bright students* may be used to describe an intelligent student because intelligence is the most salient meaning in the context of *students* (Giora, 1997), or because the feature *bright* is projected onto *students* and the meaning of intelligence is inferred (Bowdle & Gentner, 2005), or because bright is a member of the figurative category intelligence (or intelligent things) of which *students* is also a member (Glucksberg & Keysar, 1990). Conversely, bright may be used to describe intelligence because both concepts can be compressed into one cross-modal dimension (Buckner & Krienen, 2013; Simmons & Barsalou, 2003); or because it may activate some abstract features apart from the concrete feature of emitting light, which could be used to describe intelligence (e.g., making things clear to see; Simmons & Barsalou, 2003); alternatively, the word *bright* may often co-occur with *clever* or *quick-witted* or other words referring to intelligence which consolidates this meaning (for a review on conceptual abstraction, see Barsalou, 2016; Louwerse, 2011; Zwaan, 2016). Moreover, simulation is an unconscious mental representation which loses information when brought to conscious awareness (Connell & Lynott, 2016), and so it is not possible to rate the ease of such complex simulation directly, as, for example, the ease of understanding a metaphor (i.e., comprehensibility: Campbell & Raney, 2015; Katz, Paivio, Marshark, & Clark, 1988) or of generating mental imagery for a metaphor (Cardillo, Schmidt, Kranjec, & Chatterjee, 2010). To solve this problem, we propose to measure ease of simulation indirectly by assessing the end product of successful metaphor processing in terms of three criteria that, together, offer a proxy for the ease of simulation: sensibility, usability and imaginability of metaphors.

These three criteria tap into different, but complementary, aspects of language comprehension and conceptual representation. Sensibility, that is, how much sense a sentence makes if encountered in reading or in conversation (e.g., *Students can be bright*), is a useful measure that relates to whether the process of language comprehension seems to have produced a functional end product. By successfully "making sense" of a metaphor, people will manage to construct a conceptual representation of the metaphoric meaning in question. The drawback of a sensibility rating is that it is often conflated with frequency. Studies using a sensibility judgement task, where participants decide whether or not a sentence makes sense, have found that performance of the task could be well explained by the distributional frequency of constituent words (Connell & Lynott, 2013; Gagné & Shoben, 1997; Storms & Wisniewski, 2005). This relationship could be because a sensibility judgement task requires relatively shallow processing as a simple yes/no answer would suffice (Barsalou, Santos, Simmons, & Wilson, 2008; Connell & Lynott, 2013). When it comes to sensibility ratings, people may also base their decisions on the statistical, distributional patterns of the constituent words, such that, if the constituent words often occur together (e.g., *account* and *freezing*), they would find it easier to pass it as something they can understand (however vaguely), compared to words that rarely co-occur (e.g., *illness* and *bright*). Thus, in order to encourage people to engage in deeper processing of meaning, we adopted the second criterion: usability, that is, how easy it is to use the sentence in conversation or in text. Usability supplements sensibility because it requires people not only to understand and make sense of a sentence, but also to imagine scenarios where they could use the sentence in language production: a feat that would be very difficult without thorough and confident grasp of what the metaphor actually means. Thus, the usability rating will require deeper processing than the sensibility rating.

In contrast to the sensibility and usability criteria which focuses on language processing, the imaginability criterion – that is, how easy it is to imagine the concept described in a metaphor – is introduced to give a lay definition of simulation. Although imagining a concept requires conscious effort while simulation is unconscious, both processes depend on generating, retaining and even manipulating mental representations. Note that imaginability is not imageability, the latter of which is a criterion biased towards concrete concepts and especially the visual modality (Connell & Lynott, 2012; Paivio, Yuille, & Madigan, 1968). In contrast, imaginability (i.e., how easy it is to imagine the concept described in a metaphor) can be easily applied to abstract concepts. People could find it as easy to imagine an abstract concept such as a bright future as a concrete concept such as a bright student.

These three measurements can offer a proxy of how easy it is to simulate a concept in a successful representation of metaphoric meaning. If a metaphor is considered as sensible, easy to use, and easy to imagine, such a metaphor should normally be regarded as easy to simulate. We can therefore combine these three ratings into a single component measure that captures overall ease-of-simulation

(EoS); in other words, we propose that ease of simulation provides an overall index of the comprehension difficulty or processing demands of a given metaphor. If the ease-of-simulation measure is a critical predictor of metaphor processing, it should outperform the individual component scales in predicting how fast and accurately metaphors are processed. Therefore, following details of the norming study, we subsequently validate the norms by analysing response times from a metaphor sensibility judgement task using the ease-of-simulation measure and the three individual scales separately as predictors.

Norming Procedure

This norming study aimed to operationalise ease of simulation in metaphor processing. For this purpose, we generated metaphoric sentences in the form of "A can be B" in which A was a noun as the target word and B was a sensory adjective as the source word (e.g., Student can be bright). We asked participants to rate these sentences on their sensibility, usability and imaginability, and we combined these three sets of ratings using principal components analysis to obtain a single score for each item (i.e., ease-of-simulation measure). The full set of metaphors, along with ratings, the ease-of-simulation measure, and distributional frequencies, will be made public online (https://osf.io/xgysz).

Method

Participants. We recruited 171 students from Lancaster University to complete the survey. All were native speakers of English. They took part in the study either online for a chance to win a £10 Amazon voucher (N = 112) or in a lab for a reward of £3 (N = 59).

Materials. A basic set of 113 object properties (i.e. adjectives), that were unimodal in nature (i.e., had a single dominant perceptual modality: see also Liu, Connell, & Lynott, 2018b), were selected from the modality exclusivity norms of Lynott and Connell (2009). Each adjective was paired with four nouns (chosen from various sources such as the Kenkyusha Dictionary of English Collocations (Katsumata, 1980), Macmillan English Dictionary (Rundell, 2007), Oxford Dictionary of English (Stevenson, 2010)) in order to generate sentences with different metaphoric interpretations, giving a total of 452 metaphors. When generating the metaphors, we made sure that it was the adjectives, rather than the nouns, that were interpreted metaphorically, and that the metaphors did not have a literal meaning readily accessible (e.g., *Feet can be itchy*, meaning wanting to travel, was rejected as a possible item because both *feet* and *itchy* were interpreted metaphorically and the literal meaning of an itching sensation was readily accessible too).

The combination of each adjective with four nouns enabled the metaphors to vary on two dimensions: that is, the authors' initial intuitions of the ease of simulation and a measure of linguistic distributional frequency (i.e., how often the adjective and noun co-occurred). Linguistic distributional frequency has been shown to affect conceptual representations and the speed of language processing (Andrews, Vigliocco, & Vinson, 2009; Connell & Lynott, 2013; Landauer & Dumais, 1997; Louwerse & Jeuniaux, 2010), because things that often appear together in the world also tend to be mentioned together in language (Kintsch & Dijk, 1978; Louwerse & Jeuniaux, 2010). Therefore, when norming the ease of simulation for the adjective metaphors in this study, we also need to take the distributional frequency into account to make sure that any measure of ease of simulation is not entirely confounded by linguistic distributional frequency. In order to do so, linguistic distributional frequency and ease of simulation were varied independently when creating metaphors for each adjective, so that items could be roughly categorised into 4 groups (easy-high, hard-high, easy-low, hard-low: see Table 1).

Table 1
Sample metaphors, and their scores for EoS and LDF

Metaphor	EoS	LDF
Illness can be bright.	-1.32	2.95
Supply can be bright.	-1.02	3.72
Solutions can be bright.	1.41	3.11
Students can be bright.	1.84	4.08
Minutes can be lukewarm.	-1.49	3.31
Scheme can be lukewarm.	-0.78	1.61
Supporters can be lukewarm.	0.69	2.69
Reaction can be lukewarm.	1.03	3.45

Ease of simulation ranged from intuitions of easy (e.g., the meaning of the metaphor *Students can be bright* is likely to be represented relatively easily as a quick-witted, intelligent students) to difficult (e.g., the meaning of the metaphor *Supply can be bright* is likely to be represented only with great difficulty, if at all). The metaphors also varied on the dimension of linguistic distributional frequency, which was determined by the bi- to 5- gram co-occurrence frequencies between the adjective and the noun with zero to three intervening words in the Google Web1T Corpus [Brants and Franz (2006); e.g., "bright students" or "students bright", "bright ... students" or "students ... bright", etc.]. Frequencies were log transformed as $\log_{10}(\text{frequency} + 41)$ because} 41 is the lowest non-zero frequency in the corpus (i.e., the raw frequency of 0 in fact means that the constituent words co-occurred 41 times;} e.g., Connell & Lynott, 2013). Thus, we obtained a linear scale of linguistic distributional frequency for all metaphors normed (M = 2.95, SD = 0.97). Linguistic distributional frequency ranged from high (e.g., *Student can be bright*) to low (e.g., *Solution can be bright*).

Procedure. In order to operationalise ease of simulation, we asked participants to rate the sentences on three scales – sensibility (i.e., how much sense the expression makes as if read in a story or heard in conversation), usability (i.e., how easy it would be to use the expression in writing or conversation), and imaginability (i.e., how easy it is to imagine the concept described in the expression). In the survey, each metaphoric expression was rated on these three criteria on a 1-7 Likert scale, with 1 being "making no sense at all", "very hard to use/imagine" and 7 being "making perfect sense", "very easy to use/imagine" (full instructions are given in Appendix A).

With the four groups of adjective-noun pairs (easy/hard to simulate x high/low linguistic distributional frequency) and three scales of rating, we generated twelve list conditions of items to which each participant was assigned based on their month of birth. Each participant saw 113 items that ranged across a full spectrum from easy to hard to simulate, and from high to low distributional frequency, and were distributed in approximately equal proportions across the three ratings scales. Thus, all items were rated on all three scales, but each participant rated a particular adjective only once. The order of item presentation under each scale was randomised prior to the study and the order of scale presentation was randomised for each participant. Participants were not compelled to rate all items but could complete as many as they wanted to.

The survey was administered both online and in a lab setting with the permission of the Lancaster University's Department of Psychology Ethics Committee and Lancaster University's Research Ethics Committee. In both instances, participants completed the entire study via the Qualtrics online survey platform.[1] Here, we report how we determined our sample size, all data exclusions,

[1] We compared ratings in the lab versus online using mixed-effect linear models with the test setting as the fixed effect, and participants and metaphor item as random-intercept effects; overall, online ratings were higher than the ratings in the lab for both imaginability (b = -0.46, 95% CI = -0.79 - -0.12, df = 164.09, t = -2.65, p = .009) and sensibility (b = -0.51, 95% CI = -0.82 - -0.20, df = 160.40, t = -3.20, p = .002), but not for usability (b = -0.25, 95% CI = -0.58 - 0.09, df = 166.09, t = -1.44, p = .15). This comparison should be treated with caution because online participants

all manipulations, and all measures in the study.[2]

Analysis and Discussion

Data exclusion and treatment. A total of 17,329 data points was collected. Participants who completed less than one-third of the items each scale (i.e., 13 items or fewer) were removed (three participants for imaginability, nine for sensibility, eight for usability). Those who provided the same rating value for 13 or more items in a row were removed too (one for imaginability, one for sensibility, four for usability). Two participants' ratings on imaginability were removed because their small number of data points, combined with low variability, meant that calculation of Cronbach's alpha could not proceed (error message: "missing value in the correlation R-package {psych}: Revelle, 2017). Thus, 16,888 data points remained after exclusion, with a minimum of 10 points per item per scale. The inter-rater

greatly outnumbered the participants in the lab, with some items rated more than ten times online but only once or twice in the lab.

[2] We used R (Version 3.5.0; R Core Team, 2017) and the R-packages *abind* (Version 1.4.5; Plate & Heiberger, 2016), *arm* (Version 1.10.1; Gelman & Su, 2016), *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2015), *bindrcpp* (Version 0.2.2; Müller, 2017), *bookdown* (Version 0.7; Xie, 2016), *broom* (Version 0.4.4; Robinson, 2017), *citr* (Version 0.2.0; Aust, 2016), *coda* (Version 0.19.1; Plummer, Best, Cowles, & Vines, 2006), *contrast* (Version 0.21; Kuhn, Steve Weston, Wing, Forester, & Thaler, 2016), *cowplot* (Version 0.9.2; Wilke, 2017), *data.table* (Version 1.11.4; Dowle & Srinivasan, 2017), *doBy* (Version 4.6.1; Højsgaard & Halekoh, 2016), *dplyr* (Version 0.7.5; Wickham, Francois, Henry, & Müller, 2017), *Formula* (Version 1.2.3; Zeileis & Croissant, 2010), *ggplot2* (Version 2.2.1; Wickham, 2009), *gridExtra* (Version 2.3; Auguie, 2017), *Hmisc* (Version 4.1.1; Harrell Jr, Charles Dupont, & others., 2018), *interplot* (Version 0.1.5; Solt & Hu, 2015), *knitr* (Version 1.20; Xie, 2015), *lattice* (Version 0.20.35; Sarkar, 2008), *lme4* (Version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.0.1; Kuznetsova, Brockhoff, & Christensen, 2017), *magrittr* (Version 1.5; Bache & Wickham, 2014), *MASS* (Version 7.3.50; Venables & Ripley, 2002), *Matrix* (Version 1.2.14; Bates & Maechler, 2017), *mgcv* (S. N. Wood, 2003, 2004, Version 1.8.23; 2011; S. Wood, N., Pya, & S"afken, 2016), *multcomp* (Version 1.4.8; Hothorn, Bretz, & Westfall, 2008), *MuMIn* (Version 1.40.4; Bartoń, 2017), *mvtnorm* (Version 1.0.8; Genz & Bretz, 2009), *nlme* (Version 3.1.137; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017), *papaja* (Version 0.1.0.9735; Aust & Barth, 2017), *plyr* (Wickham, 2011; Version 1.8.4; Wickham et al., 2017), *psych* (Version 1.8.4; Revelle, 2017), *purrr* (Version 0.2.5; Henry & Wickham, 2017), *rcartocolor* (Version 0.0.22; Nowosad, 2017), *rms* (Version 5.1.2; Harrell Jr, 2018), *SparseM* (Version 1.77; Koenker & Ng, 2017), *survival* (Version 2.42.3; Terry M. Therneau & Patricia M. Grambsch, 2000), *TH.data* (Version 1.0.8; Hothorn, 2017), and *tidyr* (Version 0.8.1; Wickham & Henry, 2017) for all our analyses and the writing up of this manuscript.

Table 2
Correlations among variables with their means and standard deviations (SD)

	1.	2.	3.	4.	5.
1. EoS	–	–	–	–	–
2. Imaginability	.95	–	–	–	–
3. Sensibility	.96	.86	–	–	–
4. Usability	.96	.86	.87	–	–
5. LDF	.27	.23	.26	.28	–
Mean	0.00	3.98	3.83	3.63	2.96
SD	1.00	1.44	1.55	1.48	0.97

consistency of each scale was calculated per group (months of birth from January to December) per scale, and showed that response was highly consistent among participants (imaginability average $\alpha = 0.888$, sensibility average $\alpha = 0.907$, usability average $\alpha = 0.896$). Detailed results of Cronbach's alpha are reported in Supplementary Materials (https://goo.gl/UT76mL).

Extracting the ease-of-simulation (EoS) measure. We first calculated means and standard deviations of each item per scale, and then we performed principal components analysis via covariance matrix on the three sets of ratings in order to extract one factor that can quantify the ease of simulation. This single factor had an eigenvalue of 2.73, was the only principal component with eigenvalue > 1, and explained 91% of the total variance of the ratings. The fact that sensibility, usability, and imaginability ratings could be collapsed so successfully into a single component suggests that these three rating scales essentially captured the same underlying information: that is, successful simulation of meaning.

We then used the factor score of this single principal component (estimated via regression as a standardised variable) as the EoS score for each metaphor item. That is, the higher the EoS score, the easier it was to simulate the metaphoric meaning. Table 2 shows the descriptive statistics of and inter-correlations between the EoS measure and the three rating scales. It is notable that the three individual ratings scales correlate equally well with each other but less well than they do with the extracted ease of simulation measure, suggesting that each rating scale captures unique variance as well as that related to ease of simulation. We test the utility of this unique variance in the norms validation.

Comparison with linguistic distributional frequency. In creating our metaphors, we attempted to ensure that EoS was not confounded by linguistic distributional frequency; that is, the ease of simulating a particular metaphoric meaning should not be a function of how frequently the words in that metaphor co-occur in language. In order to examine if we were successful in our efforts, we calculated the relationship between the log linguistic distributional frequency for each item and the corresponding measures from our norming procedure. As shown in Table 2, linguistic distributional frequency had a relatively consistent, positive relationship with EoS scores and their constituent imaginability, sensibility, and usability ratings. However, the relationship was weak, with linguistic distributional frequency explaining only 7.3% of the variance in EoS scores. As such, ease of simulation is not strongly influenced by linguistic distributional frequency: the words that comprise an easy-to-simulate metaphor may co-occur seldom or often, as may those that comprise a hard-to-simulate metaphor. Examples of metaphors varying on the EoS measure and linguistic distributional frequency are shown in Table 1.

Norms Validation

If ease of simulation is a critical factor in metaphor processing, as assumed by grounded theories for all language processing, then our newly created EoS measure should outperform the three individual scales (imaginability, sensibility, usability) in predicting the time-course of metaphor processing. That is, since each of the three individual rating scales captures only some of the information that exemplifies a successful perceptual-motor-affective simulation of metaphoric meaning, their individual ability to predict metaphor processing performance will be less than that of the composite EoS score. We therefore validated our norms by analysing response time (RT) data from a typical metaphor comprehension study: a sensibility judgement task, taken from a previous study by Liu, Connell, and Lynott (2018a). In that experiment, 28 participants read the 452 metaphors normed here and judged whether or not the sentences made sense. The sentences were presented one word at a time, and RT was measured from the onset of the adjectives to the onset of response.

Liu et al. (2018a) found that the EoS scores obtained from the present norming study were a significant predictor of the RT to judge the metaphors as sensible. Here, we directly compare ease of simulation with imaginability, sensibility, and usability ratings in terms of their abilities to predict these sensibility judgement times. We analysed each predictor separately in a mixed-effect linear regression, where the key predictor (EoS, imaginability, sensibility, or usability) was a fixed effect, and participant and item were crossed random factors modelling random intercept. The imaginability, sensibility and usability ratings were all centred before analysis.

Furthermore, we also investigated whether a visual dominance exists in the EoS norms. There is an overall visual dominance in language (San Roque et al., 2015), which means that visual metaphors may be more commonly encountered, and thus may be easier to make sense of, imagine or use. Therefore, it is possible that the EoS measure may be a good predictor for visual metaphors only. To examine the possibility, we coded the data by the perceptual modality of the adjective source domain (audition, haptics and vision), and analysed whether the EoS effect on auditory and haptic metaphors differed from that on visual metaphors.

Table 3
Results from the mixed-effects linear regressions of RT (ms) in a sensibility judgement task

	b	95% CI	*p*
EoS	-149.85	-177.95 – -121.74	< .001
Imaginability	-98.12	-117.70 – -78.53	< .001
Sensibility	-88.29	-107.02 – -69.55	< .001
Usability	-96.09	-115.18 – -76.99	< .001

Note. *b* is non-standardised regression coefficient
Variables were fixed predictors in separate models.

EoS Measure versus Individual Rating Scales

All four variables (EoS, imaginability, sensibility, usability) had a significant negative relationship with RT in their respective analyses. That is, the easier it is to simulate the meaning of a metaphor (or the more sensible, easy to use, or easy to imagine the metaphor seemed to be), the faster it was accepted as sensible. Table 3 shows the effects of the variables separately in each model.

From the table, it can be seen that the EoS measure had a numerically larger effect size (b = -149.846) than the other variables, suggesting it as a better predictor of the response times compared to the other measures. To evaluate evidence for the advantage of the EoS measure over others, we calculated the relevant Bayes Factors for the purpose of model comparison. Bayes factor (BF) is a powerful tool to compare models and shows the ratio of the likelihood of data given different hypotheses. It could be used to make a statement about the amount of evidence for the alternative hypothesis compared against the null hypothesis, in contrast to a simple null-hypothesis significance testing using p-values (Jarosz & Wiley, 2014; Wagenmakers, 2007). In the current study, we contrasted the model with the EoS measure (alternative hypothesis H_1) with each of those models with individual ratings (imaginability, sensibility or usability: null hypotheses H_0) and calculated the amount the evidence that favoured the EoS model (BF_{10}).

We used the Bayesian Information Criterion (BIC) to estimate the likelihood of data under each hypothesis (Wagenmakers, 2007). We extracted BIC for each model, and then calculated BF10 as follows.

$$BF_{01} = e^{\frac{(BIC(H_1) - BIC(H_0))}{2}}$$

In all three comparisons, the Bayes Factors showed very strong evidence in favour of the EoS variable, in comparison with imaginability rating ($BF_{10} = 101.36$), sensibility rating ($BF_{10} = 5174.10$), and usability rating ($BF_{10} = 95.15$). Figure 1 visually illustrates the advantage of ease of simulation over the other measures (for details of model comparison, see Supplementary Materials). In other words, ease of simulation was the best predictor of the speed of accepting metaphors as sensible. This finding suggests that the unique variance in each individual rating scale is noise rather than useful information regarding processing difficulty, and that the shared variance between the scales – as captured by the composite EoS measure – succeeds in capturing useful information about meaning representation that indexes the latency of comprehension.

Examining the Dominance of Visual Metaphors

Although the EoS measure was shown to be the best predictor of RT, it could be argued that its effect is mainly driven by visual metaphors (i.e., metaphors where the literal meaning of the adjective relates to vision, such as bright), which make up 20% of our item set, and could be easier to process simply due to the fact that most sensory language concerns vision as opposed to other modalities (San Roque et al., 2015). We analysed the sensibility judgment times in a mixed-effect hierarchical linear regression with EoS as well as the perceptual modality of the source adjectives.

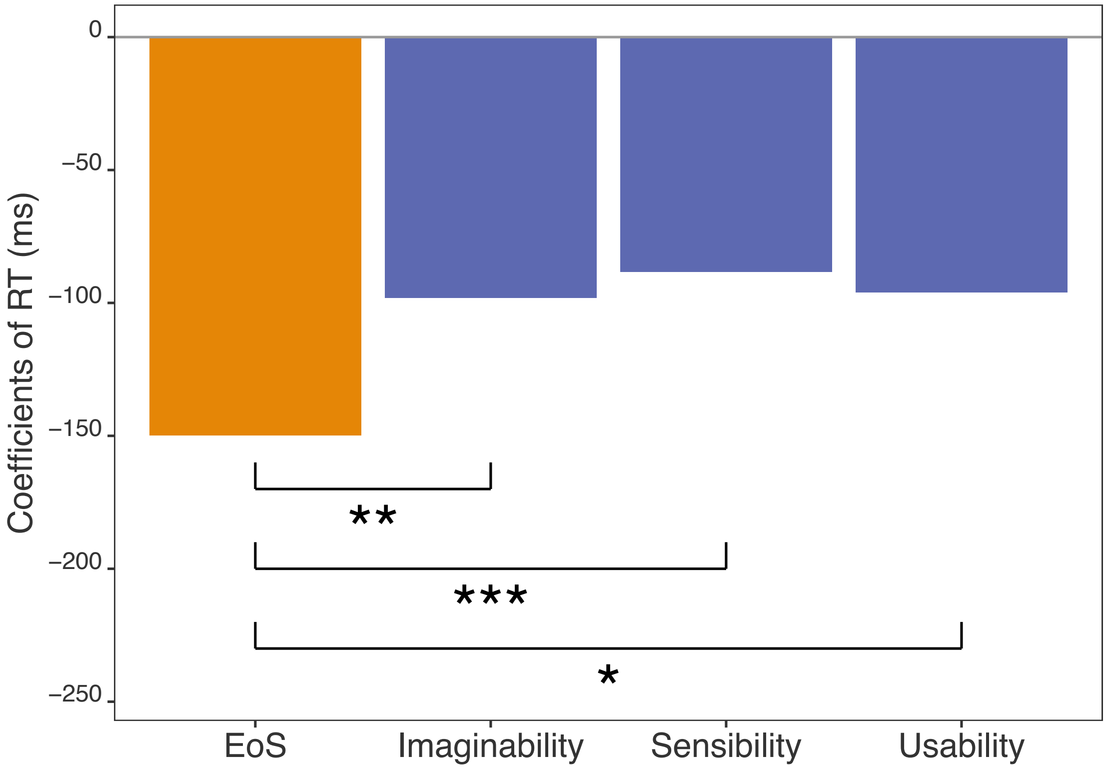


Figure 1. Effects of EoS measure, imaginability, sensibility and usability on RT (in order of the predicting power). The asterisks show the amount of evidence in favour of the EoS model. *: $BF_{10} > 10$; **: $BF_{10} > 100$; ***: $BF_{10} > 1000$.

Model 1 included the EoS measure and perceptual modality as fixed variables and participants and items as crossed random-intercept variables, and Model 2 included the critical EoS x perceptual modality interaction as additional fixed variables. The perceptual modality was dummy coded with the visual modality as the reference level. The EoS measure remained a significant predictor of RT (see Table 4). Furthermore, response to auditory metaphors were significantly slower than visual metaphors, while RT to haptic metaphors was not significantly different from that to visual metaphors. However, the interactions between source perceptual modality and EoS measure were not significant and the marginal R^2 (Bartoń, 2016) showed that the interaction term could not account for the data better. In other words, EoS did not have different effects on metaphors of different source modalities.

Table 4
Summary of Hierarchical Model Comparison for Variables to Predict Sensibility Judgement RT

	b	95% CI	R^2	ΔR^2
Model 1			0.06	
EoS	-147.77	-176.34 – -119.20		
Perceptual modality				
Auditory - visual	37.90	-20.23 – 96.04		
Haptic - visual	52.16	-17.47 – 121.79		
Model 2			0.06	-5.33 x 10^{-05}
EoS	-145.4640	-188.03 – -102.89		
Perceptual modality				
Auditory - visual	41.00	-25.69 – 107.70		
Haptic - visual	51.77	-35.08 – 138.62		
EoS x perceptual modality				
EoS x Auditory - visual	-6.34	-67.76 – 55.09		
EoS x Haptic - visual	0.70	-75.49 – 76.90		

Note. b is non-standardised regression coefficient
R^2 stands for marginal R^2, which represents the variance explained by the fixed factors only.

General Discussion

We presented ease of simulation norms for 452 metaphors with 113 unimodal sensory adjectives as the source terms (e.g., *Student can be bright*). In literal use, such adjectives relate to strong perceptual experiences in one of the visual, auditory or haptic modalities (Lynott & Connell, 2009). However, when these unimodal adjectives become part of a metaphor, the resulting mental representation is no longer unimodal, but becomes multimodal and complex. Moreover, multimodal simulations are prone to information loss when consciously inspected (Connell & Lynott, 2016), which creates difficulty in directly measuring how easy it is to simulate such metaphoric constructions. Thus, the norms presented here adopt a novel approach of taking indirect measures of how successfully the metaphoric meaning had been represented (i.e., the end product of the simulation process) in order to offer a proxy for ease of simulation. If the meaning of a phrase has been successfully simulated, it should be easy to imagine the concept it conveys, it should make sense, and it should be easy to re-use the phrase in language production. From these three different rating scales (i.e. imaginability, sensibility and usability), we used principal components analysis to create a composite measure that explained most of the underlying variance in a single variable, and termed this component ease of simulation. We validated this measure by using it to predict metaphor processing times in a sensibility judgement task. Critically, the predictive power of the EoS measure was higher than that of any individual scales. It suggests that the EoS measure can capture more comprehensively the information that exemplifies the simulation of meaning in metaphor processing. Thus, EoS scores provide a useful index of metaphor comprehension difficulty. Lastly, the EoS measure was found to be equally predictive for metaphors with visual target words as for those with auditory or haptic target words, thus rejecting the hypothesis that the EoS measure was a better measurement for visual metaphors because of the visual dominance.

Considering that the data for the norms validation was from a sensibility judgement task, it was striking to find that the EoS measure was a better predictor than the separate measure of sensibility ratings. If the sensibility judgement task and the sensibility ratings rely on the same mechanism (i.e., deciding whether/how much a metaphor makes sense to the reader), it should be natural and even trivial to find that sensibility ratings were an ideal predictor of the response time in the sensibility judgement task. That is, metaphors that are normally considered to make a lot of sense should be quickly judged as sensible, while metaphors that are considered to make little sense should take longer to be judged as sensible, if not rejected altogether. If that had been the case, our norming study would have had little theoretical value. However, since sensibility ratings involve consciously inspecting the contents of a mental representation, and since it is not possible to inspect the full contents of a simulation without losing information (Connell & Lynott, 2016), we suspected that sensibility ratings – like usability and imaginability ratings – would capture only some of the necessary criteria that determine how quickly metaphors are processed and judged. Indeed, we found that EoS scores, which were derived from usability and imaginability as well as from sensibility, was the best predictor of sensibility judgement times. This finding suggests that sensibility ratings include a good deal of noise that is unrelated to the ease of metaphor comprehension, as well as useful information (shared with the other ratings of usability and imaginability) that is ultimately passed onto the EoS measure. By contrast, the EoS norms capture a useful measure of the end-product of successful metaphor comprehension (i.e., a grounded simulation), and hence provide a useful index of metaphor comprehension difficulty.

With the EoS norms, we manage to measure the effort to arrive at a mental representation of a metaphor (i.e., a simulation of sensorimotor, affective and introspective experience), while remaining agnostic to the precise content or mechanism of such a representation. The sensory adjectives in this norming study are known to relate to strong sensory experiences in one of the visual, auditory and haptic modalities, at least for their literal meanings. However, when put in a metaphor, they generate mental representations that are multimodal and complex, which raises many further questions. What is simulated exactly? In which aspects of experience are these representations of metaphoric meaning grounded? How might one investigate these representations directly? By operationalizing and quantifying ease of simulation in metaphor comprehension, we believe we have taken the first step to answer these questions. Moreover, by making the norms public, it is our sincere hope that they can be a stepping stone for the research community to study simulated representations in metaphor processing.

References

Allport, D. A. (1985). Distributed memory, modular subsystems and dysphasia. In S. Newman & R. Epstein (Eds.), *Current perspectives in dysphasia* (pp. 207–244). Edinburgh: Churchill Livingstone. Retrieved from https://www.scribd.com/document/17676263/Allport-1985-Distributed-Memory-Modular-Subsystems-and-Dysphasia http://scholar.google.com/scholar?hl=en{\&}btnG=Search{\&}q=intitle:Distributed+mem

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. doi:10.1037/a0016261

Aravena, P., Hurtado, E., Riveros, R., Cardona, J. F., Manes, F., & Ibáñez, A. (2010). Applauding with closed hands: neural signature of action-sentence compatibility effects. *PloS One*, *5*(7), e11751. doi:10.1371/journal.pone.0011751

Auguie, B. (2017). *GridExtra: Miscellaneous functions for "grid" graphics.* Retrieved from https://CRAN.R-project.org/package=gridExtra

Aust, F. (2016). *Citr: 'RStudio' add-in to insert markdown citations.* Retrieved from https://CRAN.R-project.org/package=citr

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Bache, S. M., & Wickham, H. (2014). *Magrittr: A forward-pipe operator for r.* Retrieved from https://CRAN.R-project.org/package=magrittr

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*(4), 577–609; discussion 610–60. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1693222{\&}tool=pmcentrez{

Barsalou, L. W. (2016). On staying grounded and avoiding Quixotic dead ends. *Psychonomic Bulletin & Review*, *23*, 1122–1142. doi:10.3758/s13423-016-1028-3

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In A. M. Glenberg & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 245–283). Oxford, England: Oxford University Press. Retrieved from http://psychology.emory.edu/cognition/barsalou/papers/Barsalou{_}et{_}al{_}chap{

Bartoń, K. (2016). *MuMIn: Multi-Model Inference*. Retrieved from https://cran.r-project.org/package=MuMIn

Bartoń, K. (2017). *MuMIn: Multi-model inference*. Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., & Maechler, M. (2017). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bonner, M. F., & Grossman, M. (2012). Gray matter density of auditory association cortex relates to knowledge of sound concepts in primary progressive aphasia. *Journal of Neuroscience*, *32*(23), 7986–7991. doi:10.1523/JNEUROSCI.6241-11.2012

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., & Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, *46*(2), 743–756.

doi:10.1016/j.neuropsychologia.2007.10.007

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*(1), 193–216. doi:10.1037/0033-295X.112.1.193

Brants, T., & Franz, A. (2006). Web 1T 5-gram Version 1 LDC2006T13. DVD. Philadelphia: Linguistic Data Consortium.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cognitive Brain Research*, *24*(3), 355–63. doi:10.1016/j.cogbrainres.2005.02.020

Buckner, R. L., & Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. *Trends in Cognitive Sciences*, *17*(12), 648–665. doi:10.1016/j.tics.2013.09.017

Campbell, S. J., & Raney, G. E. (2015). A 25-year replication of Katz et al.'s (1988) metaphor norms. *Behavior Research Methods*, *48*(1), 330–340. doi:10.3758/s13428-015-0575-2

Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods*, *42*(3), 651–64. doi:10.3758/BRM.42.3.651

Connell, L. (2007). Representing object colour in language comprehension. *Cognition*, *102*(3), 476–85. doi:10.1016/j.cognition.2006.02.009

Connell, L., & Lynott, D. (2012). Principles of representation: Why you can't represent the same concept twice. *Under Review*, *526*. Retrieved from http://scholar.google.com/scholar?hl=en{\&}btnG=Search{\&}q=intitle:Principles+of+R

Connell, L., & Lynott, D. (2013). Flexible and fast: linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, *20*(3), 542–50. doi:10.3758/s13423-012-0368-x

Connell, L., & Lynott, D. (2016). Do we know what we're simulating? Information loss on transferring unconscious perceptual simulation to conscious imagery. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(8), 1218–32. doi:10.1037/xlm0000245

Connell, L., Lynott, D., & Dreyer, F. (2012). A functional role for modality-specific perceptual systems in conceptual representations. *PloS One*, *7*(3), e33321. doi:10.1371/journal.pone.0033321

Dowle, M., & Srinivasan, A. (2017). *Data.table: Extension of 'data.frame'.* Retrieved from https://CRAN.R-project.org/package=data.table

Gagné, C. L., & Shoben, E. J. (1997). Influence of Thematic Relations on the Comprehension of Modifier–Noun Combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 71–87. doi:10.1037/0278-7393.23.1.71

Gelman, A., & Su, Y.-S. (2016). *Arm: Data analysis using regression and multilevel/hierarchical models*. Retrieved from https://CRAN.R-project.org/package=arm

Genz, A., & Bretz, F. (2009). *Computation of multivariate normal and t probabilities*. Heidelberg: Springer-Verlag.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183–206. Retrieved from http://www.tau.ac.il/{~}giorar/files/Giora{_}97{_}GradedSalienceHypothesis.pdf

Glenberg, A. M., & Gallese, V. (2012). Action-based language: a theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905–22.

doi:10.1016/j.cortex.2011.04.010

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., & Buccino, G. (2008). Processing abstract language modulates motor system activity. *The Quarterly Journal of Experimental Psychology*, *61*(6), 905–919. doi:10.1080/17470210701625550

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, *97*(1), 3–18. Retrieved from http://psycnet.apa.org/journals/rev/97/1/3/

Harrell Jr, F. E. (2018). *Rms: Regression modeling strategies*. Retrieved from https://CRAN.R-project.org/package=rms

Harrell Jr, F. E., Charles Dupont, & others. (2018). *Hmisc: Harrell miscellaneous*. Retrieved from https://CRAN.R-project.org/package=Hmisc

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, *41*(2), 301–307. doi:10.1016/S0896-6273(03)00838-9

Henry, L., & Wickham, H. (2017). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr

Hothorn, T. (2017). *TH.data: TH's data archive*. Retrieved from https://CRAN.R-project.org/package=TH.data

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Højsgaard, S., & Halekoh, U. (2016). *DoBy: Groupwise statistics, lsmeans, linear contrasts, utilities*. Retrieved from https://CRAN.R-project.org/package=doBy

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes Factors. *The Journal of Problem Solving*, *7*, 2–9. doi:10.7771/1932-6246.1167

Katsumata, S. (1980). *Kenkyusha's new dictionary of English collocations*. Kenkyusha. Retrieved from https://books.google.co.uk/books?id=7UBvswEACAAJ

Katz, A. N., Paivio, A., Marshark, M., & Clark, J. M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor and Symbolic Activity*, *3*(4), 191–214.

Kintsch, W., & Dijk, T. V. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. Retrieved from http://psycnet.apa.org/journals/rev/85/5/363/

Koenker, R., & Ng, P. (2017). *SparseM: Sparse linear algebra*. Retrieved from https://CRAN.R-project.org/package=SparseM

Kuhn, M., Steve Weston, Wing, J., Forester, J., & Thaler, T. (2016). *Contrast: A collection of contrast methods*. Retrieved from https://CRAN.R-project.org/package=contrast

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–40. doi:10.1037/0033-295X.104.2.211

Liu, P. Q., Connell, L., & Lynott, D. (2018a). Can illness be bright? Metaphor comprehension depends on linguistic and embodied factors. *Manuscript*

Under Review, 1–54.

Liu, P. Q., Connell, L., & Lynott, D. (2018b). Switching costs between both perceptual and linguistic modalities. *Manuscript in Preparation*.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*(2), 273–302. doi:10.1111/j.1756-8765.2010.01106.x

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*(1), 96–104. doi:10.1016/j.cognition.2009.09.002

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41*(2), 558–64. doi:10.3758/BRM.41.2.558

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of bayes factors for common designs*. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Müller, K. (2017). *Bindrcpp: An 'rcpp' interface to active bindings*. Retrieved from https://CRAN.R-project.org/package=bindrcpp

Niedenthal, P. M. (2007). Embodying Emotion. *Science*, *316*(5827), 1002–1005. doi:10.1126/science.1136930

Nowosad, J. (2017). *Rcartocolor: 'CARTOColors' palettes*. Retrieved from https://CRAN.R-project.org/package=rcartocolor

Paivio, A., Yuille, J., & Madigan, S. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1), 1–25. Retrieved from http://psycnet.apa.org/journals/xge/76/1p2/1/

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2017). *nlme:*

Linear and nonlinear mixed effects models. Retrieved from https://CRAN.R-project.org/package=nlme

Plate, T., & Heiberger, R. (2016). *Abind: Combine multidimensional arrays.* Retrieved from https://CRAN.R-project.org/package=abind

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for mcmc. *R News*, *6*(1), 7–11. Retrieved from https://journal.r-project.org/archive/

R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality research.* Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Robinson, D. (2017). *Broom: Convert statistical analysis objects into tidy data frames.* Retrieved from https://CRAN.R-project.org/package=broom

Rundell, M. (2007). *Macmillan English Dictionary for Advanced Learners.* Macmillan. Retrieved from https://books.google.co.uk/books?id=OHLOLAAACAAJ

San Roque, L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M., … Majid, A. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics*, *26*(1), 31–60. doi:10.1515/cog-2014-0089

Sarkar, D. (2008). *Lattice: Multivariate data visualization with r.* New York: Springer. Retrieved from http://lmdvr.r-forge.r-project.org

Simmons, W. K., & Barsalou, L. W. (2003, May). The similarity-in-topography principle: Reconciling theories of conceptual deficits. doi:10.1080/02643290342000032

Solt, F., & Hu, Y. (2015). interplot: Plot the effects of variables in interaction terms. Available at The Comprehensive R Archive Network (CRAN). Retrieved from https://CRAN.R-project.org/package=interplot

Speed, L. J., & Majid, A. (2017). Dutch modality exclusivity norms: Simulating perceptual modality in space. *Behavior Research Methods*, *49*(6), 2204–2218. doi:10.3758/s13428-017-0852-3

Stevenson, A. (2010). *Oxford Dictionary of English*. OUP Oxford. Retrieved from https://books.google.co.uk/books?id=TaZZSAAACAAJ

Storms, G., & Wisniewski, E. J. (2005). Does the order of head noun and modifier explain response times in conceptual combination? *Memory & Cognition*, *33*(5), 852–61. doi:10.3758/BF03193080

Terry M. Therneau, & Patricia M. Grambsch. (2000). *Modeling survival data: Extending the Cox model*. New York: Springer.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth.). New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex*, *24*(7), 1767–1777. doi:10.1093/cercor/bht025

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. doi:10.3758/BF03194105

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag

CHAPTER 4

Embodied and Linguistic Components in Metaphor Processing

If metaphor comprehension relies on conceptual representations that combine both embodied simulation and linguistic distributional patterns, the two components of conceptual representations (i.e., embodied and linguistic components) should both be critical predictors of performance in metaphor processing. The work in this chapter aims to test just this hypothesis, that is to demonstrate that the embodied component (operationalised as ease of simulation in the previous chapter) and the linguistic component (operationalised as the n-gram co-occurrence frequencies of component words) can each account for a unique portion of variance in metaphor processing.

Furthermore, the work in this chapter presents the first stab at the linguistic shortcut hypothesis in this thesis. If the linguistic component is fast and superficial, it should be an ideal shortcut to be utilised during shallow processing. Thus, the empirical component of this chapter is composed of two contrasting experiments, one of a sensibility judgement task (relatively shallower processing) and another an interpretation generation task (relatively deeper processing). The difference in the depth of processing demanded by the task should lead to heavier reliance on the linguistic component in the shallow task, and heavier reliance on the embodied component in the deep task. The paradigm set by the empirical study of this chapter will be reproduced in the studies of Chapters 5 and 7.

An abridged version of this work has been published as conference proceeding in: Liu, P. Q., Connell, L., & Lynott, D. (2017). Can illness be bright? Metaphor comprehension depends on linguistic and embodied factors. In G. Gunzelmann, A. Hows, T. Tenbrink, & E. J. Davelaar (Eds.), 39th Annual Conference of Cognitive Science Society (pp. 2604–2609). Austin, TX: Cognitive Science Society.

Draft was submitted to Cognitive Science in October 2017 and was produced in collaboration with Dr Louise Connell and Dr Dermot Lynott as co-authors.

Paper 2: Can illness be bright? Metaphor comprehension depends on linguistic and embodied factors

Page numbers: 100 - 152

Can illness be bright? Metaphor comprehension depends on linguistic and embodied factors

Pei Q. Liu[1] & Louise Connell[1] & Dermot Lynott[1]

[1] Department of Psychology, Lancaster University

Author Note

Department of Psychology, Lancaster University, UK.

Correspondence concerning this article should be addressed to Pei Q. Liu, Department of Psychology, Fylde College, Lancaster University, Bailrigg, LA1 4HX, UK. E-mail: p.liu1@lancaster.ac.uk

Abstract

Conceptual representation is both linguistic and embodied, at least when processing literal language. Here, we aim to demonstrate that metaphor processing is no exception. We investigated how the linguistic component (i.e., linguistic distributional frequency: the co-occurrence frequency of the constituent words in a metaphor) and the embodied component (i.e., how easy it is to simulate the concept: operationalized in a previous norming study) affect metaphor comprehension. Results show that ease of simulation contributes to metaphor processing in both shallow sensibility judgement and deep interpretation generation tasks, but has a larger effect on response times in deeper processing. Linguistic distributional frequency also contributes to metaphor processing in both tasks, but affects the speed of response only in shallow sensibility judgement. Specifically, distributional frequency acts as a linguistic shortcut, both to speed up responses to accept metaphors as sensible when the frequency is high, and to flag up potentially unsuccessful processing when it is low. Overall, these results support embodied views that emphasize the importance of both linguistic and embodied components according to task goals.

Keywords: metaphor processing, embodied simulation, linguistic distributional pattern, linguistic shortcut hypothesis

Word count: 9616 words

Can illness be bright? Metaphor comprehension depends on linguistic and embodied factors

Metaphors add warmth and color to language, and can be understood so effortlessly by language users that their use does not register as an artifice. In linguistic terms, a metaphor is a figure of speech where a word or a phrase (the source) is applied to an object or an action (the target) to which it cannot be literally applied. For example, we can easily comprehend and frequently use the expression *bright students* even though a student is not an object to which the visual property of bright can be applied. Such metaphors are an essential part of language, not least in assisting us to express abstract concepts such as time, emotions and relationships (e.g., Boroditsky, 2001; Chan, Tong, Tan, & Koh, 2013; Lakoff & Johnson, 1999).

Traditionally, much research on metaphor comprehension has focused on identifying particular factors that could make a metaphor easier – and thus faster – to understand, such as familiarity, conventionality and aptness (Blasko & Connine, 1993; Bowdle & Gentner, 2005; Cardillo, Schmidt, Kranjec, & Chatterjee, 2010; Giora, 1997; L. L. Jones & Estes, 2006; Pierce & Chiappe, 2008; Roncero & Almeida, 2014b). However, as we shall discuss below, these factors suffer from problems of theoretical specificity and empirical operationalization that make it difficult to draw clear conclusions about their roles in metaphor comprehension. Thus, it is necessary to study metaphor processing under a theoretical construct with better operationalized factors. In this paper, we look at the embodied/grounded perspective of conceptual representation, which suggests that language comprehension is an interactive conglomerate of activating linguistic symbols and simulating physical experiences. We suggest that metaphor processing should be viewed as an aspect of language processing and propose how these two components affect metaphor processing and investigate their roles with reference to how they relate to depth of processing during comprehension.

Traditional Views of Metaphor Processing

Several factors are often studied in research on metaphor processing, namely familiarity, conventionality and aptness, all of which have been demonstrated to affect the speed of metaphor comprehension. The graded salience hypothesis suggests that familiar or conventional metaphors are processed automatically because they enjoy high salience, compared with unfamiliar or novel metaphors (Blasko & Connine, 1993; Giora, 1997, 2007; Glucksberg, Gildea, & Bookin, 1982; Laurent, Denhières, Passerieux, Iakimova, & Hardy-Baylé, 2006; Peleg, Giora, & Fein, 2001). Moreover, conventionality could be a crucial factor in determining the mechanism with which metaphor is processed, according to an influential theory called the career of metaphor (Bowdle & Gentner, 2005). In this theory, metaphors (e.g., *My lawyer is a shark*) first start out as novel or unconventional constructions and are processed by comparing the source concept (e.g., *shark*) with the target (e.g., *lawyer*) and identifying potential similarities to form the basis of the metaphoric meaning (e.g., *aggressive*). As a metaphor is used more often, it becomes conventionalized (e.g., aggressive becomes a conventional attribute for "shark") and the metaphor is processed by categorizing the target as a member of this conventional category (e.g., *shark* → *aggressive beings*). In contrast, proponents of the theory of class inclusion suggest that it is aptness, rather than conventionality, that decides whether a metaphor is processed by categorization (Glucksberg & Haught, 2006b; Haught, 2014; L. L. Jones & Estes, 2005). That is, *shark* can be categorized as a member of "aggressive beings" because the metaphor is apt (i.e., it works well), but a less apt metaphor would require comparing the target and source to find similarities.

Although these factors have been shown to affect metaphor processing to a certain degree, when it comes to understanding exactly how these three factors affect the process of metaphor comprehension, the problems are two-fold: first in their theoretical specificity, and second in their operationalization. In terms of theoretical specificity, these three factors are frequently confounded with one another and are internally inconsistent. Although some have suggested that familiarity and conventionality are two different theoretical constructs (Blasko & Connine, 1993; Bowdle & Gentner, 2005; Glucksberg et al., 1982; L. L. Jones & Estes, 2005, 2006; Roncero & Almeida, 2014a), they are often used interchangeably in the literature (Cardillo, Watson, Schmidt, Kranjec, & Chatterjee, 2012; Giora, 1997; Giora & Fein, 1999; Glucksberg & Haught, 2006b; Lai & Curran, 2013). Both terms are used to describe metaphors that at one extreme are entirely novel and have never been encountered before by participants, while at the other extreme are very common, conventional, and familiar to participants. Furthermore, aptness can also be confounded with familiarity and conventionality. Not only are familiarity and conventionality both dependent on usage patterns of a metaphor across a language, but aptness itself can also be associated with usage patterns. Only apt metaphors are likely to become conventionalized or familiar, as a metaphor that does not work well is unlikely to become widely used by speakers of a language. Hence, familiar/conventional metaphors are typically apt, whereas novel metaphors might be either apt or not. There is a core dependency between the factors that is not trivial to disentangle. Indeed, ratings of aptness and familiarity are highly correlated (r = .73-.82: Katz, Paivio, Marshark, & Clark, 1988; Roncero & Almeida, 2014a), as are ratings of aptness and corpus frequency counts of the metaphoric expression (r = .41-.57: Roncero & Almeida, 2014b; Thibodeau & Durgin, 2011). This core dependency means that we cannot investigate how these factors jointly or independently affect metaphor processing.

Moreover, the definitions of these critical factors are internally inconsistent. It is ambiguous whether familiarity/conventionality describe a metaphoric expression, or its metaphoric meaning. For example, familiarity is sometimes assumed to refer to how often people have encountered the metaphoric expression itself (e.g., how often is *bright* used to describe *students*? Blasko & Connine, 1993; Bowdle & Gentner, 2005; Cardillo et al., 2012; Glucksberg, 2003; Mashal, 2013), and sometimes to how accustomed people are to relating the expression to its metaphoric meaning (e.g., how often is *bright* understood to mean *intelligent* and *quick-witted*: Campbell & Raney, 2015; Cardillo et al., 2010; Giora, 1997; Giora & Fein, 1999; Katz et al., 1988; Peleg et al., 2001; Roncero & Almeida, 2014a), but these are two very different and dissociable theoretical constructs. A particular linguistic expression might be encountered reasonably often but remain poorly understood (e.g., *hard Brexit* may have high expression familiarity but low meaning familiarity), or a metaphoric meaning might be encountered reasonably often via a different expression to the one supplied (e.g., *bright solutions* may have high meaning familiarity but low expression familiarity).

Because of the inconsistency in definition, these factors are not always consistently or rigorously operationalized in empirical studies that seek to examine their influence on metaphor comprehension. Typically, researchers measure each factor by asking participants to rate metaphors on a particular scale. Participants providing familiarity ratings are sometimes asked to rate the extent to which they have read or heard the expression in the past (Mashal, Faust, & Hendler, 2005; Roncero & Almeida, 2014a), sometimes to rate the familiarity of the ideas expressed in the metaphors (Blasko & Connine, 1993; Campbell & Raney, 2015; Katz et al., 1988), and at other times to rate the frequency of experience with both the expression and the meaning together (Cardillo et al., 2010). It is therefore unclear whether the reported effects of familiarity on metaphor processing times reflect experience at the word level (i.e., similar to word or phrase frequency) or experience with understanding the metaphoric meaning in question (i.e., experience in representing the metaphoric concept).

The same problem exists with the definition and operationalization of conventionality. In studies that test its empirical effects on metaphor processing, conventionality is sometimes operationalized as a function of the entire metaphoric expression by actually asking participants to rate familiarity rather than conventionality (e.g.: how familiar is the metaphor? Cardillo et al., 2010, 2012; Faust & Mashal, 2007; Lai, Curran, & Menn, 2009; Mashal et al., 2005), but sometimes as how often the metaphoric meaning is associated with the source (e.g., how conventional is it to use *bright* to mean *intelligent*?: Bowdle & Gentner, 2005; Gentner & Wolff, 1997; L. L. Jones & Estes, 2006; Roncero & Almeida, 2014a). While the former is confounded with familiarity, the latter operationalization has been challenged by findings that conventionality is context-dependent and cannot be defined for the source term independent of the target (Thibodeau & Durgin, 2011). That is, just because *Students can be bright* is a conventional metaphor does not mean all metaphors that use bright as a source domain for *intelligent* and *quick-witted* (e.g., *Solutions can be bright*) are equally conventional. In short, it is unclear whether conventionality is a theoretically valid construct, and whether the reported effects of conventionality on metaphor processing are truly distinct from familiarity (which has its own problems, as discussed above).

Investigation of the importance of aptness faces a similar challenge. It has been ambiguously defined as either a very general, high-level quality or goodness of a metaphor and is often operationalized as such (Blasko & Connine, 1993; Bowdle & Gentner, 2005; Campbell & Raney, 2015; Glucksberg & Haught, 2006b; Katz et al., 1988), or as a much more low-level specification of how well the metaphoric meaning (e.g., *intelligent* and *quick-witted*) fits or overlaps with the target (e.g., *students*:

Chiappe & Kennedy, 1999; L. L. Jones & Estes, 2005, 2006; Pierce & Chiappe, 2008; Roncero & Almeida, 2014b). Thibodeau and Durgin (2011), in a detailed examination of aptness ratings, argue that people are influenced by extraneous factors such as processing fluency when rating aptness and hence these ratings do not capture the theoretical construct they are meant to operationalize.

In summary, familiarity, conventionality and aptness have all been shown to affect metaphor processing. However, they have several theoretical and operational problems that mean they have limited utility in enhancing our understanding of what makes a metaphor easier to process. Rather than continue to vary and refine how these factors are conceptualized, we propose that a different approach is needed to seek clearer predictors of metaphor processing that (a) are theoretically and operationally distinct, and (b) are able to independently account for speed and accuracy performance in metaphor processing.

Grounded Views of Language Processing

Research in conceptual representation and language processing has tended to operate in parallel to that of traditional metaphor processing, and therefore takes quite a different perspective on how access to meaning takes place. Essentially, two components are employed in the mental representation of meaning when people process language (Barsalou, Santos, Simmons, & Wilson, 2008; Borghi & Binkofski, 2014; Connell & Lynott, 2014b; Louwerse & Jeuniaux, 2008; Lynott & Connell, 2010; Vigliocco, Meteyard, Andrews, & Kousta, 2009). The first component is linguistic and relies on the statistical, distributional pattern of how words co-occur across contexts. For example, the words *bright* and *students* occur together more often in language than do *bright* and *illness*, and such distributional frequency patterns have been shown to be powerful enough to capture many aspects of language processing, such as priming and reading times (Vigliocco et al., 2009), language acquisition from early infancy (Aslin, Saffran, & Newport, 1998; Kirkham, Slemmer, & Johnson, 2002), and even to some extent distinguishing words that relate to different perceptual modalities (Louwerse & Connell, 2011). The second component of conceptual representation is embodied (also known as the grounded, sensorimotor or situated), which relies on the process of simulation; that is the partial reactivation of past perceptual, motor, affective, introspective and other experiences (Allport, 1985; Barsalou, 1999; Glenberg & Gallese, 2012; Niedenthal, 2007). Evidence for embodied simulation includes shared activation between brain areas involved in perceptual or action experience and their equivalents in language comprehension. For example, reading sound-related words like “thunder” activates the auditory association cortex, and their processing is selectively impaired in patients with atrophy of the auditory association cortex (Bonner & Grossman, 2012; Boulenger et al., 2008; Goldberg, Perfetti, & Schneider, 2006; Hauk, Johnsrude, & Pulvermüller, 2004).

Together, the linguistic and embodied components can explain language processing better than either alone (Andrews, Vigliocco, & Vinson, 2009; Johns & Jones, 2012; Louwerse & Connell, 2011). In particular, linguistic distributional information provides a powerful tool for superficial language processing because the linguistic component is faster than the embodied simulation component (Barsalou et al., 2008; Louwerse & Jeuniaux, 2008). In any language processing task, both linguistic and embodied components are activated but the activation of the linguistic component will peak before that of the embodied component (Barsalou et al., 2008). Therefore, Louwerse and Jeuniaux (2008) argued that people are more likely to rely on the embodied component when deeper processing is specifically cued in the task (Connell & Lynott, 2014b); but people will be reliant upon the linguistic component to generate a good-enough approximation (Ferreira, Bailey, & Ferraro, 2002) when shallow processing can suffice. Evidence for a speed advantage of the linguistic over the embodied component has been found in behavioral (Louwerse & Connell, 2011; Santos, Chaigneau, Simmons, & Barsalou, 2011), electrophysiological (Louwerse & Hutchinson, 2012), and neuroimaging (Simmons et al., 2007) studies.

In line with these arguments, Connell and Lynott (2013) proposed that information from the linguistic component could act as a cognitive triage mechanism during language processing by providing a guide to whether it is worth expending effort on relatively costly embodied simulation (Connell, 2018). Since the co-occurrence of words in language tends to reflect the associations of objects, events, and ideas in the real world, the linguistic component can provide a rough and ready approximation of whether the embodied component will actually succeed in simulating a sensible meaning. Moreover, since activation in the linguistic component is faster than in the embodied component, the linguistic component has the potential to provide a shortcut by guiding task processing before the embodied component is fully engaged. If the linguistic component indicates that future processing is likely to fail or entail unnecessary costs (e.g., the words rarely co-occur in the same context and so their combined meaning might not be simulated successfully), then it could abandon the processing before any more cognitive effort is expended by the embodied component. On the other hand, if the linguistic component indicates that future processing is likely to succeed relatively easily (e.g., the words often co-occur in the same context and so their combined meaning can probably be simulated successfully), then it could either inform a response immediately (i.e., based on the linguistic shortcut alone) or allow the embodied component to continue developing a detailed simulation of meaning. Although the linguistic shortcut can provide a cognitive triage mechanism in any sort of cognitive task, it is more likely to form the basis of responses in circumstances where a detailed response is not required and a good-enough heuristic will therefore suffice, such as tasks that allow relatively shallow or superficial processing, or individual trials where deep and detailed processing has been abandoned.

To test this proposal, Connell and Lynott (2013) asked participants to perform two tasks with novel noun-noun compounds (e.g., cactus beetle): a shallow processing task where participants judged if the compounds made sense, or a deep processing task where they decided if they could generate specific interpretations for them. The results showed that, consistent with linguistic shortcut predictions, the linguistic distributional frequency of the constituent words (i.e., how often cactus and beetle co-occurred) predicted both the likelihood of accepting versus rejecting a compound as sensible and the time course of both successful and unsuccessful processing in the shallow sensibility judgement task. However, in the deeper interpretation generation task, distributional frequency only predicted the likelihood of accepting versus rejecting a compound as interpretable and the timecourse of unsuccessful (i.e., abandoned) processing, but not the timecourse of successful processing that resulted in a detailed interpretation. In other words, while rejecting a noun-noun compound as nonsensical or uninterpretable needed only the linguistic shortcut to quickly flag up unsuccessful processing, the time needed to accept a compound as interpretable may require more detailed simulation which cannot be captured by the linguistic distributional frequency.

Present Study

Connell and Lynott's proposal can be applied directly to the study of metaphor processing, because the interplay of the linguistic and embodied components, and the role of the linguistic shortcut as a cognitive triage mechanism, operates in theory across all types of language comprehension. However, Connell and Lynott's study was unable to directly investigate the interplay between linguistic distribution and embodied simulation during language processing because it lacked at the time a direct measure of successful simulation (i.e. the ability to mentally represent the meaning of the phrase or sentence). The current study also faces the same obstacle in operationalizing how easy it is to generate successful simulation, because the meaning of a metaphor, by definition, cannot be accessed directly by looking at the literal meaning of its constituent words (Liu, Connell, & Lynott, 2018). Furthermore, because simulations are unconscious mental representations that lose information when brought to conscious awareness (Connell & Lynott, 2016), the ease of simulation cannot be directly rated as, for example, the ease of understanding a metaphor or of generating mental imagery. Therefore, in the present study, we used a new norming metric by Liu et al. (2018) that was explicitly designed to quantify how easy people find it to arrive at a mental representation of metaphoric meaning. It proposed three indirect measures of successful metaphor processing that, together, offer a proxy for ease of simulation: sensibility (How much sense does the sentence make if you read it in a story or heard it in conversation?), usability (How easy it would be for you to use the sentence in writing or in conversation?), and imaginability (How easy it is for you to imagine the concept described in the sentence?). Using principle components analysis, these three ratings scales were combined into a single ease of simulation measure that captured their common variance.

The ease-of-simulation measure (EoS) is agnostic towards the content of the simulated representation. Instead, it measures the end-product of simulation – a complete and coherent representation of meaning – and uses it to quantify the effort of comprehension. If people find it easy to make sense of, use and imagine a metaphor, they would necessarily find the concept easy to simulate as well. Indeed, this ease-of-simulation measure can not only capture the majority of the variance (91%) of the original subscales but also predict metaphor processing time better than any of the individual subscales (Liu et al., 2018), supporting the idea that it captures the underlying mechanism of conceptual representation. Thus, we expect it to be a powerful predictor of both speed and accuracy of metaphor processing in the current study as well.

Moreover, the grounded views of conceptual representation also predict that the linguistic component (i.e., linguistic distributional frequency (LDF) calculated from a large corpus) would be able to account for metaphor processing independently of the embodied component (i.e., ease of simulation). Although it remains a contentious issue how much the linguistic component contributes to the representation of meaning, it could be suggested that these two components are of distinct natures and perform different functions (Barsalou et al., 2008; Connell, 2018; Louwerse, Hutchinson, Tillman, & Recchia, 2015; Riordan & Jones, 2011). If our study can show that ease of simulation and LDF can each explain a unique, distinct portion of variance in performance of metaphor processing, this should provide strong support for the grounded approach of conceptual representation. Hence, we operationalised LDF in the current paper as the co-occurrence frequency of constituent words within a small 5-word window (i.e., the LDF of *Students can be bright* is the co-occurence frequency of *students* and *bright* with zero to three words in between). By keeping the operational definition simplistic, any effect of LDF could be counted as strong evidence for the existence of the linguistic component beyond embodied simulation.

Furthermore, as the cognitive triage mechanism suggests, the effects of these two factors should vary according to required depth of processing predicated by the tasks. Thus, in two experiments, we asked participants to process metaphors that systematically varied in LDF and EoS in a shallow task (Experiment 1: sensibility judgement task) or in a deep task (Experiment 2: interpretation generation task). Since the role of the linguistic and embodied components in language comprehension varies according to available resources, task demands, and processing goals (Connell & Lynott, 2014b), their roles in metaphor processing will vary across experiments.

Based on the operation of each component (Barsalou et al., 2008; Louwerse & Connell, 2011; Louwerse & Jeuniaux, 2008), and the operation of the linguistic shortcut mechanism (Connell, 2018; Connell & Lynott, 2013), we propose the following hypotheses:

1. Simulation hypothesis: People will rely on the embodied component in both tasks. A metaphor that is easy to simulate will be accepted as sensible or interpretable more often. The effect of EoS on response time will be in opposite directions for "yes" and "no" responses. That is, it will take people less time to accept a metaphor that is easy to simulate, and less time to reject a metaphor (as nonsensical or uninterpretable) that is difficult to simulate (i.e., the harder a metaphor is to simulate, the quicker people will give up trying to process it). However, EoS should have a larger effect on accept ("yes") than reject ("no") response times because the latter reflects abandoned processing in the embodied component.

2. Linguistic distribution and linguistic shortcut hypothesis: People will also rely on the linguistic component in both tasks, where the linguistic shortcut will be used as a cognitive triage mechanism to identify processing that is unlikely to succeed and should be abandoned to avoid unnecessary costs. A metaphor with high LDF will be more likely to be accepted as sensible or interpretable than a metaphor with lower LDF (i.e., the less often words co-occur in language, the more likely people will be to abandon processing rather than engage in further, costly attempts at simulation). Furthermore, similar to the simulation hypothesis, the effects will be in opposite directions for "yes" and "no" responses: it will take people less time to accept a metaphor that has high LDF, and less time to reject a metaphor that has low LDF.

3. Depth of processing hypothesis: People will rely on EoS more for

interpretation generation than for sensibility judgement because the former requires deeper processing and more detailed simulation. In contrast, people will rely on LDF more in shallow sensibility judgement, where "good enough" processing will suffice, than in deep interpretation generation.

In both experiments, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Experiment 1: Sensibility Judgement Task

In this study, participants were encouraged to process metaphors in a shallow way by judging whether or not the sentences made sense, where they made a forced choice between "yes" and "no". Sensibility judgement can be considered a relatively shallow form of language processing because it involves only a single yes/no response and participants are not required to specify any further details of how or why the sentence makes sense (Connell & Lynott, 2013; Lynott & Connell, 2010).

Method

Participants. Twenty-eight participants (five male and 23 female) took part in the study, same as Connell and Lynott (2013), all of whom were students at Lancaster University and native speakers of English with mean age of 19.1 years (*SD* = 1.1). Participation took approximately 20 minutes in exchange for £3.00 or course credits.

Materials. We used a total of 452 metaphoric sentences taken from Liu et al., (2016; see Table 1 for examples[1]). All sentences took the form "Noun can be

---

[1] All data and analysis scripts can be viewed through https://goo.gl/k1Q4jq. We used R (Version 3.5.0; R Core Team, 2017) and the R-packages *abind* (Version 1.4.5; Plate & Heiberger, 2016), *arm* (Version 1.10.1; Gelman & Su, 2016), *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2018),

adjective" (e.g., *Student can be bright*), and were composed of 113 perceptual adjectives (e.g., *bright*: Lynott & Connell, 2009), each paired with four nouns (e.g., *students*/*solutions*/*supply*/*illness*) that were capable of eliciting metaphoric (i.e., non-literal) meanings, which varied systematically along the two critical dimensions of EoS and LDF.

Ease of simulation (EoS) scores for each sentence ranged from easy to difficult ($M = 0.00$, $SD = 1.00$) and were taken from the norming study by Liu et al. (2017), where higher scores represent greater EoS. For example, both *Students can be bright* and *Solutions can be bright* were considered relatively easy to simulate, whereas both Illness can be bright and Supply can be bright were considered relatively difficult to simulate.

Linguistic distributional frequency (LDF) for each sentence ranged from low to high ($M = 2.95$, $SD = 0.97$), and was calculated as the log of the summed bi- to five-gram frequencies of the sentence's noun and adjective in the Google Web1T Corpus (Brants & Franz, 2006). To take the metaphor *Students can be bright* as an example, the LDF was the sum of the frequencies of "student … bright" and "bright … student" with zero, one, two, and three intervening words. Frequencies were log

bookdown (Version 0.7; Xie, 2016), *broom* (Version 0.4.4; Robinson, 2017), *citr* (Version 0.2.0; Aust, 2016), *coda* (Version 0.19.1; Plummer, Best, Cowles, & Vines, 2006), *data.table* (Version 1.11.4; Dowle & Srinivasan, 2017), *ggplot2* (Version 2.2.1; Wickham, 2009), *interplot* (Version 0.1.5; Solt & Hu, 2015), *knitr* (Version 1.20; Xie, 2015), *lme4* (Version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.0.1; Kuznetsova, Brockhoff, & Christensen, 2017), *magrittr* (Version 1.5; Bache & Wickham, 2014), *MASS* (Version 7.3.50; Venables & Ripley, 2002), *Matrix* (Version 1.2.14; Bates & Maechler, 2017), *MuMIn* (Version 1.40.4; Bartoń, 2017), *nlme* (Version 3.1.137; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017), *papaja* (Version 0.1.0.9735; Aust & Barth, 2017), *psych* (Version 1.8.4; Revelle, 2017), *purrr* (Version 0.2.5; Henry & Wickham, 2017), and *rcartocolor* (Version 0.0.22; Nowosad, 2017) for all our analyses and the writing up of this manuscript.

Table 1

Sample metaphors, and their scores for EoS and LDF

Metaphor	EoS	LDF
Illness can be bright.	-1.32	2.95
Supply can be bright.	-1.02	3.72
Solutions can be bright.	1.41	3.11
Students can be bright.	1.84	4.08
Minutes can be lukewarm.	-1.49	3.31
Scheme can be lukewarm.	-0.78	1.61
Supporters can be lukewarm.	0.69	2.69
Reaction can be lukewarm.	1.03	3.45

transformed as $\log_{10}$(frequency + 41) because 41 is the lowest non-zero frequency in the item set (i.e., the raw frequency of 0 in fact means that the constituent words co-occurred 41 times; Connell & Lynott, 2013). Within the four metaphors created for each adjective, LDF varied independently of EoS (see Table 1). For example, both *Students can be bright* and *Supply can be bright* had relatively high LDF, whereas both *Illness can be bright* and *Solutions can be bright* had lower LDF.

The sentences were split into four lists of 113 items each, where each adjective appeared only once per list, and the distribution of easy/difficult to simulate and high/low LDF was approximately equal across lists (EoS: $F_{(3,440)} = 1.70$, $p = .166$; LDF : $F_{(3,440)} = 0.43$, $p = .734$). Each participant saw only one list.

Procedure. There were five practice trials (whose results were not recorded or analyzed) before the 113 test trials. Participants read one sentence in each trial and decide whether or not the sentence made sense, based on oral and verbal instructions. All trials had the same structure (see Figure 1). Participants first saw a fixation cross for 1000 milliseconds (ms), followed by the noun for 500 ms, followed by the phrase “can be” for 500 ms, and then followed by the adjective. The adjective remained on the screen until participants made a response. Participants pressed either the comma key (“,”) if they judged that the sentence made sense; or the full stop key (“.”) if they judged that it did not make sense. They were allowed to make their response naturally, using the finger(s) most comfortable for them. The response could be made without a time limit; and participants were told explicitly that there were no right or wrong answers to the question. Both the response decision (“yes” to accept the metaphor as sensible; or “no” to reject the metaphor as nonsensical), and response time in milliseconds (RT) from onset of the adjective, were recorded as dependent variables.

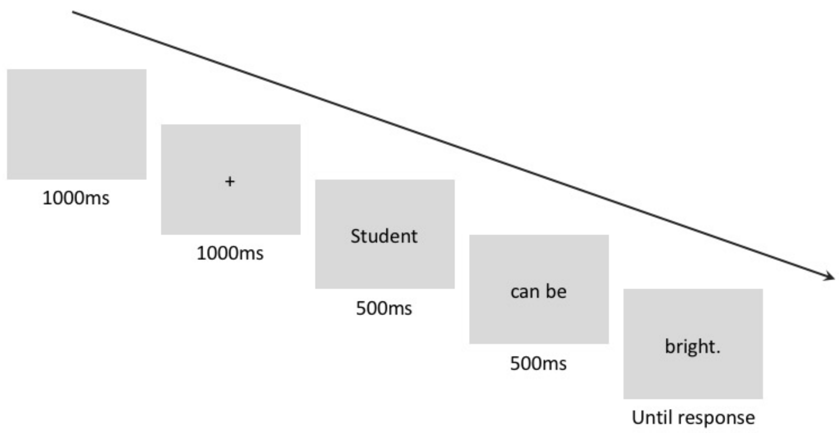


Figure 1. Trial structure in the sensibility judgement task.

Design and analysis. Response decisions were analyzed in a mixed effects logistic regression (binomial distribution with logit link), with the dependent variable of response (coded as 1 for “yes”, accepting the metaphor as sensible; and 0 for “no”, rejecting the metaphor as nonsensical), crossed random-intercept effects of participants and items,[2] and fixed variables of LDF, EoS, and their interaction. Both LDF and EoS scores were centred before analysis.

Response times (RTs) were analyzed using mixed effects linear regressions, firstly in an omnibus analysis with crossed random-intercept effects of participants and items,[3] and fixed variables of response, LDF, EoS, and their interactions. Secondly, we ran separate analyses on acceptance (“yes”) and rejection (“no”) responses because we expected the effects of the fixed variables to be in opposite directions for different response decisions. We initially considered the inclusion of word length and lexical frequency (i.e., sum of the log frequencies of nouns and adjectives, retrieved from Elexicon separately: Balota et al., 2007, http://elexicon.wustl.edu), but they were excluded from final analyses because their zero-order correlations with RT were near zero (see Supplementary Materials: https://goo.gl/trTu18).

Results and Discussion

All participants had mean response times within 3SD of the overall mean and so all were included in analysis. Two trials were removed because of motor error (RT

[2] The inclusion of participants as a random factor improved model fit above the empty model, $\chi^2_{(1)} = 427.39$, $p < .001$, as did the inclusion of items as a crossed random factor above the participants-only model, $\chi^2_{(1)} = 456.76$, $p < .001$.

[3] The inclusion of participants as a random factor improved model fit above the empty model, $\chi^2_{(1)} = 853.15$, $p < .001$. The inclusion of items as a crossed random factor did not improve model fit above the participants-only model $\chi^2_{(1)} = 0.00$, $p = .998$. However, in order to keep the models consistent, both participants and items were included as crossed random factors in the linear models of RT as well as in the logistic models of response decision. It should be noted that analyses of RT with participant-only models did not substantially change results; these analyses are reported in supplementary materials.

< 200ms). Furthermore, individual trials with RT more than three standard deviations from each participant's mean per response decision were removed as outliers: 1.33% of "yes" responses and 2.20% of "no" responses.

Response decision. Among 3105 valid trials, 1413 (45.51%) were accepted as sensible ("yes" responses) and 1692 (54.59%) were rejected as nonsensical ("no" responses). As predicted, logistic regression showed that EoS had a positive effect on the response decision, $z_{(1,3101)} = 24.60$, $p < .001$, $b = 1.253$, 95% CI = [1.153, 1.353]. As EoS increased by one unit, the odds of accepting a metaphor as sensible (as opposed to rejecting it as nonsensical) increased 3.501 times. In contrast, LDF had a negative effect on the response decision, $z_{(1,3101)} = -2.02$, $p = .043$, $b = -0.093$, 95% CI = [-0.183, 0.003], meaning that each increased unit of LDF made the decision to accept a metaphor as sensible less likely (odds ratio = 0.911). Since this partialled relationship between response decision and LDF in the regression model was the inverse of their unpartialled relationship outside the regression model (i.e., mean LDF was higher for "yes" than "no" responses: Supplementary Materials), it suggests that net suppression was present in the model.[4] That is, the shared error variance between LDF and EoS was effectively hiding the real relationship between LDF and response decision, such that the net effect of LDF in logistic regression was to enhance the effect of EoS by suppressing its unhelpful error variance (J. Cohen, Cohen, West, & Aiken, 2003).

In order to establish the true relationships between response decision and our independent variables, we therefore removed the shared variance between LDF and EoS (currently correlated at $r = .27$) by orthogonalizing the variables. We used principal components analysis (PCA) to rotate the two original variables into two

[4] A logistic regression model with linguistic associative frequency as the only fixed factor confirmed this possibility as it returns a positive coefficient, $t_{(1,3103)} = 4.03$, $p < .001$, $b = 0.36$, 95% CI = 0.19 - 0.54.

Table 2

Logistic mixed-effect regression of response decision in Experiment 1

	b	95% CI	p
Intercept	-0.25	-0.62 - 0.13	.198
EoS	1.47	1.34 - 1.60	$<$ .001
LDF	0.11	0.01 - 0.21	.027
EoS x LDF	-0.00	-0.11 - 0.10	.946

Note. b is non-standardised regression coefficient.

Both predictors orthogonalized.

orthogonal components, which (rather than the traditional PCA use of dimension reduction) allowed us to disentangle two correlated variables without losing any information (Connell & Lynott, 2014a; Glantz & Slinker, 1990). Using varimax rotation with Kaisar normalization on a covariance matrix, Component 1 corresponded to EoS (r = .99) and Component 2 corresponded to LDF (r = .99). These two orthogonalized variables accounted for 100% of the original variance while themselves were uncorrelated (i.e., r = .00). We re-ran the logistic regression with these orthogonalized variables and obtained results as follows (Table 2).

Logistic regression with these orthogonalized variables showed that EoS had a positive effect on response decision (Table 2). As the orthogonalized EoS increased by one unit, the odds of accepting a metaphor as sensible increased 3.421 times. Moreover, the effect of the orthogonalized LDF was also positive, but only as a weak trend. This time, as the orthogonalized LDF increased, the odds of accepting a metaphor as sensible increased slightly (1.084 times) as well.

Table 3

Omnibus linear mixed-effect model of RT in Experiment 1

	b	95% CI	*p*
Intercept	1129.66	1011.04 - 1248.29	< .001
EoS	67.06	41.76 - 92.36	< .001
LDF	14.76	-9.81 - 39.32	.239
Response	0.76	-36.86 - 38.38	.969
EoS x LDF	24.39	1.18 - 47.61	.040
EoS x response	-181.55	-218.91 - -144.17	< .001
LDF x response	-38.54	-74.48 - -2.61	0.036
EoS x LDF x response	-32.50	-67.07 - 2.08	0.066

Note. *b* is non-standardised regression coefficient.

Both predictors orthogonalized.

In other words, when people were asked to make a relatively shallow judgement about whether or not a metaphor made sense, their yes/no decision was influenced predominantly by EoS (easy-to-simulate metaphors were more likely to be judged as sensible) but also by the LDF of the words used in the metaphor (frequently co-occurring words were slightly more likely to be judged as sensible; infrequently co-occurring words were slightly more likely to be rejected as nonsense). Both the linguistic and embodied components played a role in metaphor processing.

Response time. As in the logistic regression of response decision, linear analysis of response times (M = 1139 ms, SD = 587 ms) also produced evidence of net suppression (see supplementary materials for details). For that reason, and to maintain the independence between the predictors in all our models, we orthogonalized the variables per response type. Table 3 shows full results of the

Table 4

Regression model on RT in Experiment 1 per response decision

	Yes			No		
	b	95% CI	*p*	*b*	95%CI	*p*
Intercept	1172.16	1051.79 - 1292.52	< .001	1155.41	1010.94 - 1299.88	< .001
EoS	-134.88	-161.98 - -107.79	< .001	72.28	47.27 - 97.29	< .001
LDF	-27.52	-53.10 - -1.93	0.036	17.72	-6.40 - 41.83	0.152
EoS x LDF	-13.00	-37.90 - 11.91	0.307	23.33	0.51 - 46.14	0.046

Note. *b* is non-standardised regression coefficient.

Both predictors orthogonalized per response decision.

omnibus analysis across all responses. Overall, EoS had a positive effect on RT, meaning that response times were generally slower for easy-to-simulate metaphors. However, EoS also interacted negatively with response decision, suggesting that the direction of the EoS effect differed by response type. LDF had no overall main effect, but interacted with response decision to indicate that the direction of LDF differed for "yes" and "no" RTs.

Since we had separate hypotheses for "yes" and "no" RTs, we divided the dataset by response decision and analyzed their RT separately. Results are given in Table 4. For "yes" responses (i.e. metaphors that were accepted as sensible; RT: M = 1150 ms, SD = 589 ms), EoS had the predicted negative effect, such that the easier a metaphor was to simulate, the less time people took to accept it as sensible Table 4. LDF also had a negative effect on acceptance times as predicted. That is, the more often the words in a metaphor co-occurred in language, the faster people were to accept it as sensible. Figure 2 shows the standardized coefficients of each variable on RT per response decision.

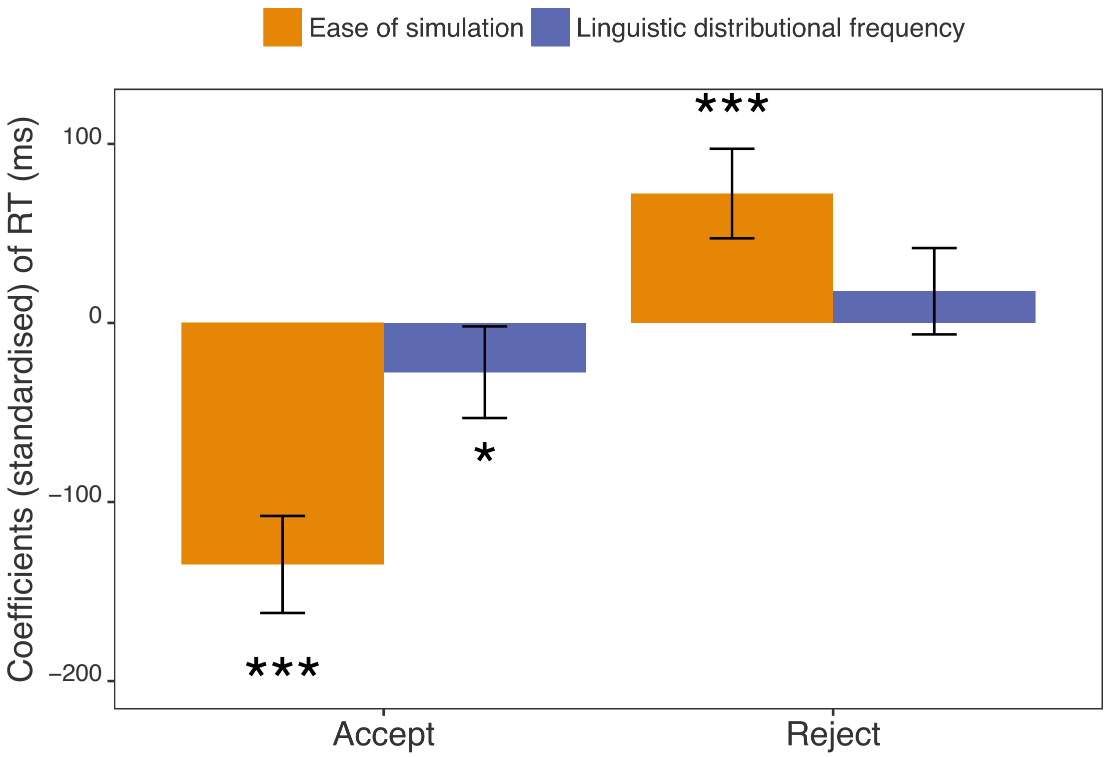


Figure 2. Effects of EoS and LDF, expressed as standardized regression coefficients, on acceptance and rejection times for sensibility judgement. Error bars represent 95% confidence intervals, and asterisks represent the significance of directional predictions where specified (*** $p < .001$, ** $p < .01$, * $p < .05$).

For "no" responses (metaphors rejected as nonsensical; RT: M = 1114 ms, SD = 603 ms), the effects ran in the opposite direction (Table 4). As predicted, EoS had a positive effect on RT, meaning that people were faster to reject metaphors that were normally regarded as difficult to simulate. LDF had no significant main effect but did positively interact with EoS, such that the effect of EoS was enhanced when LDF was high, but was reduced when LDF was low. Figure 3 shows the effects of EoS on the RT of reject ("no") responses as LDF increased. Overall, people were faster to reject a metaphor as nonsensical when its words rarely co-occurred and its meaning was difficult to simulate, and slower to reject metaphors whose words often co-occurred but whose meaning was generally considered easy to simulate.

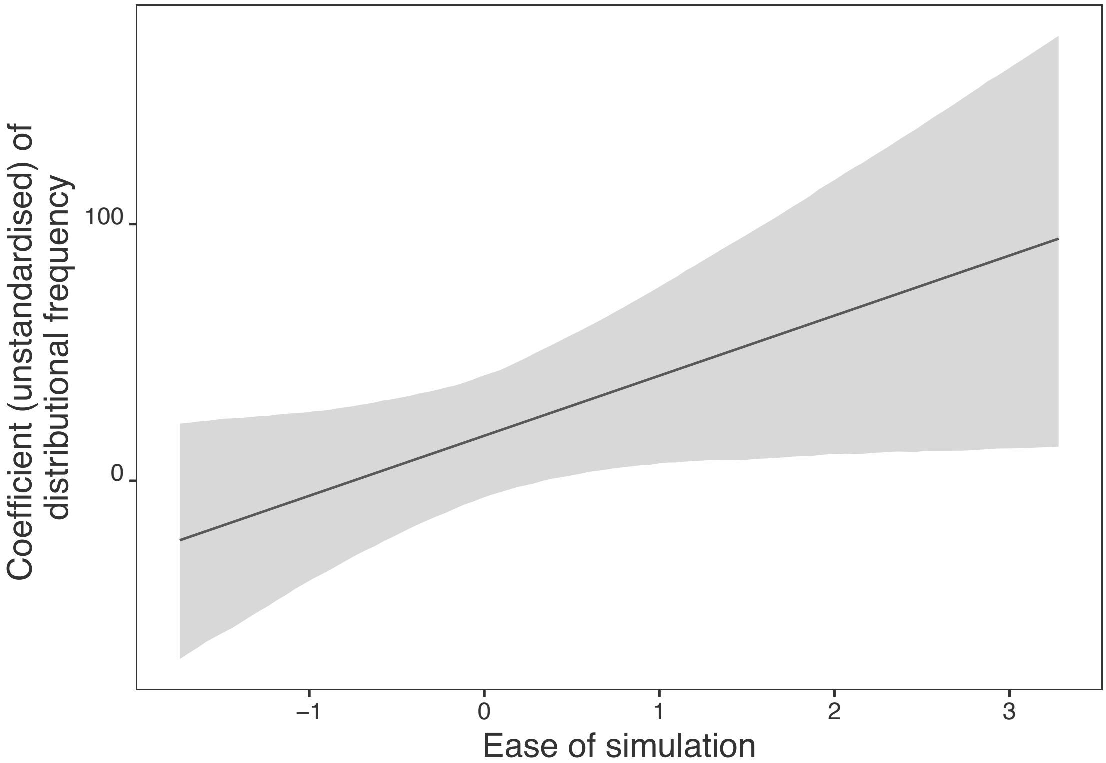


Figure 3. Change in the coefficients of LDF as EoS increased when people rejected metaphors as nonsensical. The shaded area indicates the 95% confidence interval.

Summary. Overall, results showed that people relied on both the embodied and linguistic components when judging metaphor sensibility. Following our simulation hypothesis, metaphors whose meanings were easy to simulate (e.g., *Students can be bright*; *Solutions can be bright*) were more likely to be accepted as sensible and were accepted more quickly, whereas metaphors whose meanings were typically regarded as difficult to simulate (e.g., *Supply can be bright*; *Illness can be bright*) were rejected more quickly. Moreover, EoS had a greater effect on metaphor acceptance times (“yes” b = –134.88) than on rejection times (“no” b = 72.28), as predicted, due to the fact that acceptance times reflect simulation that has successfully completed in the embodied component whereas rejection times reflect simulation attempts that have been abandoned. In addition, as predicted by the linguistic distributional hypothesis, metaphors whose words often co-occurred in context (e.g., *Students can be bright*; *Supply can be bright*) were more likely to be accepted as sensible (although only marginally) and were accepted more quickly, while metaphors whose words rarely co-occurred (e.g., *Illness can be bright*) were not rejected directly. Instead, LDF moderated the effect of EoS (i.e. the effect of EoS was reduced when the LDF was low), which suggested that people rejected the metaphors as nonsensical based on their low LDF, without further processing in the embodied component. This pattern of findings supports the idea that the linguistic shortcut was used as a cognitive triage mechanism during metaphor comprehension to identify processing that was unlikely to succeed (i.e., produce a coherent simulation of meaning) and should therefore be abandoned rather than wasting resources on further processing.

Experiment 2: Interpretation Generation Task

In this study, participants were encouraged to process metaphors in a relatively deep way by completing an interpretation generation task. As they read each metaphor sentence, participants tried to think of a possible meaning for the sentence. If they could think of a meaning, participants were required to specify their interpretation. Interpretation generation can therefore be considered as a deeper form of metaphor processing than sensibility judgement because it requires participants to represent a particular meaning in detail (Connell & Lynott, 2013; Lynott & Connell, 2010). All metaphors were the same as in Experiment 1, but since participants would be engaging in deeper processing when generating interpretations than when simply judging sensibility, we expected they would be more likely to successfully simulate a meaning. Thus, we predicted a higher acceptance rate (i.e., more “yes” responses) in this study than in Experiment 1 (as per Connell & Lynott, 2013), as well as our hypotheses earlier outlined.

Method

Participants. Forty native speakers of English (11 males and 29 females; age M = 19.65, SD = 2.08) were recruited from the same population as Experiment 1 to match the coefficient of variation in Experiment 1. Participation took approximately 20 minutes in exchange for £3.00 or course credits.

Materials. Same as in Experiment 1.

Procedure. The procedure was identical to Experiment 1, except that instructions asked participants to come up with a meaning for each sentence, and to be clear and specific in their meanings. The trial structure was the same as Experiment 1 (see Figure 1) up to the point where participants were required to make a decision. In this study, if they pressed the comma (“,”) key to indicate that they could think of a meaning for the metaphor, they would be presented with a blank screen where they could type in the meaning just generated; if they pressed the full stop (“.”) key to indicate that they could not think of a meaning, they would proceed to the next trial.

The item lists from Experiment 1 were split in half pseudo-randomly while preserving the equal distribution of items along the dimensions of EoS and LDF, in order to reduce the possibility of fatigue, resulting in 8 lists of 56-57 items each. As before, each participant saw only one list. The whole task took approximately 20 minutes to complete and participants were given one break halfway through.

Design and analysis. Data were analyzed in the same way as in Experiment 1. Mixed effects logistic regression of response decisions contained crossed random-intercept factors of participants and items, as did mixed effects linear regressions of response times.[5] As in Experiment 1, there was evidence of net suppression in the regression models due to the shared error variance between predictors (see supplementary materials for details). For that reason, and to maintain consistency in predictors between models and experiments, we therefore report the analysis using orthogonalized variables of EoS and LDF.

Results and Discussion

All participants had mean response times within 3SD of the overall mean so none were excluded on that criterion. We also identified accept ("yes") trials with invalid interpretations, namely those that were: blank or single characters, participant-flagged errors (e.g., "mistake" or "don't know"), simple repetition of the original metaphor (e.g., *Students can be bright* → "students are bright"), replacement of the adjective with a synonym or an adjective that the researchers judged to be nonsensical in context (e.g., *Borders can be deafening* → "borders can be loud"; *Economy can be freezing* → "The economy can be dry"), and clear misreading of words (e.g. *Charge can be rustling* → "coins can be heard to make a [rustling] sound in pockets", consistent with misreading charge as change). Two participants were excluded from analysis for providing more than 50% invalid interpretations. Amongst the remaining participants, 2.33% of interpretations (31

[5] In logistic regression, the inclusion of participants as a random factor improved model fit above the empty model, $\chi^2_{(1)} = 368.76$, $p < .001$), as did the inclusion of items as a crossed random factor, $\chi^2_{(1)} = 156.50$, $p < .001$. In linear regressions, the inclusion of participants as a random factor improved model fit above the empty model $\chi^2_{(1)} = 638.85$, $p < .001$, although the inclusion of items as a crossed random factor did not improve model fit above the participants-only model, $\chi^2_{(1)} = 0.00$, $p = 1.00$. However, as in Experiment 1, both participants and items were included as crossed random factors in the linear models of RT as well as in the logistic models of response decision in order to keep the analyses consistent. Analyses of RT with participant-only models did not substantially alter results, and are reported in supplementary materials.

trials) were identified as invalid and removed from analysis. No trials were removed due to motor error, but RTs more than 3 standard deviations from each participants' mean per response decision were removed as outliers: 2.10% of "yes" responses and 2.00% of "no" responses.

Response decision. Among 2103 valid trials, 1302 (61.91%) were accepted as interpretable ("yes" responses) whereas 801 (38.09%) were rejected as uninterpretable ("no" responses). Logistic regression showed that both EoS and LDF had a positive effect on response decision (see Table 5). For every unit of increase in EoS, the odds of accepting a metaphor as interpretable increased 2.826 times; and for every unit of increase in LDF, it increased 1.286 times.

In short, when people were asked to make a relatively deep judgement about whether they could think of a meaning for a metaphor, their yes/no decision was influenced by both EoS (easy-to-simulate metaphors were more likely to be interpretable) and by the LDF of the words used in the metaphor (frequently co-occurring words were more likely to give rise to an interpretable metaphor). Hence, both the linguistic and embodied components played a role in metaphor processing, as predicted.

Response time. Linear regression of RT across all responses (M = 2837 ms, SD = 2469 ms) found no overall effect of EoS (see Table 6). However, EoS interacted negatively with response decision, indicating the effect of EoS for "yes" was opposite to that for "no" responses. LDF did not have any significant effects or interactions.

As in Experiment 1, we had separate hypotheses for "yes" and "no" RTs and so analysed them separately; results are shown in Table 7. For "yes" responses (i.e. accepting metaphors as interpretable; RT: M = 3083 ms, SD = 2638 ms), EoS had a negative effect, as predicted, meaning that people were faster to accept a metaphor as interpretable when it was typically considered easy to simulate

Table 5

Logistic mixed-effect regression of response decision in Experiment 2

	b	95% CI	*p*
Intercept	0.96	0.42 - 1.50	$<$ .001
EoS	1.53	1.35 - 1.71	$<$ .001
LDF	0.34	0.20 - 0.48	$<$ .001
EoS x LDF	-0.01	-0.15 - 0.13	0.887

Note. *b* is non-standardised regression coefficient.

Both predictors orthogonalized.

Table 6

Omnibus linear mixed-effect model of RT in Experiment 2

	b	95% CI	*p*
Intercept	2796.78	2341.82 - 3251.76	$<$ .001
EoS	125.28	-20.79 - 271.35	.093
LDF	44.47	-97.02 - 185.95	.538
Response	58.38	-143.64 - 260.40	.571
EoS x LDF	-34.60	-167.90 - 98.71	.6111
EoS x response	-589.63	-774.55 - -404.70	$<$ .001
LDF x response	-13.52	-193.60 - 166.55	.883
EoS x LDF x response	4.82	-166.13 - 175.76	.956

Note. *b* is non-standardised regression coefficient.

Both predictors orthogonalized.

compared to difficult to simulate. LDF did not affect the speed of interpretation, nor was there an interaction (Figure 4).

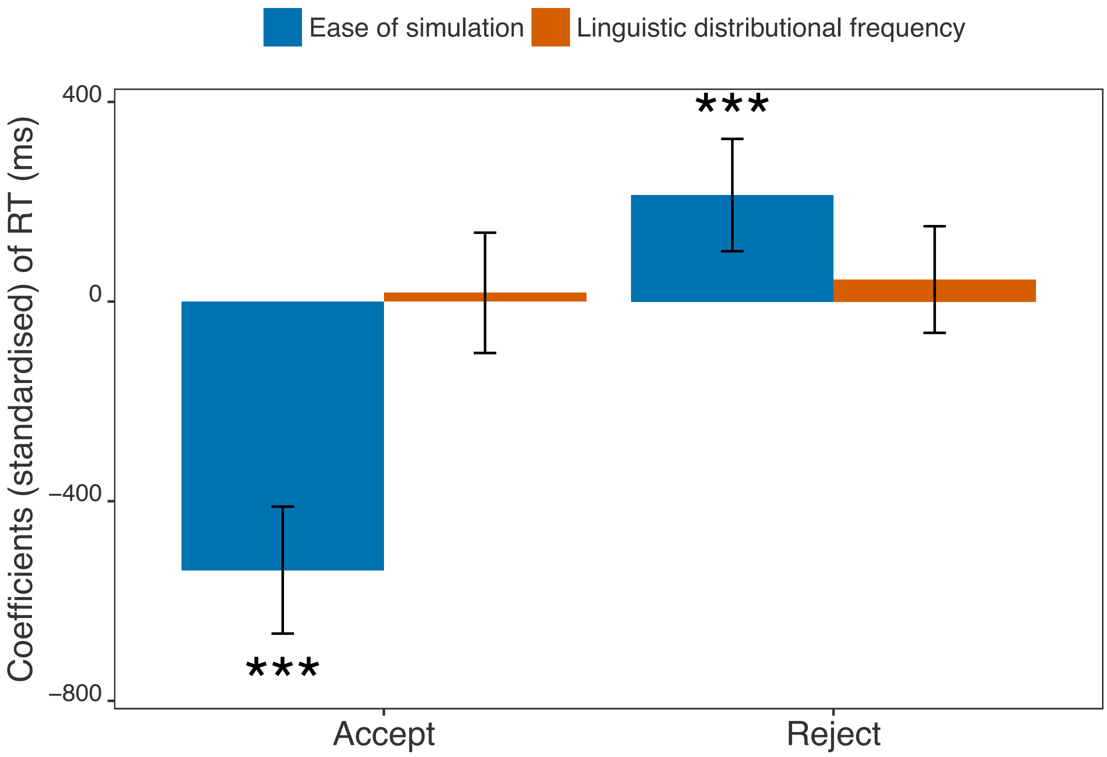


Figure 4. Effects of EoS and LDF, expressed as standardised regression coefficients, on acceptance and rejection times for interpretation generation. Error bars represent 95% confidence intervals, and asterisks represent the significance of directional predictions where specified (*** $p < .001$, ** $p < .01$, * $p < .05$).

For "no" responses that rejected the metaphor as uninterpretable (RT: $M =$ 2436 ms, $SD = 2105$ ms), EoS had a positive effect on RT: as predicted, people were faster to reject a metaphor as uninterpretable when it was normally considered difficult to simulate. LDF did not affect rejection speed, nor did it interact with EoS (Figure 4).

Summary. Results of Experiment 2's interpretation generation task were similar to those of Experiment 1's sensibility judgement, in showing that people

Table 7

Regression model on RT in Experiment 2 per response decision

	Yes			No		
	b	95% CI	*p*	*b*	95%CI	*p*
Intercept	2961.19	2507.85 - 3414.54	< .001	3245.15	2507.68 - 3982.63	< .001
EoS	-538.01	-665.18 - -410.83	< .001	213.38	100.81 - 325.94	< .001
LDF	17.48	-103.06 - 138.02	0.776	44.11	-62.70 - 150.91	0.419
EoS x LDF	-34.12	-149.71 - 81.46	0.563	-56.58	-158.13 - 44.98	0.276

Note. *b* is non-standardised regression coefficient.

Both predictors orthogonalized per response decision.

relied on both the embodied and linguistic components when processing metaphors. EoS affected both the likelihood and processing speed of generating an interpretation for a metaphor, supporting the simulation hypothesis: the easier it was to simulate the meaning of a particular metaphor, the more likely it was to be interpreted and the faster people accepted it as interpretable. Conversely, metaphors whose meanings were difficult to simulate were rejected as uninterpretable relatively quickly. As predicted, EoS had a greater effect on metaphor acceptance times ("yes" $b = -538.00$) where simulation was ultimately successful, than on rejection times ("no" $b = 213.38$) where processing in the embodied component was abandoned. Following the linguistic distributional hypothesis, metaphors whose words often co-occurred in context (e.g., *Students can be bright*) were more likely to be accepted as interpretable than metaphors whose words rarely co-occurred (e.g., *Solution can be bright*). The effects of LDF did not carry over into interpretation times, however, unlike in Experiment 1's sensibility judgement times. Hence, support for the use of the linguistic shortcut as a cognitive triage mechanism in deep metaphor processing is limited to its effect on the likelihood of interpretation. We examine such task differences further in cross-experiment analyses.

Cross-Experiment Analyses

Since some of our hypotheses concern differences between shallow and deep processing tasks, we conducted a meta-analysis of the data from Experiments 1 and 2 in order to compare effect sizes between sensibility judgement and interpretation generation tasks.

Method

The data from Experiments 1 and 2 were combined for analysis: response decision in logistic mixed effects regression, and response time in linear mixed effects regression. All analyses used the same models with orthogonalized variables that were previously employed in separate experiments, this time including an additional fixed predictor variable of task (coded 0 for sensibility judgement, 1 for interpretation generation) that interacted with all other predictors.

Results and Discussion

Response decision. Results are shown in Table 8. The likelihood of accepting versus rejecting a metaphor varied by task: the odds of accepting a metaphor in the interpretation generation task were 3.24 times more than in the sensibility judgement task, as predicted. EoS had a positive effect on response decisions (i.e., easy simulation lead to more successful processing), which stayed constant between shallow and deep tasks. LDF also had a positive effect, and its positive interaction with task indicated that LDF had a larger effect on response decision in deep than shallow metaphor processing.

Response time. Results are shown in Table 9. Deep interpretation generation took significantly longer time than shallow sensibility judgement. EoS had opposite effects on acceptance versus rejection times, which was negative for "yes" responses but positive for "no" responses. Furthermore, EoS interacted with task, showing that its effect is larger for deep interpretation generation than for shallow sensibility judgement across both types of responses, as predicted by the depth of processing hypothesis. The effects of LDF were of the same pattern (despite producing effects in Experiment 1 but not in Experiment 2), though were not significant either as a main effect or in interaction with task. LDF interacted with EoS and task in a three-way interaction when response was "no", which resulted from the presence of an interaction between LDF and EoS in shallow processing but not in deep processing.

Summary. In summary, depth of processing affected the roles of both EoS and LDF. As predicted by the depth of processing hypothesis, EoS had a larger effect during deep processing than during shallow processing: specifying the meaning of a metaphor led people to rely more on embodied simulation than simply judging whether or not a metaphor makes sense. However, this greater reliance on the embodied component influenced only response times, and not the likelihood of accepting the metaphor as sensible or interpretable. Also as predicted, LDF had a larger effect on shallow processing than one deep processing, but only in terms of how it moderated the effect of EoS in "yes" response times: in successful metaphor processing, people relied on the linguistic shortcut to avoid unnecessarily effortful simulation more during sensibility judgement than during interpretation generation. While LDF also affected response decisions, the effect ran contrary to expectations by being larger for deep processing than for shallow processing: people relied on the linguistic shortcut more when attempting to generate a meaning for a metaphor than when deciding whether or not it made sense. Therefore, while our results

Table 8

Logistic mixed-effect regression of response decision for cross-experiment comparison

	b	95% CI	p
Intercept	-0.25	-0.76 - 0.26	.337
EoS	1.48	1.36 - 1.60	$<$ .001
LDF	0.11	0.02 - 0.21	.023
Task	1.18	0.50 - 1.86	$<$ .001
EoS x LDF	-0.00	-0.11 - 0.10	.926
EoS x task	0.01	-0.18 - 0.20	.919
LDF x task	0.22	0.06 - 0.38	.006
EoS x LDF x task	0.00	-0.16 - 0.16	.979

Note. b is non-standardised regression coefficient. Both predictors orthogonalized.

supported the depth of processing hypothesis regarding EoS, the picture of more complex for LDF.

Table 9

Regression model on RT for cross-experiment comparison per response decision

	Yes			No		
	b	95% CI	*p*	*b*	95%CI	*p*
Intercept	1172.56	770.22 - 1574.91	$<$ .001	1156.14	489.11 - 1823.17	.001
EoS	-134.85	-220.25 - -49.45	.002	72.07	24.80 - 119.35	.003
LDF	-27.62	-108.23 - 53.00	.502	18.19	-27.19 - 63.57	.432
Task	1785.91	1253.26 - 2318.55	$<$ .001	2127.20	1243.17 - 3011.23	$<$ .001
EoS x LDF	-13.12	-91.63 - 65.38	.743	23.54	-19.54 - 66.63	.284
EoS x Task	-402.94	-526.50 - -279.38	$<$ .001	145.00	60.73 - 229.27	$<$ .001
LDF x Task	44.64	-72.09 - 161.37	.454	27.39	-52.80 - 107.58	.503
EoS x LDF x Task	-20.72	-133.57 - 92.12	.719	81.67	-157.94 - -5.41	.036

Note. *b* is non-standardised regression coefficient.

General Discussion

The current study investigates what makes metaphors easier to process from the perspective of grounded language comprehension (Barsalou et al., 2008; Connell & Lynott, 2014b; Louwerse & Jeuniaux, 2008; Vigliocco et al., 2009), with a view to establishing how the roles of the embodied component (i.e., simulation of sensorimotor, affective, and other experience) and linguistic component (i.e., distributional patterns of how words co-occur in language) varied with depth of processing. Our goal in taking this approach was to move the investigation of metaphor processing beyond the traditional factors, such as familiarity, conventionality, and aptness, which – while having a long history of use – have been increasingly criticized for theoretical and operational problems that limit their utility in explaining what makes one metaphor easier to understand than another. Following previous research (Connell & Lynott, 2013; Louwerse & Connell, 2011), we used LDF to operationalise processing in the linguistic component, which was calculated based on bi- to 5-gram co-occurrence frequencies from a large corpus. In addition, we proposed a new variable construct called EoS to operationalize processing in the embodied component (Liu et al., 2018), which reflected how easily people find it to arrive at a mental representation of metaphoric meaning.

Following our predictions, we found that EoS and LDF played unique and distinct roles in metaphor processing. They independently predicted the likelihood of accepting a metaphor in both shallow (Experiment 1) and deep (Experiment 2) processing tasks. People were more likely to decide a metaphor was sensible or interpretable if it was normally considered easy to simulate and if its words often appeared together in context. EoS could successfully predict the speed of processing in both tasks (i.e., people were faster to accept a metaphor as sensible or interpretable when it was easy to simulate, and faster to reject it as nonsensical or uninterpretable when it was difficult, regardless of LDF), and the effect was larger for deep interpretation generation than shallow sensibility judgement. In contrast, LDF only predicted the speed of processing in shallow sensibility judgement but not in deep interpretation generation. The more often a metaphor's words co-occurred in language, the faster people were to accept the metaphor when the depth of processing was shallow and good-enough processing (Ferreira et al., 2002) could suffice. Moreover, although LDF did not directly influence the speed of rejecting a metaphor as nonsensical, it moderated the effect of EoS by reducing its influence at low distributional frequencies. This pattern of effects supported the linguistic shortcut hypothesis that LDF can be used as a cognitive triage mechanism to identify metaphors that are unlikely to result in successful simulation, and so costly processing in the embodied component can be abandoned to avoid wasting resources, where task demands and processing goals allow (Connell, 2018; Connell & Lynott, 2013, 2014b).

Our results are consistent with the conceptual combination study of Connell and Lynott (2013), which supported the cognitive triage mechanism of LDF. The critical differences between the conceptual combination study and the current one lies in that LDF had a smaller effect in our study than in the conceptual combination study. Specifically, in our study, LDF did not affect response times at all in the interpretation generation task, whereas in the conceptual combination study it had a significant positive effect on rejection time and overall response time in both tasks. One reason for these differences could be our addition of EoS as a predictor, independent from LDF. Specifically, we found in Experiment 1 that LDF was not entirely orthogonal to EoS, which means that both variables could account for a mutual portion of variance. Therefore, the LDF variable in the previous conceptual combination study did not only account for the linguistic component of conceptual representation, but may also have accounted partially for the embodied component. In contrast, the present study made sure that EoS and LDF were not only theoretically but also statistically distinct.

More broadly, results from the current study supports grounded views of language processing, according to which the linguistic and the embodied components combine to form conceptual representations of meaning (Barsalou et al., 2008; Connell & Lynott, 2014b; Louwerse & Jeuniaux, 2008). Previous research has demonstrated the roles of the linguistic and embodied components in processing literal sentences about sensory features (Louwerse & Connell, 2011), noun-noun conceptual combination (Connell & Lynott, 2013), and property generation (Santos et al., 2011). The present paper showed for the first time that both components are also critical to processing metaphoric language.

Moreover, the present paper showed that the effects of the two components were independent and distinct. In spite of the fact that LDF was kept minimal, we found evidence that it affected the speed and outcome of metaphor processing above and beyond the effect of EoS. People were able to decide whether a metaphor was sensible/meaningful based on the linguistic distributional patterns alone. Thus, we can argue people use lingusitic information heuristically (Barsalou et al., 2008; Connell, 2018; Louwerse & Connell, 2011). Of course, just because two words frequently co-occur does not necessarily mean that they were intended to form a metaphor. However, words tend to appear close together in language because their concepts form part of the same simulation of an idea or situation, which at least some of the times will be metaphorical. Conversely, words that seldom appear in the same context are unlikely to form a metaphor. As such, LDF represents a relatively coarse-grained, but nonetheless highly useful, approximation of whether a particular source and target can potentially form a metaphor.

However, contrary to the linguistic shortcut hypothesis, our study did not support the difference between shallow and deep processing in the cross-experiment analysis of the linguistic component. LDF had similar effects on RT during both deep and shallow processing tasks, and it even had greater effects on response decision during deep processing than shallow processing tasks. In other words, shallow processing did not encourage the use of the linguistic shortcut. One possible reason could be because people had as much time as needed to make a response; they had unlimited time resource to form a mental representation using the embodied component and were not under pressure to conserve processing resources. In future research, we plan to impose a time constraint on metaphor processing in order to further examine the linguistic shortcut hypothesis and the circumstances that incentivize its use.

A key motivation for adopting a grounded approach to metaphor processing is the problems and criticisms of traditional research the factors that have traditionally been the focus of metaphor processing (i.e. familiarity, conventionality, and aptness). However, it is undeniable, as some may argue, that the linguistic component and the embodied component have several parallels with those factors. Specifically, LDF captures the extent to which familiarity, conventionality, and aptness are based on frequency of exposure, and EoS captures the extent to which aptness is based on successful representation of metaphoric meaning. Nevertheless, the grounded approach we propose offers several advantages that make it stand out from traditional approaches.

First, EoS and LDF are better operationalized. Both factors have clear definitions, unlike familiarity, conventionality, or aptness, which have different definitions to different research communities. EoS only refer to how easy it is to simulate the meaning of the sentence. It takes the sentence as a whole, without selectively emphasizing the source or the target terms. In contrast, the definition of conventionality tends to focus on the source term (see Introduction), even though ratings of conventionality are context dependent (i.e. influenced by the target term).

Also, LDF only refers to the co-occurrence of the constituent words, unlike traditional familiarity or conventionality, which could concern either the metaphor sentence itself or the concept it refers to. Thus, the operational definition of the two factors are unambiguous, improving the validity of the construct.

Second, we have demonstrated that these two predictors can exert statistically independent effects. Although the raw variables correlate to some degree, and there is evidence for net suppression in the models, it can be corrected by orthogonalizing the variables to remove the correlation between them. The grounded views of language processing state that the embodied and linguistic components are theoretically distinct, and our statistical treatment reflects just that. In other words, if our two predictors had the same core dependency as familiarity, conventionality and aptness, their effects would not have been unique and separable.

Finally, under the theoretical construct of the grounded views, the present studies produced findings that cannot be explained by the traditional factors, namely the interplay between EoS and LDF according to the required depth of processing. Previous studies have not considered metaphor processing as an interactive process that adapts to the situation. Based on the grounded view, we predicted that people would process metaphors using different strategies according to the requirement of the tasks. Our study found results that support this hypothesis. Indeed, previous studies which have shown an effect of aptness over conventionality often adopt a deep processing task, such as generating interpretations (Glucksberg & Haught, 2006a; L. L. Jones & Estes, 2006), listing attributes (Glucksberg & Haught, 2006b), and creating new metaphors based on examples (Pierce & Chiappe, 2008). Given the parallels we draw between aptness and EoS, this previous work is consistent with our finding that deeper processing relies more on EoS.

In conclusion, metaphor processing relies on conceptual representation that encompasses the two components proposed by the grounded account, that is the linguistic component and the embodied component. These two components interplay flexibly in order to produce representations that satisfy the requirement of the tasks.

References

Allport, D. A. (1985). Distributed memory, modular subsystems and dysphasia. In S. Newman & R. Epstein (Eds.), *Current perspectives in dysphasia* (pp. 207–244). Edinburgh: Churchill Livingstone. Retrieved from https://www.scribd.com/document/17676263/Allport-1985-Distributed-Memory-Modular-Subsystems-and-Dysphasia http://scholar.google.com/scholar?hl=en{\&}btnG=Search{\&}q=intitle:Distributed+mem

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. doi:10.1037/a0016261

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics By 8-Month-Old Infants. *Psychological Science*, *9*(4), 321–324.

Aust, F. (2016). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from https://CRAN.R-project.org/package=citr

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Bache, S. M., & Wickham, H. (2014). *Magrittr: A forward-pipe operator for r*. Retrieved from https://CRAN.R-project.org/package=magrittr

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–59. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3193910{\&}tool=pmcentrez{

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*(4), 577–609; discussion 610–60. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1693222{\&}tool=pmcentrez{

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In A. M. Glenberg & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 245–283). Oxford, England: Oxford University Press. Retrieved from http://psychology.emory.edu/cognition/barsalou/papers/Barsalou{_}et{_}al{_}chap{

Bartoń, K. (2017). *MuMIn: Multi-model inference.* Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., & Maechler, M. (2017). *Matrix: Sparse and dense matrix classes and methods.* Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 295–308. doi:10.1037//0278-7393.19.2.295

Bonner, M. F., & Grossman, M. (2012). Gray matter density of auditory association cortex relates to knowledge of sound concepts in primary progressive aphasia. *Journal of Neuroscience*, *32*(23), 7986–7991. doi:10.1523/JNEUROSCI.6241-11.2012

Borghi, A. M., & Binkofski, F. (2014). *Words as Social Tools: An Embodied View on Abstract Concepts.* New York, NY: Springer New York. doi:10.1007/978-1-4614-9539-0

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*, 1–22.

doi:10.1006/cogp.2001.0748

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., & Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, *46*(2), 743–756. doi:10.1016/j.neuropsychologia.2007.10.007

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*(1), 193–216. doi:10.1037/0033-295X.112.1.193

Brants, T., & Franz, A. (2006). Web 1T 5-gram Version 1 LDC2006T13. DVD. Philadelphia: Linguistic Data Consortium.

Campbell, S. J., & Raney, G. E. (2015). A 25-year replication of Katz et al.'s (1988) metaphor norms. *Behavior Research Methods*, *48*(1), 330–340. doi:10.3758/s13428-015-0575-2

Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods*, *42*(3), 651–64. doi:10.3758/BRM.42.3.651

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: tuning the brain for metaphors. *NeuroImage*, *59*(4), 3212–21. doi:10.1016/j.neuroimage.2011.11.079

Chan, K. Q., Tong, E. M. W., Tan, D. H., & Koh, A. H. Q. (2013). What do love and jealousy taste like? *Emotion*, *13*(6), 1142–1149. doi:10.1037/a0033758

Chiappe, D. L., & Kennedy, J. M. (1999). Aptness predicts preference for metaphors or similes, as well as recall bias. *Psychonomic Bulletin & Review*, *6*(4), 668–76. doi:10.3758/BF03212977

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple*

Regression/Correlation Analysis for the Behavioral Sciences. Mahwah, NJ: Erlbaum. Retrieved from https://books.google.co.uk/books?id=fAnSOgbdFXIC

Connell, L. (2018). What have labels ever done for us? The lingusitic shortcut in conceptual processing. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2013). Flexible and fast: linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, *20*(3), 542–50. doi:10.3758/s13423-012-0368-x

Connell, L., & Lynott, D. (2014a). I See/Hear What You Mean: Semantic Activation in Visual Word Recognition Depends on Perceptual Attention. *Journal of Experimental Psychology. General*, *143*(2), 527–533. doi:10.1037/a0034626

Connell, L., & Lynott, D. (2014b). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*(3), 390–406. doi:10.1111/tops.12097

Connell, L., & Lynott, D. (2016). Do we know what we're simulating? Information loss on transferring unconscious perceptual simulation to conscious imagery. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(8), 1218–32. doi:10.1037/xlm0000245

Dowle, M., & Srinivasan, A. (2017). *Data.table: Extension of 'data.frame'*. Retrieved from https://CRAN.R-project.org/package=data.table

Faust, M., & Mashal, N. (2007). The role of the right cerebral hemisphere in processing novel metaphoric expressions taken from poetry: a divided visual field study. *Neuropsychologia*, *45*(4), 860–70. doi:10.1016/j.neuropsychologia.2006.08.010

Ferreira, F., Bailey, K., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. Retrieved from http://cdp.sagepub.com/content/11/1/11.short

Gelman, A., & Su, Y.-S. (2016). *Arm: Data analysis using regression and multilevel/hierarchical models*. Retrieved from https://CRAN.R-project.org/package=arm

Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language*, *37*(3), 331–355. doi:10.1006/jmla.1997.2527

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183–206. Retrieved from http://www.tau.ac.il/{~}giorar/files/Giora{_}97{_}GradedSalienceHypothesis.pdf

Giora, R. (2007). Is metaphor special? *Brain and Language*, *100*(2), 111–4. doi:10.1016/j.bandl.2006.08.001

Giora, R., & Fein, O. (1999). On understanding familiar and less-familiar figurative language. *Journal of Pragmatics*, *31*(12), 1601–1618. doi:10.1016/S0378-2166(99)00006-5

Glantz, S. A., & Slinker, B. K. (1990). *Primer of Applied Regression and Analysis of Variance* (p. 777). McGraw-Hill Education. Retrieved from https://books.google.com/books?id=UcR6QgAACAAJ{\&}pgis=1

Glenberg, A. M., & Gallese, V. (2012). Action-based language: a theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905–22. doi:10.1016/j.cortex.2011.04.010

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, *7*(2), 92–96. doi:10.1016/S1364-6613(02)00040-2

Glucksberg, S., & Haught, C. (2006a). Can Florida become like the next Florida? When metaphoric comparisons fail. *Psychological Science*, *17*(11), 935–938.

doi:10.1111/j.1467-9280.2006.01807.x

Glucksberg, S., & Haught, C. (2006b). On the Relation Between Metaphor and Simile: When Comparison Fails. *Mind and Language*, *21*(3), 360–378. doi:10.1111/j.1468-0017.2006.00282.x

Glucksberg, S., Gildea, P., & Bookin, H. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, (21), 85–98. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022537182904674

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual Knowledge Retrieval Activates Sensory Brain Regions. *Journal of Neuroscience*, *26*(18), 4917–4921. doi:10.1523/JNEUROSCI.5389-05.2006

Haught, C. (2014). Spain is Not Greece: How Metaphors are Understood. *Journal of Psycholinguistic Research*, *43*(4), 351–356. doi:10.1007/s10936-013-9258-2

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, *41*(2), 301–307. doi:10.1016/S0896-6273(03)00838-9

Henry, L., & Wickham, H. (2017). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr

Johns, B. T., & Jones, M. N. (2012). Perceptual Inference Through Global Lexical Similarity. *Topics in Cognitive Science*, *4*(1), 103–120. doi:10.1111/j.1756-8765.2011.01176.x

Jones, L. L., & Estes, Z. (2005). Metaphor comprehension as attributive categorization. *Journal of Memory and Language*, *53*(1), 110–124. doi:10.1016/j.jml.2005.01.016

Jones, L. L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language*, *55*(1), 18–32. doi:10.1016/j.jml.2006.02.004

Katz, A. N., Paivio, A., Marshark, M., & Clark, J. M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor and Symbolic Activity*, *3*(4), 191–214.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), 4–5. doi:10.1016/S0010-0277(02)00004-5

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13

Lai, V. T., & Curran, T. (2013). ERP evidence for conceptual mappings and comparison processes during the comprehension of conventional and novel metaphors. *Brain and Language*, *127*(3), 484–496. doi:10.1016/j.bandl.2013.09.010

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, *1284*, 145–155. doi:10.1016/j.brainres.2009.05.088

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books. Retrieved from http://books.google.co.uk/books?id=KbqxnX3{_}uc0C

Laurent, J. P., Denhières, G., Passerieux, C., Iakimova, G., & Hardy-Baylé, M. C. (2006). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research*, *1068*(1), 151–160. doi:10.1016/j.brainres.2005.10.076

Liu, P. Q., Connell, L., & Lynott, D. (2018). Ease-of-Simulation norms for 452

Adjective Metaphors. *Manuscript Under Review*.

Louwerse, M. M., & Connell, L. (2011). A taste of words: linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, *35*(2), 381–98. doi:10.1111/j.1551-6709.2010.01157.x

Louwerse, M. M., & Hutchinson, S. (2012). Neurological evidence linguistic processes precede perceptual simulation in conceptual processing. *Frontiers in Psychology*, *3*(October), 385. doi:10.3389/fpsyg.2012.00385

Louwerse, M. M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. In A. Glenberg & A. C. Graesser (Eds.), *Embodiment and meaning: A debate* (pp. 309–326). Oxford, England: Oxford University Press. Retrieved from http://www.madresearchlab.org/Selected{_}Publications{_}files/LouwerseJeuniauxBoo

Louwerse, M. M., Hutchinson, S., Tillman, R., & Recchia, G. (2015). Effect size matters: the role of language statistics and perceptual simulation in conceptual processing. *Language, Cognition and Neuroscience*, *30*(4), 430–447. doi:10.1080/23273798.2014.981552

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41*(2), 558–64. doi:10.3758/BRM.41.2.558

Lynott, D., & Connell, L. (2010). Embodied conceptual combination. *Frontiers in Psychology*, *1*(November), 212. doi:10.3389/fpsyg.2010.00212

Mashal, N. (2013). The role of working memory in the comprehension of unfamiliar and familiar metaphors. *Language and Cognition*, *5*(4), 409–436. doi:10.1515/langcog-2013-0024

Mashal, N., Faust, M., & Hendler, T. (2005). The role of the right hemisphere in processing nonsalient metaphorical meanings: application of principal components analysis to fMRI data. *Neuropsychologia*, *43*(14), 2084–100. doi:10.1016/j.neuropsychologia.2005.03.019

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of bayes factors for common designs*. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Niedenthal, P. M. (2007). Embodying Emotion. *Science*, *316*(5827), 1002–1005. doi:10.1126/science.1136930

Nowosad, J. (2017). *Rcartocolor: 'CARTOColors' palettes*. Retrieved from https://CRAN.R-project.org/package=rcartocolor

Peleg, O., Giora, R., & Fein, O. (2001). Salience and Context Effects: Two Are Better Than One. *Metaphor and Symbol*, *16*(3-4), 173–192. doi:10.1080/10926488.2001.9678894

Pierce, R. S., & Chiappe, D. L. (2008). The Roles of Aptness, Conventionality, and Working Memory in the Production of Metaphors and Similes. *Metaphor and Symbol*, *24*(1), 1–19. doi:10.1080/10926480802568422

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2017). *nlme: Linear and nonlinear mixed effects models*. Retrieved from https://CRAN.R-project.org/package=nlme

Plate, T., & Heiberger, R. (2016). *Abind: Combine multidimensional arrays*. Retrieved from https://CRAN.R-project.org/package=abind

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for mcmc. *R News*, *6*(1), 7–11. Retrieved from https://journal.r-project.org/archive/

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*(2), 303–345. doi:10.1111/j.1756-8765.2010.01111.x

Robinson, D. (2017). *Broom: Convert statistical analysis objects into tidy data frames*. Retrieved from https://CRAN.R-project.org/package=broom

Roncero, C., & Almeida, R. G. de. (2014a). Semantic properties, aptness, familiarity, conventionality, and interpretive diversity scores for 84 metaphors and similes. *Behavior Research Methods*. doi:10.3758/s13428-014-0502-y

Roncero, C., & Almeida, R. G. de. (2014b). The importance of being apt: metaphor comprehension in Alzheimer's disease. *Frontiers in Human Neuroscience*, *8*, 973. doi:10.3389/fnhum.2014.00973

Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Language and Cognition*, *3111*, 83–119. doi:10.1515/

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Marin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802–2810. doi:10.1016/j.neuropsychologia.2007.05.002.A

Solt, F., & Hu, Y. (2015). interplot: Plot the effects of variables in interaction terms. Available at The Comprehensive R Archive Network (CRAN).

Retrieved from https://CRAN.R-project.org/package=interplot

Thibodeau, P. H., & Durgin, F. H. (2011). Metaphor Aptness and Conventionality: A Processing Fluency Account. *Metaphor and Symbol*, *26*(November), 206–226. doi:10.1080/10926488.2011.583196

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth.). New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247. doi:10.1515/LANGCOG.2009.011

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.name/knitr/

Xie, Y. (2016). *Bookdown: Authoring books and technical documents with R markdown*. Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://github.com/rstudio/bookdown

CHAPTER 5

Linguistic Shortcut under Time Constraints

From the previous chapter, we found evidence for independent influences of both linguistic and embodied components during metaphor processing. These two theoretically distinct and statistically separable components can each uniquely account for the performance of metaphor comprehension tasks. It also indeed suggested that people used different strategies to process metaphors when different depths of processing was required. Specifically, they will draw on the embodied component more heavily for relatively deep processing. However, the role of the linguistic component seemed to have been largely unaffected by the depth of processing. If anything, the effect of LDF was enhanced in the deep interpretation generation task, contrary to our prediction based on the linguistic shortcut hypothesis.

Therefore, this chapter sets out to examine whether people could be induced to take the linguistic shortcut. One possible condition for people to take the linguistic component as a shortcut is when processing resources are limited, such as when people are under time pressure. In the work presented in this chapter, I will introduce constraints on the response interval when participants perform the same two tasks as Chapter 4. The limited time resources, combined with the need for only superficial processing in the sensibility judgement task, should prompt people to take the shortcut of linguistic distributional patterns.

Paper 3: Effect of time constraints on the grounded representation of metaphor processing

Page numbers: 156 - 211

Effect of time constraints on the conceptual representation of metaphor processing

Pei Q. Liu[1] & Louise Connell[1] & Dermot Lynott[1]

[1] Department of Psychology, Lancaster University

Author Note

Department of Psychology, Lancaster University, UK.

Correspondence concerning this article should be addressed to Pei Q. Liu, Department of Psychology, Fylde College, Lancaster University, Bailrigg, LA1 4HX, UK. E-mail: p.liu1@lancaster.ac.uk

Abstract

What shapes the conceptual representations during metaphor processing? In this paper, we investigate this question by studying the roles of both embodied simulation and linguistic distributional patterns. Researchers have proposed that the linguistic component is shallow and speedy, ideal as a shortcut to construct crude representations and conserve valuable cognitive resources. Thus, during metaphor processing, people should rely on the linguistic component more if the goal of processing is shallow and the time available is limited. Here, we present two pre-registered experiments which aim to evaluate this hypothesis. The results supported the role of simulation in metaphor processing, but not the linguistic shortcut hypothesis: the effect of linguistic distributional frequency increased as people had more time to process the metaphors, and as they engaged in deep processing. Furthermore during shallow processing, the processing was easier when the embodied and linguistic components support each other. These findings indicate a complex interaction between the embodied and linguistic components during metaphor processing.

Keywords: metaphor processing, embodied simulation, linguistic distributional pattern, linguistic shortcut hypothesis

Word count: 9609 words

Effect of time constraints on the conceptual representation of metaphor processing

Metaphor is a crucial aspect of human language and cognition. A metaphoric expression applies to words and phrases where its meaning departs from the literal meaning of the expression. For instance, *Students can be bright* uses the word *bright* in a non-literal sense (i.e., "clever or intelligent"), while its literal meaning of "emitting much light" does not apply to the concept of *students*. Hence, it is not immediately apparent how people achieve comprehension in such a case. In this paper, we will view metaphor processing as a process that relies on both coarse-grained linguistic distributional information as well as fine-grained embodied simulation (e.g., Barsalou, Santos, Simmons, & Wilson, 2008; Connell & Lynott, 2014; Louwerse & Jeuniaux, 2008). Linguistic distributional information describes patterns of how linguistic symbols (words or phrases) co-occur across language. As they are to do with linguistic symbols, distributional patterns do not have a direct and causal link with meaning. We will argue that distributional patterns are utilised in metaphor processing as well as the embodied component, and test whether the linguistic distributional information can be used heuristically for metaphor processing, particularly under time pressure.

Grounded Views of Language Processing

Research on conceptual representations suggests that at least two components are employed for semantic representation when people process language (Barsalou et al., 2008; Connell & Lynott, 2014; Louwerse & Jeuniaux, 2008; Lynott & Connell, 2010; Vigliocco, Meteyard, Andrews, & Kousta, 2009). On one hand, representations are formed concerning the statistical, distributional pattern of how words co-occur across contexts. For example, the words *bright* and *student* occur together more often in language than do *shining* and *student*. Such distributional patterns are able to explain to a great extent many aspects of language processing, such as priming and reading times (Vigliocco et al., 2009), language acquisition from early infancy (Aslin, Saffran, & Newport, 1998; Monaghan, Chang, Welbourne, & Brysbaert, 2017; Saffran, Johnson, Aslin, & Newport, 1999). Even a simple measure of co-occurrence frequencies within a small five-word window could explain phenomena of language processing, such as the effect of perceptual modalities of the referent (Louwerse & Connell, 2011), and comprehending novel noun-noun combinations (Connell & Lynott, 2013).

On the other hand, past experiences of perceptual, motor, affective, introspective and other features are also partially reactivated during language processing (Allport, 1985; Barsalou, 1999; Glenberg & Gallese, 2012; Niedenthal, 2007). Such reactivation of experiences, which is called simulation, forms the embodied component of a conceptual representation. Evidence for embodied simulation includes shared activation between brain areas involved in perceptual or motor experience and their equivalents in language comprehension. For example, reading sound-related words like *thunder* activates the auditory association cortex, and their processing is selectively impaired in patients with atrophy of the auditory association cortex (Bonner & Grossman, 2012; Boulenger et al., 2008; R. F. Goldberg, Perfetti, & Schneider, 2006). Together, the linguistic and embodied components can explain language processing better than either alone (Andrews, Vigliocco, & Vinson, 2009; Johns & Jones, 2012; Louwerse, 2011).

In particular, researchers have suggested that the linguistic component can provide a shortcut for superficial language processing (Barsalou et al., 2008; Connell & Lynott, 2014; Louwerse & Jeuniaux, 2008). During language processing, the activation of linguistic distributional patterns is easy and requires little effort, so it reaches the peak of its activation before embodied simulation. Furthermore, the co-occurrence of words in language often reflects the associations of objects, events, and ideas in the real world. The fact that *bright* appears with *sun* frequently in close proximity reflects the reality that the sun is usually bright. Therefore, the linguistic component is sufficient to generate a good-enough approximation (Ferreira, Bailey, & Ferraro, 2002), especially when detailed semantic representation is not required (Connell & Lynott, 2013, 2014; Louwerse, 2011; Louwerse & Connell, 2011). In contrast, embodied simulation, though also activated immediately in language processing, is slower and more costly. Therefore, it is reasonable that people would rely on information from the linguistic component for a guide to whether it is worth expending effort on costly embodied simulation (Connell & Lynott, 2013).

To test this proposal, Connell and Lynott (2013) asked participants to perform two tasks with novel noun-noun compounds (e.g., *cactus beetle*): a relatively shallow processing task where participants makes a yes/no judgement concerning whether the compounds made sense, or a deeper processing task where they decided if they could generate specific interpretations for them (and then provided the interpretation). They measured the linguistic distributional information with a simple co-occurrence frequency (i.e., times when *cactus* and *beetle* co-occur with zero to three words in between). The results showed that, when the compound rarely co-occurred, which suggested that the processing was likely to fail, people were more likely to reject the compound and rejected it more quickly, abandoning it before any more cognitive effort was expended. On the other hand, when the compound frequently co-occurred (e.g., *army* and *decision*), which suggested that future processing was likely to succeed, people's response strategy would be based on the requirement of the task. In the shallow sensibility judgment task, linguistic distributional frequency heavily affected the response decision and speed to accept a compound; whereas in the deep interpretation generation task, linguistic distributional frequency did not affect the speed to accept a compound. In other words, while rejecting a noun-noun compound can simply rely on a linguistic shortcut to quickly flag up unsuccessful processing, accepting a compound as interpretable may require more detailed simulation which cannot be captured by the linguistic distributional frequency.

The cognitive triage mechanism, though proposed for literal language processing, can equally be applied metaphor processing. Liu, Connell, and Lynott (2017) tested whether the mechanism worked in metaphor processing with a similar design. Follwing Connell and Lynott (2013), they kept the operational definition of the lingusitic component minimal; and going further, they also introduced a measure of the effort to simulate a concept (Liu, Connell, & Lynott, 2018). In this metaphor processing study, participants read metaphoric sentences such as *Students can be bright* or *Supporters can be lukewarm*, and performed either a yes/no sensibility judgement task or an interpretation generation task. The results of the metaphor processing study supported the role of the linguistic component again. While ease of simulation affected processing profusedly in both tasks, linguistic distributional frequency still contributed independently to the speed of processing in the relatively shallow, sensibility judgment task (as well as the outcome of processing in the relatively deep, interpretation generation task). That is, the co-occurrence of the words *students* and *bright* predicted how quickly people judged the metaphor *Students can be bright* to be sensible, but did not predict how quickly they could generate a specific interpretation for the metaphor (e.g., "clever students").

These results, on the face value, seemed to support the triage mechanism and the linguistic shortcut hypothesis. However, contrary to the idea that people would rely on linguistic distributional frequency more in shallow processing, linguistic distributional frequency did not have a reliably larger effect on response time in the shallow sensibility judgment task than the deep interpretation generation task. Furthermore, linguistic distributional frequency had a greater effect on the outcome of processing in the deep interpretation generation task than the in shallow sensibility judgment task. This pattern of results seems to suggest the opposite of the linguistic heuristic hypothesis: that is, people relied on the linguistic component more heavily when they were required to process a metaphor deeply and generate a detailed mental representation.

One possible explanation for the mixed results could be that the use of the linguistic shortcut needs to be incentivised. That is to say, in order for people to rely on linguistic distributional patterns rather than detailed simulated information, they need to be given conditions that induces a speed-accuracy tradeoff. The advantage of the linguistic shortcut is that it conserves limited cognitive capacity and time compared to the more costly and time-consuming, full-fledged simulation. Therefore, people should be more reliant on the linguistic shortcut if the resources (of time or processing capacity) are limited.

Present Study

In the present study, we will continue to view metaphor processing as an aspect of language processing and examine its conceptual representation. We will test the role of lingusitic distributional patterns in particular, by putting people under different levels of time constraints while they perform the same metaphor processing tasks as the previous study (Liu et al., 2017). For a sentence such as *Supply can be bright*, the linguistic distributional patterns should first suggest that the constituent words *supply* and *bright* often co-occur, so the processing is likely to succeed. In such a case, people could be encouraged to take the linguistic shortcut if they have limited time to make a response. Therefore, they may either accept the metaphor straight away based on the high distributional frequency even though the embodied component has not produced a coherent simulation yet, or they could allow the embodied component to engage in deeper and more costly simulation. In contrast, for a sentence such as *Illness can be bright*, the linguistic component should immediately suggest that the two constituent words rarely co-occur. Based on this information, people may quickly reject the metaphor to avoid time running out or they may only allow the embodied component to process the metaphor in a very small degree. In both cases, the shorter the time is available, the more people should rely on the linguistic component, whereas if people have more time, they can be free to engage in time-consuming simulations.

Additionally, we expect to replicate the effects of linguistic distributional frequency and ease of simulation from the previous processing study. That is, we expect them to affect the judgement and speed of processing independently. People should be more likely and faster to accept a metaphor as sensible or interpretable when ease of simulation is high, and less likely and slower to reject it when ease of simulation is low. Linguistic distributional frequency should further predict a unique portion of response decision and speed, which should be in the same direction as ease of simulation. Task will further moderate the effect of both components, such that ease of simulation will have a larger effect during deeper processing than shallow processing, while vice versa for linguistic distributional frequency.

The main studies with the hypotheses and the method were preregistered on AsPredicted.org (http://aspredicted.org/zv3y2.pdf) and conducted as the pre-registration unless otherwise specified and justified. The pilot study, whose aim was to determine elements of design of the main studies, was not pre-registered due to its exploratory nature, but is reported fully in this manuscript. We report here in all three experiments how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Pilot Study

A pilot study was conducted to determine the time constraints suitable for the experiments. The constraints were selected based on the results from the previous metaphor processing study (Liu et al., 2017). The short constraints were set so that about 50% of the trials per task were responded to within the constraint. The threshold was set to give participants the pressure to respond quickly but not so much that the task became impossible. Thus, the short constraints turned out to be 1s for the sensibility judgment task and 2s for the interpretation generation task. The long constraints were set at the point where the distribution of RT would naturally end if assumed symmetric. This criterion was to ensure that participants could naturally respond to most of the trials without pressure. Thus, the long constraints turned out to be 3s for the sensibility judgement task and 8s for the interpretation generation task. Consequently, the medium constraints were set to be the middle point of the two, that was 2s for the sensibility judgement task and 5s for the interpretation generation task. In the pilot study, we tested if the short constraint was too difficult for participants. The task would be judged as too difficult if people responded "yes" for less than 30% of the trials for the sensibility judgement task and less than 40% for the interpretation generation task. This difficulty threshold was also determined based on the previous study in which the acceptance rate was 45.51% for the sensibility judgement task and 61.91% for the interpretation generation task. That is to say, we expected the task to be more difficult in the current study because of the time constraint. Therefore, the threshold of acceptance rate was lowered in the present study so that people were allowed to fail to process the majority of the metaphors, but not so low as to suggest that participants would fail completely.

Participants

Sixteen native speakers of English from the Department of Psychology, Lancaster University were tested for the pilot study (age: M = 26.26 years, SD = 4.10 years; male: six; left-handed: one), eight for the sensibility judgement task and eight for the interpretation generation task.

Material

The material contained 452 metaphoric sentences from the ease-of-simulation norms (see also Liu et al., 2017, 2018). Table 1 contains examples of the materials, while the full list is in the Supplementary Material (https://goo.gl/sRkXun). All sentences took the form "Noun can be adjective" (e.g., *Students can be bright*), composed of 113 uni-modal perceptual adjectives selected from the modality exclusivity norms of Lynott and Connell (2009) and Dantzig, Cowell, Zeelenberg, and Pecher (2011) (e.g., *bright*), each paired with four nouns so that the adjectives could elicit metaphoric meanings. By pairing each adjective with four nouns, the metaphors varied on the following two variables:

Ease of simulation (EoS). EoS measured indirectly the effort to successfully simulate the concept in a metaphoric sentence. For each sentence, EoS ranged from easy to difficult ($M = 0.00$, $SD = 1.00$; values automatically standardised through PCA), which was obtained through a novel norming study by Liu et al. (2018) where people rated on the metaphors based on three criteria: sensibility (How much sense does the sentence make if you read it in text or heard it in conversation?), usability (How easy it would be for you to use the sentence in writing or in conversation?), and imaginability (How easy it is for you to imagine the concept described in the sentence?). One principle component was extracted from the ratings, the factor scores of which was used as a continuous measure of how easy it is to simulate the concept. This EoS measure managed to explain 91% of the original variance of the ratings and outperformed all the separate ratings in predicting the response time of a sensibility judgement task (Liu et al., 2018). Therefore, EoS could be argued to measure the underlying mechanism of comprehension, instead of the noise related to specific linguistic tasks, which was

Table 1

Sample metaphors, and their scores for EoS and LDF

Metaphor	EoS	LDF
Illness can be bright.	-1.32	2.95
Supply can be bright.	-1.02	3.72
Solutions can be bright.	1.41	3.11
Students can be bright.	1.84	4.08
Minutes can be lukewarm.	-1.49	3.31
Scheme can be lukewarm.	-0.78	1.61
Supporters can be lukewarm.	0.69	2.69
Reaction can be lukewarm.	1.03	3.45

assumed to be simulation (Zwaan, 2004).

Linguistic distributional frequency (LDF). LDF measured the co-occurrence frequency of the constituent words in a sentence. For each sentence, LDF ranged from low to high (M = 2.95, SD = 0.97), and was calculated as the log-transformed sum of the bi- to five-gram frequencies of the metaphor's constituent words in the Google Web1T Corpus (Brants & Franz, 2006). To take the metaphor *Students can be bright* as an example, the LDF was the sum of the frequencies of "student ... bright" and "bright ... student" with zero, one, two, and three intervening words. It was then log transformed as $\log_{10}(\text{LDF} + 41)$ where 41 is the lowest non-zero frequency in the corpus (Connell & Lynott, 2013). Among the four metaphors created for each adjective, LDF varied independently from of EoS (see Table 1). The two variables had a mild correlation (r = .26, sharing only 7% of common variance), which was handled later during the analysis.

These variables did not contribute to the analysis of the pilot study data, but were used to construct balanced stimulus lists. The sentences were split into four lists of 113 items each, where each adjective appeared only once per list, and the distribution of easy/difficult to simulate and high/low distributional frequency was approximately equal across lists (EoS: $F_{(3,448)} = 0.03$, $p = .99$; LDF: $F_{(3,448)} = 0.09$, $p = .96$). Each participant saw only one list.

Procedure

Participants read the metaphoric sentences and performed either of the two tasks. They were randomly assigned to one of the tasks and judged whether or not the sentence made sense, or whether they could think of a meaning for the sentence. Each trial was conducted as in Figure 1 in the two tasks except at the final screen when people needed to make a judgement. The adjectives remained on the screen until participants responded or the time ran out. For the sensibility judgement task, participants were given 1s to respond; for the interpretation generation task, they were given 2s to respond. If their response was "yes" (i.e. they judged the sentences as sensible or meaningful), they would press the comma key (,); if "no", they would press the full stop key (.). In the interpretation generation task, they were then asked to type down the meaning if they had responded "yes". If they did not manage to give a response within the time limit, a feedback saying "***TOO SLOW***" was given for 350ms.

Design and Analysis

Response decisions and RT were recorded, but only response decisions were analysed. The "yes" response was categorised as accepted trials, the "no" response as rejected trials, and if the time elapsed without any responses, a missed trial.

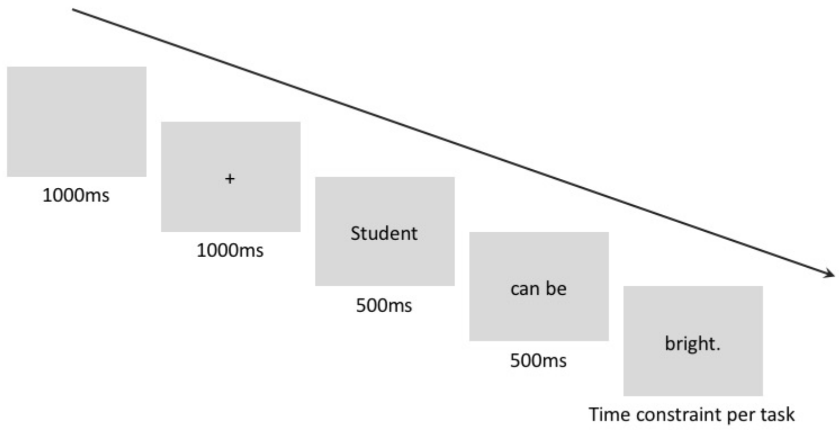


Figure 1. Trial structure of both sensibility judgement and interpretation generation.

Results

The sensibility judgement task had an average missing rate of 7.41% and an acceptance rate of 35.73%. The interpretation generation task had an average missing rate of 3.32% and an acceptance rate of 49.45%. Hence, we decided that the short time constraints for both tasks were acceptable and applied the time constraints to Experiments 1 and 2.

Experiment 1: Sensibility Judgement Task

In this experiment, people performed sensibility judgements under three levels of time constraints (i.e., a short constraint of 1s, a medium constraint of 2s and a long constraint of 3s). Since it only required a simple "yes" or "no" answer, the experiment encouraged people to engage in relatively shallow processing (Connell & Lynott, 2013; Lynott & Connell, 2010). We expected that while EoS would have an effect on the response decision and RT (i.e., high EoS, more and faster acceptance, less and slower rejection), LDF would also have an effect (i.e., high LDF, more and faster acceptance, less and slower rejection), especially under short time constraint. Furthermore, if people could simulate a metaphor extensively and in detail, the processing should end up being successful. Therefore, EoS should have a larger effect on the RT when the response was "yes" than "no". In contrast, because the linguistic shortcut could flag up potentially unsuccessful processing (i.e., those with low LDF), EoS should have a smaller effect and LDF should have a larger effect when the response was "no" than "yes".

Participants

Forty-eight students from Lancaster University participated in the study, all of whom were native speakers of English (age: 19.49 (SD = 2.91) years; male: seven; right-handed: 43). Participation took approximately 20 minutes in exchange for course credits. The sample size was determined based on a minimum acceptance rate of 35%, which would provide a minimum number of data points per participant that was comparable to Liu et al. (2017); because this acceptance rate was achieved, it was not necessary to test any extra participants.

Materials

The 452 metaphoric sentence and item lists were the same as in the pilot study.

Procedure

The study was composed of three test blocks, one for each level of time constraint: short (1s), medium (2s), and long (3s), to which the materials were randomly assigned. Participants were made aware that the time constraints might vary between blocks, although they did not know whether a given block would use a short, medium or long constraint. All items appeared under all time constraints across participants, and the order of blocks were counterbalanced. Before each block, participants were given ten practice trials in order to accustom them to the time-constrained response interval, during which their response was not recorded. The trial structure in the practice and test trials was identical to the pilot study. Participants were asked to respond "yes" if the sentence made sense and "no" if it did not make sense. Response decisions ("yes" or "no") were recorded for each trial, and response times were measured from the onset of the adjective until the keypress of the response decision.

Design and Analysis

This experiment had EoS and LDF as two continuous predictors, and three levels of time constraints (i.e., short, medium and long) were applied as within-subject categorical predictors. The time constraints were backward difference coded so that Contrast 1 was the difference of long minus medium constraints(with medium constraint as the reference level), and Contrast 2 is the difference of medium minus short constraints (with short constraint as the reference level). The coding was not specified in the pre-registration, but it allowed us to spot the change of the effects of the two continuous predictors as time constraint tightened. Response decisions ("yes" or "no" responses) and response time (RT in millisecond) were the dependent variables. The response decisions were coded as 1 for "yes" response and 0 for "no" response.

The analyses were conducted according to the pre-registration.[1] Response

[1] We used R (Version 3.5.0; R Core Team, 2017) and the R-packages *abind* (Version 1.4.5; Plate & Heiberger, 2016), *arm* (Version 1.10.1; Gelman & Su, 2016), *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2015), *bookdown* (Version 0.7; Xie, 2016), *broom* (Version 0.4.4; Robinson, 2017), *citr* (Version 0.2.0; Aust, 2016), *coda* (Version 0.19.1; Plummer, Best, Cowles, & Vines, 2006),

decisions[2] were analysed using logistic mixed-effect regressions with centred EoS, centred LDF, levels of time constraints and their interactions as fixed predictors, and participant and item as crossed random-intercept factors.[3] Furthermore (not pre-registered but consistent with our hypotheses), we also analysed the response decision per time constraint in order to model the effects of the predictors within each time constraint individually. RT[4] were analysed using linear mixed-effect regressions with the same fixed predictor plus response decision, and the

---

data.table (Version 1.11.4; Dowle & Srinivasan, 2017), *ggplot2* (Version 2.2.1; Wickham, 2009), *gridExtra* (Version 2.3; Auguie, 2017), *interplot* (Version 0.1.5; Solt & Hu, 2015), *knitr* (Version 1.20; Xie, 2015), *lme4* (Version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.0.1; Kuznetsova, Brockhoff, & Christensen, 2017), *magrittr* (Version 1.5; Bache & Wickham, 2014), *MASS* (Version 7.3.50; Venables & Ripley, 2002), *Matrix* (Version 1.2.14; Bates & Maechler, 2017), *MuMIn* (Version 1.40.4; Bartoń, 2017), *nlme* (Version 3.1.137; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017), *papaja* (Version 0.1.0.9735; Aust & Barth, 2017), *psych* (Version 1.8.4; Revelle, 2017), *purrr* (Version 0.2.5; Henry & Wickham, 2017), and *rcartocolor* (Version 0.0.22; Nowosad, 2017) for all our analyses and the writing up of this manuscript.

[2] The inclusion of participants as a random factor improved model fit above the empty model, $\chi^2(1) = 394.00$, $p < .001$, as did the inclusion of items as a crossed random factor above the participants-only model, $\chi^2(1) = 753.31$, $p < .001$.

[3] Although it has been suggested that that mixed-effect models with random intercept and slope generalises better than a random-intercept-only model (Barr, Levy, Scheepers, & Tily, 2013), we were not able to fit maximal models to our data because of several reasons. First, such a model (with two random predictors and three or four fixed predictors and their interactions) would require a huge amount of data for the parameter estimation which our study could not afford. Second, fitting such models is time consuming, often taking hours or days if performed in R, which makes tweaks and model comparisons impractical.

[4] The inclusion of participants as a random factor improved model fit above the empty model, $\chi^2(1) = 785.53$, $p < .001$. The inclusion of items as a crossed random factor did not improve model fit above the participants-only model $\chi^2(1) = 0.00$, $p = 1$. However, in order to keep the models consistent, both participants and items were included as crossed random factors in the linear models of RT as well as in the logistic models of response decision.

random-intercept factors as the mixed-effect logistic regressions. We also analysed the RT per response decision because we had different hypothesis for "yes" and "no" responses.

In these analyses, we also considered variables such as lexical frequency (sum of the log transformed frequencies of the component words) and sentence length (sum of the length of the component words) as additional predictors as pre-registered. However, the correlation between these variables and RT were near zero ($r < .10$), so they were not included in the regression models (see Supplementary Material). Moreover in the analyses of RT, we found evidence of net suppression and problems with multicolinearity [J. Cohen, Cohen, West, and Aiken (2003); see Supplementary Materials]. This means that the shared error variance between LDF and EoS is effectively hiding the real relationship between the key variables and RT, such that the net effect of LDF was to enhance the effect of EoS by suppressing the latter's unhelpful error variance. Therefore, consistent with our pre-registration, EoS and LDF were centred and orthogonalised using principle components analysis with varimax rotation and Kaissar normalisation on a model by model basis. For clarity and space, we reported only results with orthogonalised variables for RT (for results with original variables, see Supplementary Material), and referred to these orthogonalised variables with their original labels (i.e., EoS instead of orthogonalised EoS, LDF instead of orthogonalised LDF). In the analysis of response decision, we did not find evidence of net suppression, so we did not orthogonalise the variables.

Results and Discussion

All participants had their mean response time within 3SD of the overall mean, so all were included in analysis. Three trials were excluded for motor error (RT < 200ms). Individual trials were excluded as outliers if the RT was more than 3SD from each participant's mean per response decision per level of time constraint. This exclusion was done differently from the pre-registration, in which we proposed to exclude outliers per response decision only, because we realised that this approach would have affected responses under short constraint and long constraint disproportionately. The percentage of outlying trials were: for "yes" responses, 0.15% under short constraint, 0.28% for medium constraint, and none for long constraint; for "no" responses, none for short constraint, 0.64% for medium constraint, 0.36% for long constraint. Consequently, 5261 trials were counted as valid trials and were used in analysis.

Response decisions. Overall, 2059 trials were accepted as sensible (acceptance rate: 39.14%); 3202 trials were rejected as nonsensical (rejection rate: 60.86%). For each level of time constraint, the acceptance rate was 39.23% for short constraint (656 among 1672 trials), 39.18% for medium constraint (701 among 1789 trials), and 39.00% for long constraint (702 among 1800 trials).

Results of the logistic mixed-effect regression are in Table 2. Only EoS had a significant and positive effect. It means that the easier the metaphor was normally regarded to simulate, the more likely it is to be judged as sensible (odds ratio = 3.26). Furthermore, the effect of EoS changed between short time constraint and medium time constraint, being larger for medium than short. Besides, there was also a trend that the effect of EoS became slightly larger as LDF increased. When separated by the levels of time constraint (Table 3), EoS had an effect for all three levels of time constraint. Additionally, the slight trend of interaction between EoS and LDF appeared only for short time constraint but not for medium or long constraints.

Response time. The overall mean RT was 856ms for "yes" responses (*SD* = 312ms) and 838ms for "no" responses (*SD* = 305ms). Separated by time constraint, the mean RT for "yes" responses was 687ms (*SD* = 156ms) for short constraint, 863ms (*SD* = 272ms) for medium constraint, and 1008ms (*SD* = 373ms) for long

Table 2

Logistic mixed-effect regression of response decision in Experiment 1

	b	95% CI	p
Intercept	-0.64	-0.93 - -0.36	$<$ .001
Long - medium	-0.01	-0.18 - 0.16	.932
Medium - short	-0.01	-0.18 - 0.16	.896
LDF	0.06	-0.03 - 0.16	.177
EoS	1.18	1.08 - 1.28	$<$ .001
Long - medium x LDF	-0.08	-0.27 - 0.1	.383
Medium - short x LDF	-0.01	-0.19 - 0.18	.948
Long - medium x EoS	-0.12	-0.31 - 0.07	.225
Medium - short x EoS	0.38	0.19 - 0.57	$<$ .001
LDF x EoS	0.08	-0.01 - 0.17	.084
Long - medium x LDF x EoS	-0.03	-0.21 - 0.16	.776
Medium - short x LDF x EoS	-0.04	-0.22 - 0.15	.714

Note. b is non-standardised regression coefficient.

constraint; the mean RT for “no” response was 688ms (SD = 140ms) for short constraint, 839ms (SD = 251ms) for medium constraint, and 974ms (SD = 390ms) for long constraint.

In the omnibus model of RT (Table 4), we found that the time constraints did have a significant effect on RT. Participants responded more slowly as the time constraints got longer. EoS had a positive effect on “no” response when LDF was at its mean, meaning that the RT became longer as EoS increased. This effect was reversed for “yes” response (i.e., RT became shorter as EoS increased) and the effect

Table 3

Logistic regression model of response decision per time constraint in Experiment 1

		b	95% CI	p
Short	Intercept	-0.62	-0.92 - -0.32	$<$ .001
	LDF	0.09	-0.05 - 0.23	.197
	EoS	0.96	0.81 - 1.11	$<$ .001
	LDF x EoS	0.11	-0.02 - 0.24	.097
Medium	Intercept	-0.64	-0.93 - -0.36	$<$ .001
	LDF	0.08	-0.06 - 0.22	.266
	EoS	1.31	1.15 - 1.48	$<$ .001
	LDF x EoS	0.07	-0.07 - 0.22	.319
Long	Intercept	-0.70	-1.06 - -0.35	$<$ .001
	LDF	0.01	-0.14 - 0.16	.859
	EoS	1.30	1.13 - 1.48	$<$ .001
	LDF x EoS	0.06	-0.09 - 0.2	.457

Note. b is non-standardised regression coefficient.

on “yes” response was larger than that on “no” response. Furthermore, LDF moderated the effect of EoS on “no” response such that the effect of EoS became smaller (less positive) as LDF increased, which existed only for the long and medium constraints (reference levels of the contrast coding) and was smaller at medium than long constraints.

Table 4

Omnibus linear mixed-effect model of RT in Experiment 1

	b	95% CI	p
Intercept	845.92	808.68 - 883.17	$< .001$
Long - medium	129.20	106.25 - 152.16	$< .001$
Medium - short	158.75	135.54 - 181.96	$< .001$
Response	18.48	1.72 - 35.24	.031
LDF	-0.93	-11.76 - 9.9	.867
EoS	14.80	3.49 - 26.11	.010
Long - medium x Response	55.87	17.35 - 94.39	.004
Medium - short x Response	36.72	-1.8 - 75.23	.062
Long - medium x LDF	-0.55	-24.45 - 23.36	.964
Medium - short x LDF	3.13	-21.12 - 27.38	.800
Response x LDF	-1.99	-18 - 14.03	.808
Long - medium x EoS	-7.11	-31.89 - 17.67	.574
Medium - short x EoS	19.71	-5.3 - 44.71	.122
Response x EoS	-56.54	-72.8 - -40.29	$< .001$
LDF x EoS	-13.28	-23.68 - -2.88	.012
Long - medium x Response x LDF	-18.30	-57.56 - 20.95	.361
Medium - short x Response x LDF	7.21	-31.58 - 46.01	.715
Long - medium x Response x EoS	-71.38	-109.79 - -32.98	$< .001$
Medium - short x Response x EoS	-54.82	-93.23 - -16.41	.005
Long - medium x LDF x EoS	-21.05	-43.61 - 1.52	.068
Medium - short x LDF x EoS	3.26	-19.91 - 26.43	.783
Response x LDF x EoS	12.16	-3.2 - 27.53	.121
Long - medium x Response x LDF x EoS	25.89	-10.54 - 62.33	.164
Medium - short x Response x LDF x EoS	-15.00	-50.93 - 20.92	.413

Note. b is non-standardised regression coefficient.

Table 5

Regression model on RT in Experiment 1 per response decision

	Yes			No		
	b	95% CI	*p*	*b*	95%CI	*p*
Intercept	861.79	824.74 - 898.85	< .001	850.65	807.05 - 894.24	< .001
Long - medium	144.15	118.34 - 169.95	< .001	132.00	111.69 - 152.3	< .001
Medium - short	175.15	148.9 - 201.41	< .001	150.57	129.86 - 171.28	< .001
LDF	-7.45	-19.97 - 5.07	.244	3.64	-6.46 - 13.74	.481
EoS	-49.13	-61.9 - -36.37	< .001	17.26	7.06 - 27.46	.001
Long - medium x LDF	-20.66	-46.43 - 5.11	.116	4.50	-15.97 - 24.97	.667
Medium - short x LDF	1.15	-24.83 - 27.12	.931	0.53	-20.4 - 21.46	.96
Long - medium x EoS	-70.98	-97.7 - -44.27	< .001	-3.58	-24.81 - 17.64	.741
Medium - short x EoS	-38.77	-65.26 - -12.29	.004	15.85	-5.5 - 37.19	.146
LDF x EoS	-3.25	-15.49 - 8.98	.602	-11.90	-21.14 - -2.66	.012
Long - medium x LDF x EoS	0.08	-26.68 - 26.84	.995	-19.30	-38.51 - -0.1	.049
Medium - short x LDF x EoS	-5.81	-31.39 - 19.77	.656	5.66	-14.06 - 25.39	.574

Note. b is non-standardised regression coefficient.

When separated by response decision (Table 5), EoS had a negative effect on RT for “yes” response during long and medium constraints. That is, the easier a metaphor was typically considered to simulate, the faster people accepted a metaphor as sensible. This effect was larger at the longer constraints than the short constraint (Table 6 and Figure 2). Analyses per time constraint also confirmed that EoS did not have a significant effect on “yes” RT at short time constraint, but only did during medium and long constraint (Table 11 and Figure 2). In contrast, LDF did not have a significant effect on “yes” RT.

Table 6

Regression model on RT in Experiment 1 per response decision per time constraint

		Yes			No		
		b	95% CI	*p*	*b*	95%CI	*p*
Short	Intercept	695.86	670.25 - 721.47	$< .001$	695.55	673.1 - 717.99	$< .001$
	LDF	-6.18	-16.95 - 4.59	.261	1.47	-6.45 - 9.4	.716
	EoS	-0.50	-11.71 - 10.71	.931	6.80	-1.23 - 14.83	.098
	LDF x EoS	-3.74	-14.16 - 6.68	.483	-1.43	-8.95 - 6.1	.711
Medium	Intercept	874.41	824.4 - 924.42	$< .001$	846.18	800.76 - 891.61	$< .001$
	LDF	0.08	-17.9 - 18.07	.993	0.98	-11.7 - 13.66	.880
	EoS	-42.70	-61.1 - -24.3	$< .001$	23.80	10.9 - 36.7	$< .001$
	LDF x EoS	1.25	-16.32 - 18.81	.889	-6.04	-17.56 - 5.49	.305
Long	Intercept	1,001.11	932.49 - 1069.72	$< .001$	1,003.82	921.12 - 1086.51	$< .001$
	LDF	-19.14	-40.56 - 2.28	.081 †	-0.38	-18.04 - 17.29	.967
	EoS	97.66	-120.02 - -75.31	$< .001$	37.22	18.98 - 55.45	$< .001$
	LDF x EoS	-3.97	-25.92 - 17.98	.723	-17.75	-34.18 - -1.31	.035

Note. *b* is non-standardised regression coefficient.

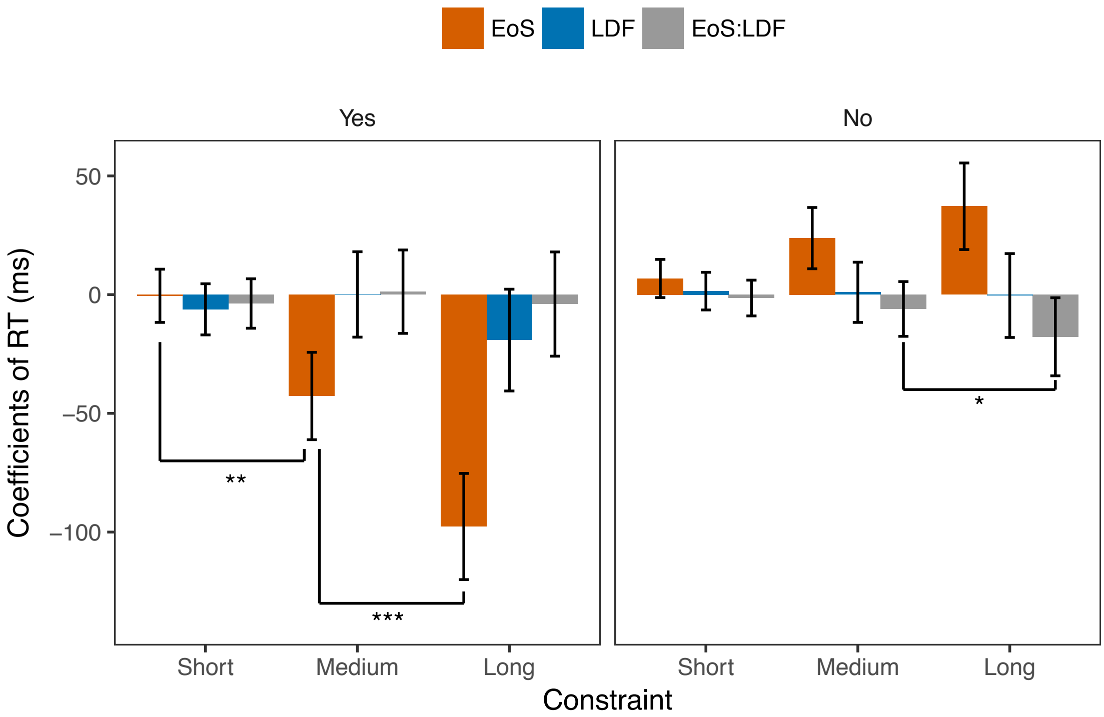


Figure 2. Regression coefficients of RT per response decision in Experiment 1. Error bars stand for 95%CI. Asterisks stand for p-values of interaction between variables and time constraint. *: $p < .05$, **: $p < .01$: ***: $p < .001$.

For "no" response, EoS had a positive effect (when LDF was at its mean) which meant that as EoS increased, RT increased as well (Table 5). In other words, metaphors which were typically considered hard to simulate were rejected more quickly than those considered easy to simulate. the harder it was typically considered to simulate a metaphor, the faster people rejected the metaphor as nonsensical (Table 5). This effect did not change significantly with time constraint, which meant that although the effect of EoS seemed to have increased from short to long constraint (Table 6 and Figure 2), the difference was not detected by the way the contrast was coded. Besides, EoS was moderated by LDF, which further interacted with the time constraints. Separate analyses per time constraint showed that the interaction between EoS and LDF was borne out by the long constraint.

Figure 3 shows the changes of the effect of EoS at four quartiles of LDF (within each quartile, the number of trials were the same). It could be seen that when the LDF was at the lower end, EoS had a positive effect. In other words, when the constituent words rarely appeared in the same context, people were still faster to reject a metaphor that was typically considered to be hard to simulate, and slower to reject one that was typically considered easy to simulate. As LDF increased, the effect of of EoS diminished and then turned to the opposite direction, so much so that when the constituent words frequently co-occurred, people were faster to reject the metaphors which were typically considered easy to simulate (though only a small number of trials were rejected in this case), but slower to reject those metaphors which were typically considered hard to simulate.

Summary. In this experiment, we found that EoS affected metaphor processing as predicted. When a metaphor was typically regarded to be easy to simulate, it was easier (more likely and faster) for people to accept the metaphor as sensible, and harder (less likely and slower) to reject it as nonsensical. The effect of EoS on acceptance speed was moderated by the time constraint, that is: people relied on simulation more when there was longer time available. Thus, it supported the simulation-linguistic based conceptual representation which claimed that the embodied component was more costly and time-consuming, and required longer time to be fully engaged.

The effect of EoS on rejection speed was more complex. It had a positive effect as expected when LDF was at its mean or particularly low. In other words, people found it easier to reject a metaphor which were typically regarded as difficult to simulate. However, the effect of EoS diminished and even was reversed if the constituent words frequently co-occurred, in which case people spent longer time to process a metaphor that was typically harder to simulate (*Supply can be bright*) than one that was typically easy to simulate (*Students can be bright* - rarely

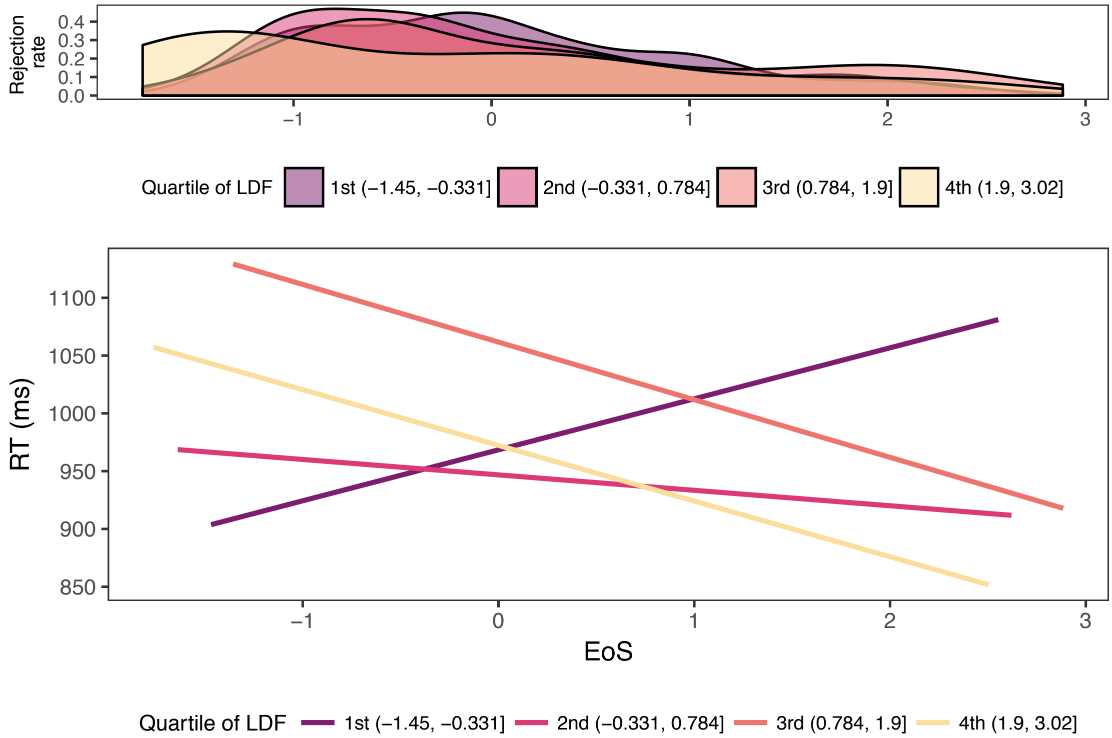


Figure 3. Effects of EoS at long constraint at four quartiles of LDF. As LDF increased from the 1st quartile (yellow) to the 4th quartile (violet), the effect of EoS changed from positive to negative. The density plot on the top shows the rejection rate along EoS per quartile of LDF. The rejection rate at the high end of EoS was less than 10%.

rejected). In other words, people were willing to expend more effort on a metaphor before rejecting it, if the constituent words frequently co-occurred. Such an interaction between EoS and LDF supported the linguistic shortcut hypothesis. The distributional patterns acted as a shortcut for identifying metaphors that could potentially be processed successfully. Metaphors, which were considered hard to simulate, were rejected quickly if their constituent words rarely co-occurred, but slowly if the constituent words often co-occurred.

However, we did not find strong support for the linguistic shortcut hypothesis in terms of how the linguistic component behaved under time constraint. LDF, contrary to our predictions, actually had a larger effect at the long constraint. It was possible that imposing time constraints did not incentivise the linguistic shortcut, but rather caused people to change their processing strategies altogether. As a result, the RT of the short constraint could not be predicted by either of the key predictors, which suggested that people may have based their judgements on variables at the lexical level such as word length and lexical frequencies (even though these variables did not have a strong correlation with the RT overall).

Experiment 2: Interpretation Generation Task

In this study, people performed an interpretation generation task under three different time constraints, a short time constraint of 2 seconds, a medium time constraint of 5 seconds, and a long time constraint of 8 seconds. If they decided they could interpret the sentences, they would be required to type down the meaning of the sentence. We predicted that EoS should have an effect on the response decision and RT (higher EoS, more and faster acceptance; lower EoS, more and faster rejection), as well as LDF (higher LDF, more and faster acceptance; lower LDF, more and faster rejection). Furthermore, EoS would have larger effect under longer time constraints, while LDF would have smaller effect under longer time constraints.

Participants

Fifty-four participants were recruited for this experiment (age: 20.43 (4.38) years; male: 15; left-handed: seven). They participated in the study for 30 minutes for £3.50 or the equivalent of course credits. The sample size was pre-determined based on a minimum acceptance rate of 45%, which would provide at least the same number of data points as Liu et al. (2017); because this acceptance rate was achieved, it was not necessary to test any extra participants.

Materials

The same materials were used in this experiment as in Experiment 1. However, because participants were now required to type interpretations, we reduced the length of the item lists from 113 items to 75-76 items in order to reduce chances of fatigue. Thus, six item lists were generated in the same way as the pilot study and Experiment 1. Each participants saw each adjective only once and the distribution of EoS and LDF were equal among the six lists (EoS: $F_{(5,446)} = 0.21$, $p = .96$; LDF: $F_{(5,446)} = 0.49$, $p = .78$).

Procedure

The procedure was the same as Experiment 1 with the following exceptions. The short, medium and long time constraints lasted 2 seconds, 5 seconds, and 8 seconds, respectively. Participants were asked to respond "yes" if they could think of a meaning for the sentence, and "no" if they could not think of a meaning for the sentence. If participants responded "yes", they were required to type the meaning of the sentence at the next screen, with no time limit for typing.

Design and Analysis

Same as Experiment 1 for response decisions[5] and RT.[6]

[5] The inclusion of participants in logistic mixed-effect model model as a random factor improved model fit above the empty model, $\chi^2(1) = 562.71$, p $< .001$, as did the inclusion of items as a crossed random factor above the participants-only model, $\chi^2(1) = 738.94$, $p < .001$.

[6] The inclusion of participants in linear mixed-effect model as a random factor improved model fit above the empty model, $\chi^2(1) = 764.04$, $p < .001$. The inclusion of items as a crossed random factor did not improve model fit above the participants-only model $\chi^2(1) = 0.00$, $p = 1$. However, in order to keep the models consistent, both participants and items were included as crossed random factors in the linear models of RT as well as in the logistic models of response decision.

Results and Discussion

The data exclusions were the same as in Experiment 1 with an additional validity check on participants' interpretations to make sure they were performing the task properly. An interpretation was marked as invalid if it was left blank, marked out as a mistake ("I don't know"), if it was clearly misread (interpretation specifies meaning of similar word, e.g. "charge" for "change") or simply replaced the adjectives with a synonym (not pre-registered but same with Liu et al. (2017), "Students can be bright" as "Students can be shiny"). All participants had a mean RT within three standard deviations of the overall mean, and had at least 50% of their interpretation marked as valid, and so all were included in the analysis. Ninety-nine trials were missed (no response), five trials were excluded as motor errors, 94 trials were marked as invalid meanings. In addition, 20 trials were outliers whose RT was 3SD from participants means per response decision and time constraint: that is for "yes" response, none for short constraint, 0.57% for medium constraint, and 0.36% for long constraint; for "no" response, 0.65% for short constraint, 0.76% for medium constraint, and 0.53% for long constraint. Consequently, 3853 trials were used in analysis: 1553 valid trials for "yes" response, and 2300 for "no" response (acceptance rate 39.30%). For short constraint, the acceptance rate was 36.37%; for mid constraint, 39.46%; and for long constraint, 41.77%.

Response decision. Logistic mixed-effect regression showed only an effect of EoS both in the overall analysis (Table 7) and analyses separated by time constraints (Table 8). As EoS increased, the decision to accept a metaphor as interpretable increased. The odds ratios were 3.97 for the overall analysis, 4.04 under the short constraint, 4.11 under the medium constraint, and 4.02 under the long constraint. Time constraints did not interact with either EoS or LDF.

Table 7

Logistic mixed-effect regression of response decision in Experiment 2

	b	95% CI	p
Intercept	-0.56	-0.97 - -0.14	.009
Long - medium	0.14	-0.07 - 0.35	.203
Medium - short	0.08	-0.14 - 0.3	.464
LDF	0.04	-0.07 - 0.16	.484
EoS	1.38	1.25 - 1.51	$<$ .001
Long - medium x LDF	0.02	-0.21 - 0.25	.889
Medium - short x LDF	-0.04	-0.27 - 0.2	.758
Long - medium x EoS	0.01	-0.23 - 0.25	.916
Medium - short x EoS	-0.05	-0.29 - 0.2	.704
LDF x EoS	-0.01	-0.13 - 0.1	.821
Long - medium x LDF x EoS	0.09	-0.14 - 0.32	.44
Medium - short x LDF x EoS	-0.01	-0.24 - 0.22	.919

Note. b is non-standardised regression coefficient.

Response times. The overall mean RT were 1334ms (SD = 790ms) for “yes” responses, and 1120ms (SD = 653ms) for “no” responses. Separated by time constraint, the mean RT for “yes” responses was 1026ms (SD = 322ms) for short constraint, 1239ms (SD = 631ms) for medium constraint, and 1698ms (SD = 1037ms) for long constraint; the mean RT for “no” response was 965ms (SD = 344ms) for short constraint, 1082ms (SD = 567ms) for medium constraint, and 1318ms (SD = 886ms) for long constraint.

Omnibus linear mixed-effect regression showed (Table 9) EoS had a significant

Table 8

Logistic regression model of response decision per time constraint in Experiment 2

		b	95% CI	p
Short	Intercept	0.65	-1.1 - -0.21	.004
	LDF	0.06	-0.11 - 0.24	.455
	EoS	1.40	1.17 - 1.62	$<$ .001
	LDF x EoS	-0.06	-0.23 - 0.1	.458
Medium	Intercept	-0.60	-1.03 - -0.18	.005
	LDF	0.02	-0.17 - 0.2	.860
	EoS	1.42	1.18 - 1.65	$<$ .001
	LDF x EoS	-0.05	-0.23 - 0.13	.568
Long	-0.43	-0.89 - 0.04	.072	
	LDF	0.10	-0.07 - 0.27	.267
	EoS	1.39	1.17 - 1.61	$<$ .001
	LDF x EoS	0.01	-0.16 - 0.18	.908

Note. b is non-standardised regression coefficient.

positive effect for “no” response on the RT regardless of LDF or time constraints, which was reversed for “yes” response. The effect was larger for “yes” than “no” response. The effect of LDF was contained in its interaction with response with a negative term, showing that although the effect of LDF was not significant for “no” response, it was indeed opposite for “yes” and “no” responses and was larger for “yes” response than “no” response.

After the omnibus analysis, the RT data was split first by response decisions and then by time constraint, same as Experiment 1. When separated by response

Table 9

Omnibus linear mixed-effect model on RT in Experiment 2

	b	95% CI	*p*
Intercept	1,174.99	1083.89 - 1266.08	$<$ .001
Long - medium	278.95	214.25 - 343.65	$<$ .001
Medium - short	94.56	30.66 - 158.46	.004
Response	179.22	130.63 - 227.82	$<$ .001
LDF	12.79	-15.5 - 41.08	.376
EoS	45.63	15.54 - 75.72	.003
Long - medium x Response	153.87	49.07 - 258.66	.004
Medium - short x Response	158.39	51.06 - 265.71	.004
Long - medium x LDF	-26.79	-92.81 - 39.24	.427
Medium - short x LDF	29.89	-36 - 95.79	.374
Response x LDF	-51.63	-95.35 - -7.91	.021
Long - medium x EoS	41.46	-28.48 - 111.41	.245
Medium - short x EoS	-25.86	-94.27 - 42.55	.459
Response x EoS	-194.84	-239.83 - -149.85	$<$ .001
LDF x EoS	-18.41	-45.81 - 9	.188
Long - medium x Response x LDF	-13.21	-120.08 - 93.67	.809
Medium - short x Response x LDF	8.69	-98.76 - 116.13	.874
Long - medium x Response x EoS	-86.62	-192.41 - 19.17	.109
Medium - short x Response x EoS	-50.12	-157.34 - 57.1	.360
Long - medium x LDF x EoS	-0.78	-64.34 - 62.78	.981
Medium - short x LDF x EoS	-23.22	-85.74 - 39.31	.467
Response x LDF x EoS	37.61	-4.03 - 79.25	.077 †
Long - medium x Response x LDF x EoS	-56.53	-155.62 - 42.56	.264
Medium - short x Response x LDF x EoS	25.92	-73.3 - 125.14	.609

Note. *b* is non-standardised regression coefficient.

decisions (Table 10), EoS had a negative effect on “yes” RT. That is, the easier it was considered to simulate a metaphor, the faster people accepted it as interpretable. Also, the effect of LDF was negative too at medium and long constraints (i.e., the higher the LDF was, the faster people accepted the metaphor as interpretable). Both effects were as predicted. EoS’s effect did not increase significantly as the time constraints became longer, but only showed a slight trend between short and medium constraints (Table 11 and Figure 4). LDF interacted with time constraints between medium and long constraints, that is: the effect of LDF was larger at the long constraint than the medium and the short constraints, which was the opposite of what was predicted by the linguistic shortcut hypothesis.

Table 10

Regression model on RT in Experiment 2 per response decision

	Yes			No		
	b	95% CI	*p*	*b*	95%CI	*p*
Intercept	1,337.16	1241.73 - 1432.58	$< .001$	1,191.85	1095.16 - 1288.54	$< .001$
Long - medium	400.77	325.09 - 476.44	$< .001$	264.85	210.39 - 319.32	$< .001$
Medium - short	210.34	132.17 - 288.52	$< .001$	104.66	50.3 - 159.01	$< .001$
LDF	-40.00	-73.29 - -6.72	.020	16.37	-8.97 - 41.71	.206
EoS	-163.58	-198.81 - -128.36	$< .001$	51.48	25.33 - 77.63	$< .001$
Long - medium x LDF	-77.96	-155.21 - -0.71	.048	-27.59	-82.25 - 27.07	.323
Medium - short x LDF	24.51	-52.9 - 101.92	.535	36.05	-18.7 - 90.81	.197
Long - medium x EoS	-54.27	-132.74 - 24.21	.176	31.58	-25.61 - 88.76	.279
Medium - short x EoS	-70.08	-151.5 - 11.34	.092 †	-16.56	-72.61 - 39.5	.563
LDF x EoS	16.12	-16.71 - 48.95	.337	-16.82	-40.45 - 6.8	.163
Long - medium x LDF x EoS	-49.26	-125.16 - 26.64	.204	3.10	-48.57 - 54.77	.906
Medium - short x LDF x EoS	-1.07	-77.47 - 75.34	.978	-21.56	-72.21 - 29.08	.404

Note. *b* is non-standardised regression coefficient.

For “no” RT, only EoS had a strong and consistent effect (Table 10). The harder it was considered to simulate a metaphor, the faster people rejected it as uninterpretable. The effect held true for all levels of time constraints. In contrast, LDF did not have an effect at all on “no” RT (Table 11 and Figure 4).

Table 11

Regression model on RT in Experiment 2 per response decision per time constraint

		Yes			No		
		b	95% CI	*p*	*b*	95%CI	*p*
Short	Intercept	1,033.83	985.88 - 1081.79	< .001	1,008.10	945.59 - 1070.6	< .001
	LDF	-6.46	-32.3 - 19.38	.625	4.43	-16.94 - 25.79	.685
	EoS	-69.95	-97.43 - -42.46	< .001	40.44	18.08 - 62.8	< .001
	LDF x EoS	2.94	-23.15 - 29.02	.825	-4.42	-24.19 - 15.36	.662
Medium	Intercept	1,264.75	1149.89 - 1379.61	< .001	1,143.92	1033.15 - 1254.7	< .001
	LDF	-19.27	-61.34 - 22.79	.37	26.48	-5.85 - 58.8	.109
	EoS	-124.02	-169.45 - -78.6	< .001	37.90	3.9 - 71.89	.029
	LDF x EoS	19.11	-21.4 - 59.63	.356	-11.95	-42.16 - 18.26	.439
Long	Intercept	1,649.91	1474.25 - 1825.57	< .001	1,466.99	1254.52 - 1679.45	< .001
	LDF	-83.54	-151.31 - -15.77	.016	12.67	-33.37 - 58.7	.59
	EoS	-202.97	-274.98 - -130.95	< .001	90.48	42.24 - 138.72	< .001
	LDF x EoS	-10.02	-76.72 - 56.68	.768	16.20	-59.5 - 27.11	.464

Note. *b* is non-standardised regression coefficient.

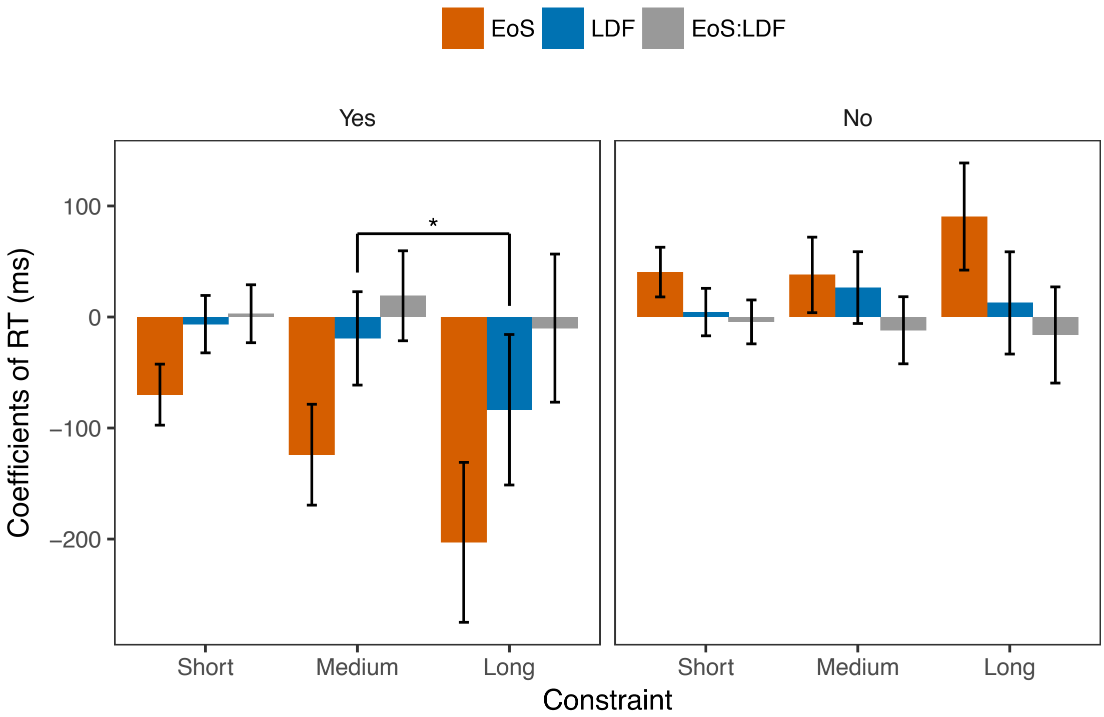


Figure 4. Regression coefficients of RT per response decision in Experiment 2 (i.e. interpretation generation). Error bars stand for 95% CI. Asterisks stand for p-values of interaction between variables and time constraint. *: $p < .05$.

Summary. EoS had a significant effect on response decision and both "yes" and "no" RT as predicted and demonstrating a similar pattern to the sensibility judgement task in Experiment 1. The easier it was to simulate a metaphor, the more likely people accepted it as interpretable, the faster they could think of a meaning, and the longer it took to reject as uninterpretable. Although EoS trended toward the prediction of having a greater effect at longer time constraint, the effect was not reliable. The lack of the time-constraint effect may simply be a result of the coding scheme adopted by my present study. As we always contrasted the two closest time constraints in both experiments (i.e., difference between long and medium constraints, difference between medium and short constraints), the coding could be overly conservative. If the long constraint could be contrasted with the short constraint, we might have found a significant increase of the effect of EoS in this experiment. Further examination of the data could address this problem (which could not be done for the lack of time).

On the other hand, the effects of LDF were mixed. As predicted, LDF had a negative effect on "yes" RT: that is, when words in a metaphor often co-occurred, people were faster to accept it as interpretable. However, there was no evidence that shorter time constraints boosted the effect of LDF, contrary to our prediction, and the effect of LDF was instead *reduced* at shorter time constraints. Finally, LDF had no effect on response decisions nor on "no" RTs.

Cross-Experiment Comparison

To test our cross-experiment hypotheses, we conducted a meta-analysis of the data from Experiments 1 and 2. We expected EoS to have a larger effect for deep interpretation generation than for shallow sensibility judgement, while the effect of LDF would be the other way round.

Method

The data from Experiments 1 and 2 were combined for analysis: response decision in logistic mixed-effects regression, and response time in linear mixed-effects regression. Task was employed as a categorical variable (coded 0 for sensibility judgement and 1 for interpretation generation). In analysis of response decision, EoS and LDF from two datasets were first combined and then centred, while in analysis of response time, they were first combined and then orthogonalised. Levels of time constraint were also included in the models with backward contrast coding, similar to Experiments 1 and 2: that is, Contrast 1 referred to long constraint minus medium constraint (regardless of tasks), and Contrast 2 referred to medium minus short constraint (regardless of tasks).

Results and Discussion

Response decision. Results of the logistic mix-effect model was shown in Table 12. Task enhanced the effect of EoS when LDF was at its mean, such that EoS had a larger, more positive effect in the deep task than the shallow task. That is, people were more likely to accept a metaphor that was typically considered easy to simulate, and even more likely to accept such a metaphor in the deep interpretation generation task than in the shallow sensibility judgement task. EoS was also moderated by time constraint during the shallow task, such that it had a larger, more positive effect at medium time constraint than short time constraint (as predicted); but this moderation effect of time constraint did not appear during the deep task, such that, in the deep task, the effect of EoS did not increase because of time constraint. Finally, LDF had little effect on response decision with a non-significant trend for LDF positively affect chance of acceptance at higher levels of EoS.

Response time. Since the predictors had effects in opposite directions for "yes" and "no" responses, we analysed RT separated by the response (Table 13). For both "yes" and "no" responses, EoS had a larger effect for the deep task than the shallow task. In other words, people relied on embodied simulation more during deeper processing than shallower processing as predicted, which held true across the time constraints. The effect of LDF appeared for "yes" response regardless of EoS, but this effect was conditional to task. That is, LDF only had a significant effect in the deep interpretation generation task but not in the shallow sensibility judgement task, which was opposite to the hypothesis that shallow processing should encourage the using of the linguistic shortcut. Apart from this effect, LDF did not have any other effects. It did not affect "no" RT. Nor was it affected by time constraints. All these null effects did not support the hypotheses concerning the linguistic component.

Table 12

Logistic mixed-effect regression of response decision for cross-experiment comparison

	b	95% CI	p
Intercept	-0.65	-1 - -0.3	$<$ .001
Task	0.12	-0.37 - 0.6	.64
Long - medium	-0.01	-0.18 - 0.16	.908
Medium - short	-0.01	-0.18 - 0.16	.922
LDF	0.06	-0.03 - 0.15	.177
EoS	1.18	1.09 - 1.28	$<$ .001
Task x long - medium	0.14	-0.13 - 0.41	.310
Task x medium - short	0.09	-0.18 - 0.36	.519
Task x LDF	-0.01	-0.13 - 0.1	.817
Long - medium x LDF	-0.08	-0.26 - 0.1	.389
Medium - short x LDF	-0.01	-0.19 - 0.18	.926
Task x EoS	0.13	0 - 0.26	.043
Long - medium x EoS	-0.12	-0.31 - 0.07	.231
Medium - short x EoS	0.38	0.19 - 0.57	$<$ .001
LDF x EoS	0.08	-0.01 - 0.17	.081 †
Task x long - medium x LDF	0.08	-0.21 - 0.37	.567
Task x medium - short x LDF	-0.02	-0.31 - 0.27	.885
Task x long - medium x EoS	0.12	-0.18 - 0.42	.418
Task x medium - short x EoS	-0.42	-0.73 - -0.12	.006
Task x LDF x EoS	-0.09	-0.21 - 0.03	.143
Long - medium x LDF x EoS	-0.02	-0.21 - 0.16	.816
Medium - short x LDF x EoS	-0.04	-0.22 - 0.15	.687
Task x Long - medium x LDF x EoS	0.12	-0.17 - 0.41	.419
Task x Medium - short x LDF x EoS	0.02	-0.27 - 0.31	.877

Note. b is non-standardised regression coefficient.

Table 13

Regression model on RT for cross-experiment comparison per response decision

	Yes			No		
	b	95% CI	*p*	*b*	95%CI	*p*
Intercept	867.38	792.5 - 942.25	$< .001$	850.87	773.07 - 928.67	$< .001$
Task	464.53	361.3 - 567.75	$< .001$	340.71	233.09 - 448.34	$< .001$
Long - medium	145.38	98 - 192.77	$< .001$	131.83	98.42 - 165.24	$< .001$
Medium - short	175.16	126.96 - 223.36	$< .001$	149.05	114.96 - 183.14	$< .001$
LDF	-7.49	-28.41 - 13.43	.483	4.14	-11.59 - 19.87	.606
EoS	-48.87	-70.57 - -27.16	$< .001$	16.00	0.07 - 31.93	.049
Task x long - medium	251.98	179.8 - 324.15	$< .001$	132.82	80.75 - 184.9	$< .001$
Task x medium - short	35.35	-38.74 - 109.45	.35	-45.48	-97.94 - 6.98	.089 †
Task x LDF	-33.74	-63.62 - -3.87	.027	11.97	-9.39 - 33.34	.272
Long - medium x LDF	-19.97	-67.13 - 27.19	.407	3.73	-30.08 - 37.53	.829
Medium - short x LDF	1.51	-46.02 - 49.05	.95	-0.40	-34.95 - 34.14	.982
Task x EoS	-114.82	-146.56 - -83.08	$< .001$	35.84	13.48 - 58.19	.002
Long - medium x EoS	-71.49	-120.32 - -22.67	.004	-4.33	-39.24 - 30.57	.808

Medium - short x EoS	-37.31	-85.69 - 11.07	.131	16.61	-18.5 - 51.72	.354
LDF x EoS	-2.21	-22.93 - 18.5	.834	-11.90	-26.4 - 2.59	.108
Task x long - medium x LDF	-60.91	-134 - 12.17	.102	-31.28	-83.6 - 21.03	.241
Task x medium - short x LDF	23.41	-50.06 - 96.88	.532	38.73	-14.14 - 91.61	.151
Task x long - medium x EoS	17.88	-56.9 - 92.67	.639	36.66	-17.87 - 91.19	.188
Task x medium - short x EoS	-32.49	-108.7 - 43.71	.403	-32.07	-86.11 - 21.97	.245
Task x LDF x EoS	18.38	-11.53 - 48.3	.228	-5.27	-25.46 - 14.92	.609
Long - medium x LDF x EoS	-0.80	-49.61 - 48	.974	-18.92	-50.63 - 12.79	.242
Medium - short x LDF x EoS	-5.26	-51.91 - 41.39	.825	7.30	-25.27 - 39.87	.661
Task x Long - medium x LDF x EoS	-46.18	-119.68 - 27.31	.218	22.38	-26.82 - 71.58	.373
Task x Medium - short x LDF x EoS	3.57	-68.93 - 76.08	.923	-28.88	-78.09 - 20.34	.250

Note. b is non-standardised regression coefficient.

General Discussion

The current study was conducted to replicate the previous study on the conceptual representation during metaphor processing and answer the specific question, whether limited time resources would affect the conceptual representation. Our findings confirmed the hypothesis that both embodied and linguistic components existed as parts of a conceptual representation. The embodied component played a prominent and consistent role in the representation of meaning. The easier a metaphor was considered to simulate, the easier it was to accept it as either sensible or interpretable; the harder it was considered to simulate, the easier it was to reject it. We also found, as predicted, that the embodied component was more engaged during the deep processing of interpretation generation than during the relatively shallower processing of sensibility judgement, and it was also generally more engaged when there was more processing time available. Thus, it confirmed the suggestion of grounded representation that the embodied component was more costly and time-consuming (Barsalou et al., 2008; Connell & Lynott, 2014).

In contrast, evidence for the linguistic component and the linguistic shortcut hypothesis was limited. We found evidence that the linguistic component affected metaphor processing independently of ease of simulation in some cases. The more often two constituent words of a metaphor co-occurred, the easier (at least faster) it was to accept it as interpretable (which was not found for shallow sensibility judgement). However, the linguistic component did not have a greater effect when the time resources were restricted as we hypothesised. In fact, we found that people were more reliant on the linguistic component when they had a longer interval for response.

Complex Processing Strategy during Metaphor Processing

The surprising results cast doubts on the linguistic shortcut hypothesis we originally proposed. Neither time constraints nor shallow processing promoted the use of the linguistic shortcut. In fact, the reverse seemed to be true. People were more likely to incorporate linguistic information into processing when they had plenty of time, or when they needed to engage in deep processing. Therefore, we argue that linguistic information was involved in metaphor processing with a complex strategy, depending on the requirement of the task.

We found that linguistic information affected processing directly under a specific condition. That is, people relied on lingusitic information when they were required to engage in deep processing. Specifically, linguistic information could influence the speed of acceptance by itself, which in a way supports the idea that the linguistic information could be used heuristically (Connell, 2018; Louwerse & Jeuniaux, 2008). However, people were also allowed for longer time to perform the deep processing task than the shallow processing task, and people relied on linguistic information more when the time constraint was most relaxed. These two effects both indicated that linguistic information contributed to the performance of metaphor processing at a later stage, rather than immediately after the processing began as the LASS theory proposed (Barsalou et al., 2008).

We also found that linguistic information affected processing indirectly when the required depth of processing was shallow. It could signal whether the processing was likely to succeed. When the distributional frequency was high, the processing should succeed. Therefore, if people could not come up with a coherent representation for the meaning (e.g., *Supply can be bright*), it would take them even longer to reject the metaphor than when the distributional frequency was low, which suggested that the processing was unlikely to succeed (e.g., *Illness can be bright*). In other words, high distributional frequency was used to safeguard against premature rejection of a metaphor. When the distributional frequency was low, in constrast, people simply judged the metaphors based on its ease of simulation (i.e., *Illness can be bright* was rejected faster than *Solutions can be bright*).

In general, we did not find evidence to support our original hypothesis concerning the linguistic component as a shortcut for metaphor processing. How much could this lack of evidence be generalised beyond this particular study and beyond metaphor processing? There are three reasons why the methods of the current study may not be sensitive enough to detect the effect of linguistic distributional patterns. First, LDF was a minimal redition of the linguistic component. LDF uses the n-gram distributional frequency, which measured how often two words co-occurred in the same context (e.g., *solution* and *bright*). It was different from other vector-based distributional models (e.g., latent semantic analysis) which measured how often words occurred in similar context. For example, although *solution* and *bright* do not co-occur frequently in the same context, thus having a low n-gram frequency, *solution* frequently co-occurred in similar contexts as words like *question*, *answer* and *students.* Therefore, the vector-based distributional models might better at capturing the distributional pattern for metaphor processing, which should be explored further in future studies. Second, the null effects raise the question whether the evidence suggests **against the linguistic shortcut hypothesis. It is impossible to clarify this point with null-hypothesis significance testing (NHST). To examine whether the evidence were equivocal or against the effect of linguistic component during shallow processing and shorter time constraints, the data needs to be reanalysed with Bayesian models.**

Third, in order to show whether the linguistic component plays a heuristic role, it is not enough to study the end-products of metaphor

processing (i.e., RT and response decisions) alone. LDF could have been activated and influential to processing from the beginning of a trial, but by the time that a response was made (at least 200ms after the adjective onset), the initial shortcut effect might have faded. To fully examine the role of the linguistic component during online processing, measurements with high temporal resolution such as EEG and MEG should be taken.

In this paper, we viewed metaphor processing from the perspective of conceptual representations during language processing in general. We argued that metaphor processing is an aspect of language processing and should involve similar conceptual representations. However, our findings concerning metaphor processing may not be able to generalise to language processing in general because metaphors are different from literal language in at least the following two ways. Firstly, distributional patterns of constituent words are not often reliable during metaphor processing, because it is common to read metaphors that are perfectly sensible and meaningful whose constituent words rarely co-occur (e.g., *abrasive personality*, *lukewarm supporters*). Therefore, people might deem the sacrifice of accuracy to be too great if they were to take the linguistic shortcut. Secondly, metaphors may encourage people to engage in the simulation of affective experiences more than their literal counterparts (Citron & Goldberg, 2014). As a result, people put a much heavier reliance on the embodied component than during literal language processing. Having the time constraints meant that full engagement of embodied simulation was not always possible. Therefore, people might have switched their processing strategies altogether to focus on the lower-level lexical variables such as lexical frequencies or word lengths. Therefore, it is necessary for future studies to compare literal language

and metaphor processing in these aspects if a general theory of conceptual representation is to be achieved.

Conclusions

In conclusion, we found that metaphor processing relies on the unique contributions of embodied simulation and linguistic distributional patterns. These two components both inform conceptual representation and could cause conflict when there is no agreement between them. Combined with previous research on metaphor processing, literal language processing and conceptual combination, these findings indicate that conceptual representation involves a complex interaction between the two components, and flexibly constructed based on the demand of the task.

References

Allport, D. A. (1985). Distributed memory, modular subsystems and dysphasia. In S. Newman & R. Epstein (Eds.), *Current perspectives in dysphasia* (pp. 207–244). Edinburgh: Churchill Livingstone. Retrieved from https://www.scribd.com/document/17676263/Allport-1985-Distributed-Memory-Modular-Subsystems-and-Dysphasia http://scholar.google.com/scholar?hl=en{\&}btnG=Search{\&}q=intitle:Distr

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. doi:10.1037/a0016261

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics By 8-Month-Old Infants. *Psychological Science*, *9*(4), 321–324.

Auguie, B. (2017). *GridExtra: Miscellaneous functions for "grid" graphics*. Retrieved from https://CRAN.R-project.org/package=gridExtra

Aust, F. (2016). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from https://CRAN.R-project.org/package=citr

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Bache, S. M., & Wickham, H. (2014). *Magrittr: A forward-pipe operator for r*. Retrieved from https://CRAN.R-project.org/package=magrittr

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). doi:10.1016/j.jml.2012.11.001

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*(4), 577–609; discussion 610–60. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1693222{\&}tool=

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In A. M. Glenberg & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 245–283). Oxford, England: Oxford University Press. Retrieved from http://psychology.emory.edu/cognition/barsalou/papers/Barsalou{_}et{_}a

Bartoń, K. (2017). *MuMIn: Multi-model inference.* Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., & Maechler, M. (2017). *Matrix: Sparse and dense matrix classes and methods.* Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bonner, M. F., & Grossman, M. (2012). Gray matter density of auditory association cortex relates to knowledge of sound concepts in primary progressive aphasia. *Journal of Neuroscience*, *32*(23), 7986–7991. doi:10.1523/JNEUROSCI.6241-11.2012

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., & Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, *46*(2), 743–756. doi:10.1016/j.neuropsychologia.2007.10.007

Brants, T., & Franz, A. (2006). Web 1T 5-gram Version 1 LDC2006T13. DVD. Philadelphia: Linguistic Data Consortium.

Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical Sentences Are More Emotionally Engaging than Their Literal Counterparts. *Journal of Cognitive Neuroscience*, *26*(11), 2585–2595. doi:10.1162/jocn_a_00654

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Mahwah, NJ: Erlbaum. Retrieved from https://books.google.co.uk/books?id=fAnSOgbdFXIC

Connell, L. (2018). What have labels ever done for us? The lingusitic shortcut in conceptual processing. *Language, Cognition and Neuroscience.* doi:10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2013). Flexible and fast: linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, *20*(3), 542–50. doi:10.3758/s13423-012-0368-x

Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*(3), 390–406. doi:10.1111/tops.12097

Dantzig, S. van, Cowell, R. A., Zeelenberg, R., & Pecher, D. (2011). A

CHAPTER 6

ERP Markers of Embodied and Linguistic Components in a Property Verification Task

To examine the interplay of embodied and linguistic components in metaphor processing, the timecourse of these two components will be studied in the next two chapters. Chapter 6 is a sidebar, which fills in the gap in the literature concerning the ERP markers of the two components during literal language processing. I will replicate a well-established effect of embodied simulation, which is the modality switching costs in the property verification task; and building on existing findings, I will explore how embodied simulation and linguistic distributional patterns independently affect the ERP of the modality switching costs.

The modality switching costs refer to a delay in response to a property of a certain modality (e.g., vision: *The SUN is bright*) in a property verification task if the previous property verified is of a different modality (e.g., audition: *A BLENDER is loud*) than if it is of the same modality (e.g., vision: *A LEMON is yellow*). The costs were thus interpreted as evidence for the recruitment of modality-specific resources during language processing (Pecher et al., 2003). However, Louwerse & Connell (2011) found that the costs could at least partially be attributed to the distributional patterns of the property words. The fact that *loud* $\rightarrow$ *bright* takes longer to process than *yellow* $\rightarrow$ *bright* is also due to the fact that *yellow* and *bright* co-occurred more frequently than *loud* and *bright*. The linguistic distributional pattern was found to account for faster responses better than slower responses, supporting the linguistic shortcut hypothesis. As a result, in this chapter, we also expect the linguistic component to act as a shortcut. In terms of timecourse, the linguistic component should reach the peak of its activation before the embodied component.

In spite of our prediction concerning the difference in the timecourse of the two components, as well as previous ERP studies that found various ERP components associated with the modality switching costs (Bernabeu et al., 2017; Collins et al., 2011; Hald et al., 2013, 2011), the study in this chapter will be of an exploratory nature, in which the ERP of property verification will be analysed in uniformed 50ms slices, electrode by electrode. The reason to conduct a component-free analysis like this is, first of all, to limit researcher's degrees of freedom as discussed in Chapter 2. Moreover, an exploratory analysis will give us a fuller picture concerning how embodied and linguistic components were activated, which could be compared with the activations during metaphor processing found in Chapter 7.

An abstract of this work has been submitted to and accepted as a talk in the Embodied and Situated Language Processing (ESLP) conference, August 2018.

Paper 4: Tracing real-time perceptual and linguistic modality switching costs in property verification task

Page numbers: 215 - 270

Continuous neural activations of simulation-linguistic representations in modality switching costs

Pei Q. Liu[1] & Louise Connell[1] & Dermot Lynott[1]

[1] Department of Psychology, Lancaster University

Author Note

Department of Psychology, Lancaster University, UK.

Correspondence concerning this article should be addressed to Pei Q. Liu, Department of Psychology, Fylde College, Lancaster University, Bailrigg, LA1 4HX, UK. E-mail: p.liu1@lancaster.ac.uk

Abstract

Modality switching costs (MSCs) are one of the classic effects that support the embodied views of conceptual representations. They refer to a delay in response time to verify a sensory property of a certain perceptual modality (e.g., visual: SUN – bright), when the previous sensory property has been of a different modality (e.g., auditory: BLENDER – loud) compared to a property of the same modality (e.g., visual: ROSE – red). Such costs indicate that conceptual representations require the recruitment of modality-specific resources. However, MSCs could also result from the distributional pattern of property words: the reason why loud → bright takes longer than red → bright could be because bright and loud do not co-occur in the same linguistic context as frequently as bright and red. In the present study, we examined how well MSCs were predicted by an embodied model (switch / no-switch between perceptual modalities) versus a linguistic model (switch / no-switch between linguistic distributional clusters), in behavioural (RT) and continuous event-related EEG potentials (ERP) paradigms. The behavioural data supported the linguistic model in explaining MSCs and found MSCs to be moderated by the target modality, the ERPs showed that linguistic distributional pattern played a crucial role in the neural activations of MSCs. What used to be found as a result of perceptual switching (e.g., "early N400" effect) could be better explained by the linguistic model. The embodied component was activated later than the linguistic component, accounting for activations associated with semantic representation (typically in N400 area). Later during processing, both components were active for decision making (often manifested as LPC).

Keywords: modality switching costs, embodied simulation, linguistic distributional pattern, EEG

Word count: 7705 words

Continuous neural activations of simulation-linguistic representations in modality switching costs

Embodied cognition (also called grounded or situated cognition) in its most consensual form suggests that conceptual processing involves the simulation (i.e., partial reactivation) of sensory, motor, affective, introspective and other bodily experiences of concepts (Barsalou, 1999). When we process a sentence such as *The sun is bright*, we "relive" the bodily experiences with the sun and specifically the visual experience of its brightness. The modality switching costs (MSCs; Pecher, Zeelenberg, & Barsalou, 2003) showed that people's response to a word pair such as *SUN* - *bright* would be delayed if the previous word pair they verified had pertained to a different sensory modality (e.g., *BLENDER* - *loud*) compared to the same modality (e.g., *ROSES* - *red*). Such costs were interpreted as a result of engaging modality-specific resources. People were slower to respond to *bright* because their attention had been previously engaged in the auditory modality when processing *loud*. Further studies also found MSCs across linguistic and perceptual stimuli (Van Dantzig, Pecher, Zeelenberg, & Barsalou, 2008) and during novel concept creation (Connell & Lynott, 2011), again confirming that people "relive" the sensory experiences (e.g., "seeing" the sun's light, "hearing" a blender's noises) when they verify these concepts.

Meanwhile, the linguistic account of conceptual representation proposes that the meaning of a word is encoded (at least partially) in the distributional pattern of the linguistic symbols (Landauer & Dumais, 1997). As Firth (1957) famously put it, "you shall know a word by the company it keeps" (p.11). Linguistic distributional patterns can encode relationship of objects, events and matters to a certain degree (Louwerse, 2011). The meaning of *sun* could be understood by its semantic neighbours such as *sky*, *hot*, *bright*, etc. This linguistic account provides another interpretation for MSCs. The reason why people were faster to respond to *bright* after *red* than after *loud* was that *bright* and *red* co-occurred in the same context more frequently than *bright* and *loud*. Indeed the co-occurrence frequencies could differentiate the sensory modality of perceptual words to some extent (Louwerse & Connell, 2011). A corpus study illustrated that the co-occurrence frequencies of sensory property words (e.g., *bright*, *loud*, etc.) varied along three dimensions, which broadly corresponded to the perceptual property of vision-haptics, olfaction-gustation, and audition. Switch between the linguistic dimensions were capable of accounting for the faster responses of MSCs, while switch between the embodied dimensions were capable of accounting for the slower responses.

Embodied-Linguistic Conceptual Representations

Many researchers have explicitly stated that conceptual representations are both embodied and linguistic (Andrews, Vigliocco, & Vinson, 2009; Barsalou, Santos, Simmons, & Wilson, 2008; Connell & Lynott, 2014; Louwerse & Jeuniaux, 2008). Such a combinatorial account can bridge the gap between the two accounts, and thus help avoid the problems faced individually by either. At the linguistic end, the extreme form of this linguistic account needs to face the challenge of the grounding problem (Harnad, 1990), which questions how the linguistic symbols link to its real-world referents. If the meaning of a word is simply defined by other words, it is hard to say that any words have meanings at all. This problem is perfectly illustrated by Searle (1980)'s rendition Chinese room argument. Suppose a person who does not know Chinese sits in a closed room, and receives input in Chinese from a slot. He follows an instruction book that produces Chinese characters as response to the input. Suppose the instructions are sufficiently well written, and the responses so convincing that the person who gives the Chinese input thought the person in the room was really Chinese. However, the case remains that the person in the room does not know the meaning of any of the Chinese words he produces. In other words, simply knowing the distributional patterns between words is not the same as knowing a language. The combinatorial account allows words to be grounded in the brain systems of perception, action and emotion (Zwaan, 2014), though the degree to which each linguistic symbol is grounded is still a contended issue (Chemero, 2009; Harnad, 1990; Kintsch & Dijk, 1978; Öttl, Dudschig, & Kaup, 2017).

On the embodied side, the combinatorial account, among many of its benefits, affords a degree of flexibility to embodied simulation (Zwaan, 2014). As language comprehension does not only occur on the level of individual words and sentences, different representations are created based on the environment, context and goal of processing (Connell & Lynott, 2014; Lebois, Wilson-Mendenhall, & Barsalou, 2015; Louwerse & Jeuniaux, 2010). One way that the linguistic component affords more flexible conceptual representation is by providing a shortcut for embodied simulation (Connell, 2018). Theories of embodied language processing, such as the Language and Situated Simulation (LASS; Barsalou et al., 2008) theory and the Symbol Interdependency theory (SIH; Louwerse, 2011), suggest that the linguistic component is more speedy and less costly, and thus peaks prior to the embodied component (although embodied simulation could be activated very quickly as shown by Hauk, Johnsrude, and Pulvermüller (2004) for example). Since the linguistic component can partially encode bodily experiences in the real world, it is useful as an indicator for whether future processing will be successful, in other words whether it is worth expending energy on the more costly embodied simulation (Connell & Lynott, 2013).

Louwerse and Connell (2011) examined whether the MSCs could be accounted for by both simulation and linguistic information, and found that the linguistic component had indeed a temporal advantage. While a switch between the perceptual modalities (e.g., haptics → vision: *warm* → *bright*) better accounted for the switching costs when people responded slowly, a switch between the linguistic dimensions (e.g., audition → haptic-vision: *loud* → *bright*) better accounted for the costs when people responded quickly. In other words, people relied on the coarse-grained difference between the linguistic distributional patterns as a shortcut to enable more speedy response. A necessary condition for the linguistic component to be a shortcut is for the linguistic component to reach the peak of its activation before the embodied component. Support for this condition requires the examination of the continuous activations of both components during online processing, which is the aim of the present study.

Present Study

The present study sets out to test whether embodied and linguistic components both contribute to MSCs, and explore their activations during online processing using event-related EEG potentials (ERP). ERP is a segment of EEG waveforms time-locked to an event, which in the case of MSCs is the presentation of the property word (e.g., *bright*; Otten & Rugg, 2005). ERP is often characterised as discrete components, which is a segment of waveform that is functionally significant, circumscribed to a certain time window. Previous studies have found many ERP components associated with MSCs, most prominent of which were N400 and P600 or late positivity complex (LPC; Bernabeu, Willems, & Louwerse, 2017; Collins, Pecher, Zeelenberg, & Coulson, 2011; Hald, Hocking, Vernon, Marshall, & Garnham, 2013; Hald, Marshall, Janssen, & Garnham, 2011). A perceptual modality switch often elicited a greater negativity around 400ms from the property onset and greater positivity after 600ms from the property onset, which was interpreted as a difference in semantic representation and decision making respectively.

However, the findings of these components have been highly inconsistent. The N400 effect alone was found in the posterior region in some cases (Bernabeu et al., 2017; Hald et al., 2013), while in anterior region in others (Collins et al., 2011, Hald et al. (2011)). In some cases, it was found in the typical window around 400ms (Collins et al., 2011), while in others an earlier window from 270ms to 370ms (Bernabeu et al., 2017; Hald et al., 2011). Furthermore, different ERP components were also found to be associated with different perceptual modalities. According to Collins et al. (2011), the N400 effect was manifested only for visual properties, but in other studies it was found for properties of all sensory modalities. These inconsistencies are most likely due to the complex nature of semantic processing, which involves not only conceptual representation but from sublexical processing to integration of semantic representations (Hauk, 2016). These previous studies did not only examine the effect of modality switching but also other linguistic factors such as negated sentences, which could easily distort observed waveforms (Luck, 2005). Furthermore, these ERP studies so far have not considered the linguistic distributional patterns as a cause for MSCs. Therefore, the present study will focus on the switching costs alone, accounting for it by either a switch between perceptual modalities or between linguistic dimension.

To achieve a better control and maximise the effect, we conducted a behavioural study first with only visual, auditory and haptic words (Experiment 1). According to Louwerse and Connell (2011), linguistic dimensions are more coarser-grained than the perceptual modalities. While there are five major sensory modalities (i.e., vision, audition, haptics, gustation, olfaction), words describing sensory experiences often fall into three linguistic clusters. For example, visual words (e.g., *bright*) and haptic words (e.g., *warm*) often co-occur in close proximity, thus belonging to the same linguistic cluster/dimension. As a result, verifying *warm* after *bright* constitutes a switch between perceptual modalities, but not a switch between lingusitic dimensions. Thus, we are able to differentiate the effect of a perceptual switch from a linguistic switch. We hypothesised that both a perceptual switch and a linguistic switch could account for the switching costs in this experiment. From Experiment 1, we selected the stimuli that reliably produced either perceptual or linguistic MSCs (or both) to be used in Experiment 2, the ERP study. Our study will be largely exploratory because of the inconsistency in previous findings, and the inability to pinpoint the time windows and the scalp regions of the linguistic dimension switch. Thus, only tentative hypotheses could be proposed, which are 1. the linguistic switch should affect the waveform prior to the embodied switch; 2. the embodied switch should mainly affect the waveform around 400ms.

Experiment 1: Behavioural Modality Switching Costs

By studying the modality switching costs, it is possible to identify the separate contributions of embodied and linguistic components to conceptual representation. Therefore, in this experiment, our aim is to examine the costs of perceptual modality switch and the costs of linguistic dimension switch separately. Participants will read pairs of concepts (nouns) and properties (adjectives), such as *SUN - bright*, and decide whether the property is usually true for the concept. The properties shown will be of three perceptual modalities, vision, audition and haptics, but only two linguistic dimensions based on how the properties group together according to their distributional patterns. Therefore, we expect there to be a cost (i.e., delay in reaction time) when there is a switch between perceptual modalities, as well as between linguistic dimensions. In this section, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Method

Participants. Based on the effect size from Louwerse and Connell (2011) and requiring statistical power of .95 with an alpha level of .05, we estimated the required sample size to be 45 participants using G*Power (Faul, Erdfelder, Buchner,

& Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007). Forty-eight native speakers of English were thus recruited for a payment of £3.50 or course credit (age: 19.82 (1.51); female: 30; right-handed: 39).

Materials. One hundred and fifty concept-property items, such as *SUN* - *bright*, were used in the present study. The property words were selected from the modality exclusivity norms of Lynott and Connell (2009) and Dantzig, Cowell, Zeelenberg, and Pecher (2011), in which the property words were rated according to how strongly they could be experienced by the five major sensory modalities, which are vision, audition, gustation, olfaction and haptics. The property words in the present study were all uni-modal, such that the difference between the ratings of the dominant modality (i.e., the modality with the highest rating) and the modality with the second highest rating was greater than 0.75. These words pertained to three dominant modalities, which are vision ($N = 60$), audition ($N = 60$) and haptics ($N = 30$). They were paired with concept words, for which the properties were not only true but also salient (e.g., *CARNATION* - *red* instead of *CARNATION* - *black*, the latter of which could be true but was not salient), so that the concepts could activate the properties automatically. The raw co-occurrence frequencies between concepts and properties were below 1 million (Netspeak.com) to avoid automatic association between words, and the words were known by more than 75% of the population.

The perceptual modality of the property words was simply the dominant modality. The linguistic dimension of the property words, on the other hand, was determined by the co-occurrence frequencies between every two property words in the same way as Louwerse and Connell (2011). Initially, we collated a set of 244 property words and extracted their bi- to 5-gram frequencies from the Google Web-1T corpus (i.e., two property words e.g, *bright* and *mumbling* with zero, one, two and three intervening words; Brants & Franz, 2006). These frequencies of each pair of properties were summed and transformed logarithmically, resulting in a 244 x 244 matrix. Then we submitted this matrix to PCA with Varimax rotation and Kaissar normalisation and extracted 2 principle components (PCs) that could explain 61.25% of the total variance (PC1: 50.76%; PC2: 10.49%). The property words were thus allocated to either one of these linguistic dimensions based on the larger factor loadings. Furthermore, we correlated the factor loadings with each word's modality strenght and found that PC1 had a significant, positive correlation with the visual strengths of the properties (Table 1), and PC2 had a significant, positive correlation with the auditory strengths; while both components had a weak positive correlation with the haptic strengths.

Each target properties were paired with three different prime properties, so that every property appeared in all three switching conditions, which are a switch between both perceptual modalities and linguistic dimensions, a switch between perceptual modalities only but no switch between linguistic dimensions, or no switch between either perceptual modalities or linguistic dimensions. For instance, for the same target property *bright*, *mumbling* → *bright* caused a switch between both perceptual modalities, that is from audition to vision, and linguistic dimensions, that is from PC2 to PC1; *rough* → *bright* caused a switch only in perceptual modalities, that is from haptics to vision, but not between linguistic dimensions, both of which were of PC1; and *colourful* → *bright* represented neither switch. A further list of 270 fillers was added, among which 45 were true and 225 were false. The true fillers were concept-property items that did not pertain to the three sensory modalities of interest; whereas false fillers were concept-property pairs in which the properties were not usually true for the concept (e.g., *WHISKY* – *fluffy*). Some of the false fillers were highly associated words (e.g., *TEMPLE* – *praying*), so that people could not decide the truthfulness of the concept-property items from the linguistic association of the words alone (Solomon & Barsalou, 2004). The participants also

Table 1

Correlation between the PCs and modality strengths

	PC1	PC2	Visual	Haptic	Auditory
PC1	1.00	.00	.64 **	.12	-.61 **
PC2	–	1.00	-.21 **	.12	.31 **
Visual	–	–	1.00	.18 **	-.72 **
Haptic	–	–	–	1.00	-.42 **
Auditory	–	–	–	–	1.00

Note. ** means $p < .01$

saw 25 practice trials, using true items not featured in the main experiment.

In the experiment, each participant saw the items and fillers in a random order. The visual and auditory items were presented only once either as prime or as target, and the haptic items were presented twice as both prime and target, resulting in 90 prime → target pairs for each participant. Across the experiment, all items were used as both prime and target, leading to 6 lists of stimuli in total (prime/target (2) x switching conditions (3)) to which participants were randomly assigned.

Procedure. The experiment was reviewed by Lancaster University's Department of Psychology Ethics Committee and approved by Lancaster University's Research Ethics Committee. After brief information about the study, the participants were sat normally from a PC screen. Participants received both verbal and written instructions (in the exact wording as reported) that they would see concept-property pairs on the screen, and the task was to decide whether the properties were **usually** true for the concepts. For example, if they saw *CARNATION* - *red*, the answer would be "yes"; whereas if they saw *CARNATION* – *black*, even though it was possible, it was not **usually** true, so the answer would be “no”. All concepts were in the uppercase (“CARNATION”) and properties in the lowercase (“red”). The concepts and the properties appeared one after another in black letters on a grey background (Figure 1).

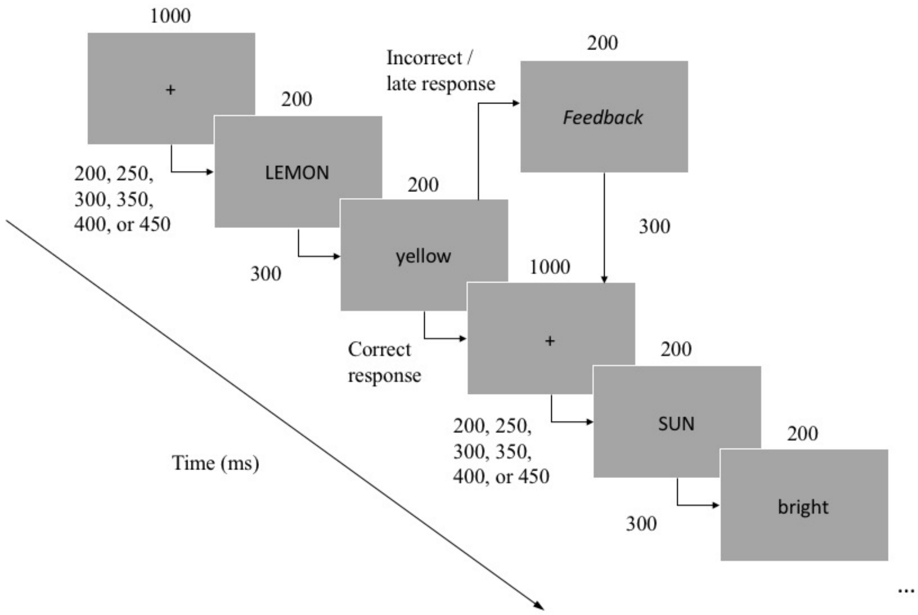


Figure 1. Sequence of presenting the CONCEPT-property pairs. Response timed out after 3000ms after the property onset.

Figure 1 shows the structure of a concept-property item. The trial started with a fixation cross for 1000ms followed by a blank screen of variable length (200ms to 450ms). Then the concept and the property appeared consecutively each for 200ms with a blank of 300ms in between. Once the property word appeared, participants could start to respond by pressing the comma key (“,”) for “yes” or the full stop key (“.”) for “no”. If the response were incorrect, the participants would receive feedback (“Error”). If they failed to respond within a 3-second interval after the onset of the property words, they would receive a feedback (“***TOO SLOW***”). The accuracy and reaction time were recorded. Afterwards, the next trial automatically started.

Analysis and design. Data were analysed and reported using R[1]. Two linear mixed-effect (LME) models were separately built for each type of switch, one for the perceptual modality switch and one for the linguistic dimension switch. The dependent variable was the reaction time (RT) to the target concept-property items from the target property onset. Participants and items were crossed random variables to predict random intercepts,[2] and the target modality (visual, auditory

---

[1] We used R (Version 3.5.0; R Core Team, 2017) and the R-packages *abind* (Version 1.4.5; Plate & Heiberger, 2016), *arm* (Version 1.10.1; Gelman & Su, 2018), *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2015), *bookdown* (Version 0.7; Xie, 2016), *broom* (Version 0.4.4; Robinson, 2017), *coda* (Version 0.19.1; Plummer, Best, Cowles, & Vines, 2006), *contrast* (Version 0.21; Kuhn, Steve Weston, Wing, Forester, & Thaler, 2016), *data.table* (Version 1.11.4; Dowle & Srinivasan, 2017), *doBy* (Version 4.6.1; Højsgaard & Halekoh, 2016), *dplyr* (Version 0.7.5; Wickham, Francois, Henry, & Müller, 2017), *Formula* (Version 1.2.3; Zeileis & Croissant, 2010), *ggplot2* (Version 2.2.1; Wickham, 2009), *gridExtra* (Version 2.3; Auguie, 2017), *Hmisc* (Version 4.1.1; Harrell Jr, Charles Dupont, & others., 2018), *interplot* (Version 0.1.5; Solt & Hu, 2015), *knitr* (Version 1.20; Xie, 2015), *lattice* (Version 0.20.35; Sarkar, 2008), *lme4* (Version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.0.1; Kuznetsova, Brockhoff, & Christensen, 2017), *magrittr* (Version 1.5; Bache & Wickham, 2014), *MASS* (Version 7.3.50; Venables & Ripley, 2002), *Matrix* (Version 1.2.14; Bates & Maechler, 2017), *mgcv* (S. N. Wood, 2003, 2004, Version 1.8.23; 2011; S. Wood, N., Pya, & S"afken, 2016), *multcomp* (Version 1.4.8; Hothorn, Bretz, & Westfall, 2008), *mvtnorm* (Version 1.0.8; Genz & Bretz, 2009), *nlme* (Version 3.1.137; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017), *papaja* (Version 0.1.0.9735; Aust & Barth, 2017), *plyr* (Wickham, 2011; Version 1.8.4; Wickham et al., 2017), *psych* (Version 1.8.4; Revelle, 2018), *purrr* (Version 0.2.5; Henry & Wickham, 2018), *rcartocolor* (Version 0.0.22; Nowosad, 2017), *rms* (Version 5.1.2; Harrell Jr, 2018), *SparseM* (Version 1.77; Koenker & Ng, 2017), *survival* (Version 2.42.3; Terry M. Therneau & Patricia M. Grambsch, 2000), *TH.data* (Version 1.0.8; Hothorn, 2017), and *tidyr* (Version 0.8.1; Wickham & Henry, 2017) for all our analyses and the writing up of this manuscript.

[2] Although model including participants as the only random intercept explained the data worse than the empty model by 843.08 times ($BF_{10} = 0.001$), model with participants and item as crossed random intercepts better explained the data than the model with only participants as random intercept 35.47 times ($BF_{10} = 35.47$).

and haptic, with auditory target as the reference level) and the switching condition (switch coded as 1, non-switch coded as 0), plus their interactions were the fixed predictors. Planned comparisons for each target modality were also run, that is perceptual/linguistic switch versus non-switch for the three target perceptual modalities separately. *p*-values for the planned comparisons were corrected using Tukey's HSD (reported in Supplementary Materials: https://goo.gl/sRkXun).

To examine the effect of switching costs and its interaction with target modality, we conducted model comparisons in a hierarchical fashion. For perceptual modality switch:

Step 1: Models with random predictors and target modality only;

Step 2: Models with random predictors and target modality + perceptual modality switch;

Step 3; Models with random predictors and target modality + perceptual modality switch and their interaction.

For linguistic dimension switch, the same three steps were carried out with linguistic dimension switch instead of perceptual modality switch.

In our study, all model comparisons were conducted using Bayes factors (BF), which denotes the ratio of likelihood of the data under both null and alternative hypothesis (Jarosz & Wiley, 2014). If the BF of H_1 against H_0 (i.e., BF_{10}) is 5, it means that the data is 5 times more likely to occur under H_1 than H_0. BF is superior to the *p*-value in two ways. First, BF could inform us how confident we can be with the hypothesis. We can be more confident with the alternative hypothesis if $BF_{10} = 100$ than if $BF_{10} = 10$. In contrast, *p*-values do not bear any significance to the power of the study, the size of the effect or the quality of the hypothesis (see Wagenmakers, 2007 for detailed discussion concerning *p*-values). Second, we do not need to conduct post hoc corrections of multiple comparisons with BF. In the

Table 2

Interpretation of BF as evidence for H 1

BF_{01}	BF_{10}	Support for H_1
1-.33	1-3	Anecdotal
.33-.10	3-10	Substantial
.10-.03	10-30	Strong
.03-.01	30-100	Very strong
<.01	>100	Decisive

present paper (especially in Experiment 2), a large number of comparisons need to be conducted for the same set of data, which causes the inflation of Type I error. BF eliminates this concern by presenting the relative likelihood of null and alternative hypothesis in each test. As a result, we will focus on BFs in this paper, only reporting the regression coefficients and BFs in the manuscript, while the *p*-values and all other results from null-hypothesis significance testing will be included in the Supplementary Materials.

The BF of LME can be easily obtained in R by extracting the Bayesian information criteria (BIC). BF_{01} (comparing H_0 against H_1) is calculated as:

$$BF_{01} = e^{\frac{(BIC(H_1) - BIC(H_0))}{2}}$$

BF_{10}, which denotes the likelihood of H_1 against H_0 is simply the inverse of BF_{01}. While the continuous scale of BF is useful enough, we could conveniently describe the results, using Jeffreys (1998)'s guideline to characterise BF as Table 2.

Results and Discussion

Seven participants' data were excluded from analysis because their accuracy to the fillers was too low (< 65%). No further participants were excluded because no one's RT was exceptionally long (i.e., > 2.5SD from the group mean). 849 prime-target pairs (23.01%) were excluded because the response to either or both items were incorrect (519 targets (14.07%) and 412 (11.17%) primes). Finally, we removed individual target items whose RT were more than 2.5SD from the condition means, which led to further exclusion of 70 trials (2.46%). Thus, 2771 trials were left and entered into the analysis.

The grand mean of RT was 805ms (SD = 255ms) across all three target modalities. Condition means of each modality were 805ms (256ms) for auditory targets, 781ms (247ms) for haptic targets, and 820ms (259ms) for visual targets.

Perceptual switching costs. As seen in Figure 2, people reacted slower to the perceptual switch condition than the perceptual non-switch condition. However, the mixed-linear model did not support the modality switching costs. The regression coefficients of models at each step is shown in Table 3. A calculation of BFs showed that the data was no better explained by the Step 2 model (with target modality + perceptual switch) than by the Step 1 model (with target modality only; BF_{21} = 1.35), nor was it better explained by the Step 3 model (with interaction) than the Step 2 model (BF_{32} = 2.20). In all, having the perceptual switch with its interaction with target modality explained the data 2.96 times better than the model with only random variables, providing only anecdotal evidence for the perceptual switching costs. Planned comparisons per target modality did not show a significant effect of perceptual switch in any perceptual modalities (Auditory b = 26.86, Haptic b = 3.70, Visual b = 15.00).

Table 3

Regression coefficients of perceptual switching models in hierarchical model comparison

	b	BF_{10}
Step 1		
Target modality		
Haptic - auditory	-22.99	
Visual - auditory	27.56	
Step 2		1.35
Target modality		
Haptic - auditory	-23.02	
Visual - auditory	27.57	
Perceptual switch	15.47	
Step 3		2.20
Target modality		
Haptic - auditory	-7.76	
Visual - auditory	35.37	
Perceptual switch	26.84	
Target modality : perceptual switch		
Haptic - auditory : perceptual switch	-23.14	
Visual - auditory : perceptual switch	-11.84	

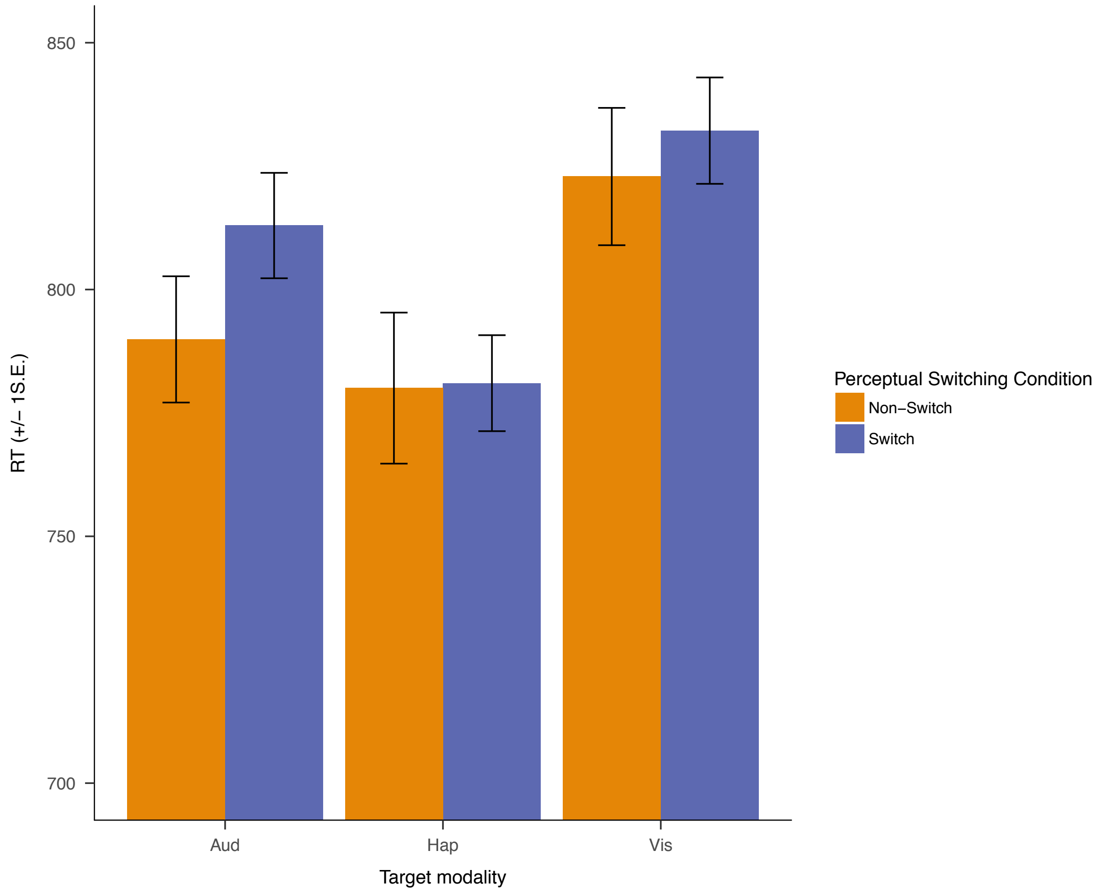


Figure 2. Mean RT of target trials per perceptual switching condition. Error bars denote +/- 1 standard error.

Linguistic switching costs. Similar to the perceptual switching condition, the linguistic switching condition also produced a general trend as predicted, apart from visual items which was faster for a linguistic switch (825ms, SD = 264ms) than a non-switch (831ms, SD = 257ms; Figure 3). The regression coefficients are shown in Table 4. Nevertheless, the BF comparing showed that although the data was no more likely under the Step 2 model than Step 1 model ($BF_{21} = 0.90$), it was 5.88 times more likely under the Step 3 model than Step 2 model, constituting substantial evidence for the interaction between linguistic dimension switch and target modality. Planned comparisons did not show a significant linguistic switching costs in any

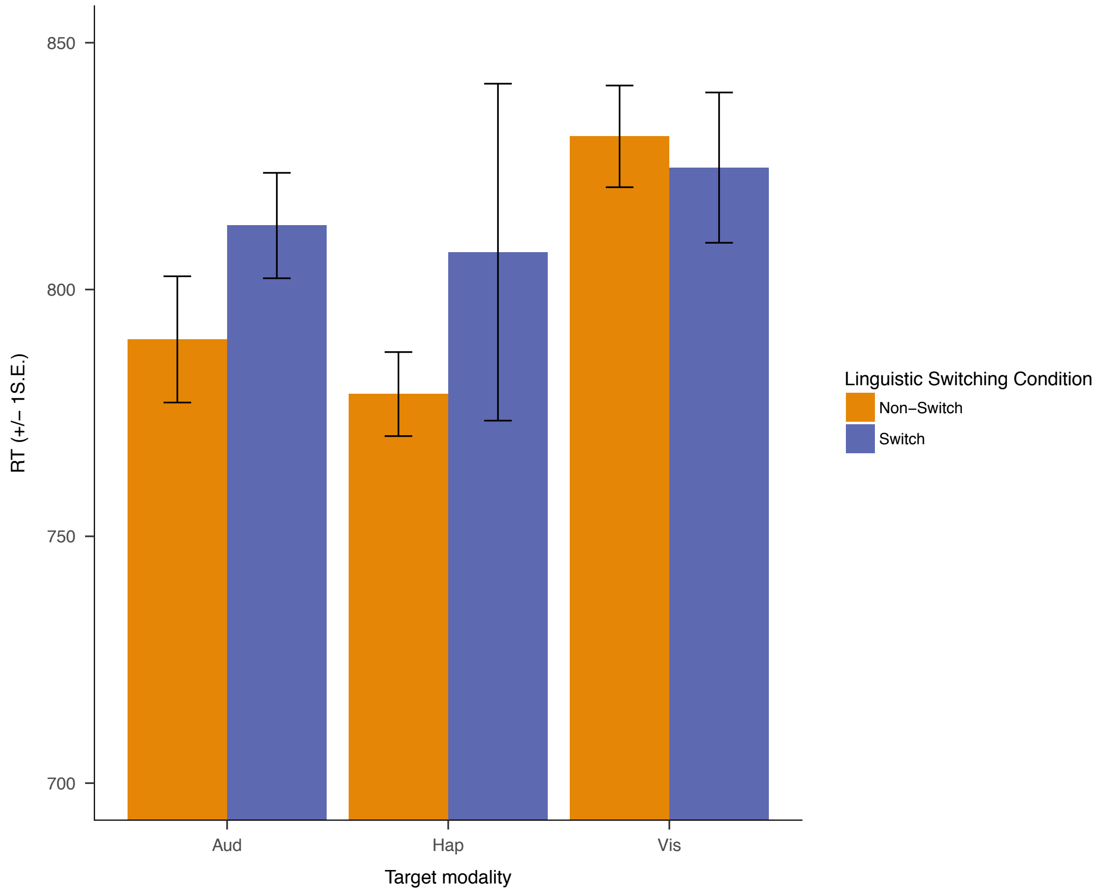


Figure 3. Mean RT of target trials per linguistic switching condition. Error bars denote +/- 1 standard error.

perceptual modality (Auditory b = 23.06, Haptic b = 28.75, Visual b = -6.33).

Further comparisons between the perceptual switching models and linguistic switching models showed that these models were on the par in explaining the modality switching costs. The Step 2 model of perceptual switch (without the interaction term) was equal to that of linguistic switch ($BF_{PL2} = 1.50$), while the Step 3 model of perceptual switch (with the interaction term) was not worse than that of the linguistic switch either ($BF_{PL3} = 0.56$).

Summary. The BFs showed some anecdotal evidence in favour of perceptual modality switching costs which interacted with the target modality, as well as

Table 4

Regression coefficients of linguistic switching models in hierarchical model comparison

	b	BF_{10}
Step 1		
Target modality		
Haptic - auditory	-22.99	
Visual - auditory	27.56	
Step 2		0.90
Target modality		
Haptic - auditory	-16.37	
Visual - auditory	31.23	
Linguistic switch	11.17	
Step 3		5.88
Target modality		
Haptic - auditory	-6.65	
Visual - auditory	47.87	
Linguistic switch	26.87	
Target modality : linguistic switch		
Haptic - auditory : linguistic switch	-5.58	
Visual - auditory : linguistic switch	-34.81	

substantial evidence in favour of the interaction between linguistic dimension switching costs and the target modality. Specifically, target modality moderated the effect of the modality switching costs for both types of switch, such that the perceptual switching costs diminished for haptic items, whereas the linguistic switching costs were reversed for visual items.

Experiment 2: ERP Effects of Modality Switching Costs

In Experiment 1, we found some evidence for both perceptual and linguistic switching costs, though neither were very strong. What was of more importance was that the two types of switch led to switching costs in different ways (i.e., in different modalities). Therefore, it was ever more crucial to study how the switch would affect the neural activities during online processing. In this experiment, we studied the ERP of modality switching costs in a 1000ms epoch across the whole scalp. We proposed some tentative hypotheses which suggested that the perceptual modality switch should affect the ERP around 400ms after the property onset, and the linguistic dimension switch should affect the ERP prior to that time window. Here we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Method

Participants. Twenty-five native speakers of English were recruited from Lancaster University for the payment of £7/hour (or the equivalent of credits). The sample size was determined based on Hald et al. (2011)'s study. Two participants were later excluded because of their high error rate to the filler trials (accuracy < 65%). Among the rest of the participants, mean age was 21.00 (*SD* = 1.38), three were male and two were left handed.

Materials. The prime → target pairs from Experiment 1 remained for this study if they resulted in a modality switching costs (i.e., the RT to a target property in a switch trial was longer than its RT in a non-switch trial), which were 114 different prime → target pairs. As in experiment 1, all the items belonged to three perceptual modalities (i.e., visual, auditory and haptic) and two linguistic dimensions (i.e., PC1 and PC2). All auditory targets were either both linguistic and perceptual switch (different prime → target pairs $N = 36$), or both linguistic and perceptual non-switch ($N = 22$); whereas visual targets could represent three different conditions: both switch ($N = 17$), both non-switch ($N = 19$), and perceptual switch but linguistic non-switch ($N = 20$). Haptic targets with their primes were all removed because the remaining sample size was too small.

All prime-target pairs were presented in a random order to each participant, as well as fillers which appeared in Experiment 1. Both prime → target pairs and fillers were presented either once or twice in the whole experiment to boost the number of trials in each condition, resulting in 137 prime-target pairs and 237 filler pairs for each participant. The experiment also included a practice session of 10 trials which did not feature in the test phase.

Procedure. Participants were tested individually in a quiet room. They were seated at a distance away from the computer screen where they could read the words on the screen and rest their both hands on the desk to press the buttons. The trial began with a fixation cross (1 second) followed by a variable blank screen between 400ms and 1600ms. Afterwards, participants saw the concept-property pairs on the screen consecutively (i.e., 200ms word followed by 300ms blank). Unlike Experiment 1, after a property disappeared, participants were required to refrain from response for 1 second until a prompt ("?") appeared, to make sure that the ERP would not be contaminated by the response readiness potential (Dehaene et al., 1998). If they had responded before the prompt appeared, a warning would appear ("Too fast. Please wait for the prompt '?'."), and the trial was dismissed. Once the prompt appeared, participants again needed to judge whether the property was **usually** true for the concept by pressing the "n" key (masked by a red sticker) with their left hand for "yes", or the "m" key (masked by a white sticker) with their right hand for "no". The same feedback for slow (> 3s) or incorrect response was given as in Experiment 1.

Participants were asked not to blink or move freely, but were given specific time for such muscle movements. Between trials, participants were free to blink for 1 second, and they were told not to blink within each trial when the words were showing. They were also given self-administered break nine times during the experiments after each five-minute block. Furthermore, each block began with 4 seconds of blank screen for the EEG recording to be recalibrate, during which participants were free to blink.

Apparatus. Words were presented on a 19-inch CRT monitor operating at 100Hz refresh rate using EPrime 1.0, and behavioural responses were recorded using a QWERTY keyboard. High-density EEG was recorded continuously using Hydrocel Geodesic Sensor Nets (Electrical Geodesics Inc., Eugene, OR, USA) at 128 locations referenced to the vertex (Cz) online. The ground electrode was at the rear of the head (between Cz and Pz). Electrophysiological signals were acquired at the sampling rate of 1000Hz by an Electrical Geodesics Inc. amplifier with a band-pass filter of 0.1-100Hz. The impedances were checked prior to the beginning of the recording and they were considered acceptable if lower than 50kΩ.

EEG analysis. The digitized EEG was further processed offline using NetStation v 4.5.4. It was band-pass filtered between 0.4-100Hz and was segmented into epochs including 200ms before adjective onset and 1000ms after adjective onset for the target trials. EEG epochs were automatically rejected for body and eye movements, as well as signals exceeding ± 200 µV at any electrode. Data were further checked through visual inspection for eye-movements, eye-blinks and other artifacts not detected by the automated algorithm. The artifact free segments were baseline-corrected with respect to the average amplitude in the 200 ms window preceding the stimulus onset, and were re-referenced to the average reference. Then, they were exported to comma-separated value (csv) files per participant per trial and analysed in R.

In order to explore the continuous activations of embodied and linguistic components, we analysed the EEG output throughout the whole 1-second epoch after the adjective onset, and across the whole scalp (Amsel, 2011; Hauk, Pulvermüller, Ford, Marslen-Wilson, & Davis, 2009; Hendrix, Bolger, & Baayen, 2017). Ninety electrodes were selected to be analysed, covering the region of a 20-20 EEG net (Appendix A). The output from each electrode were sliced into 20 uniformed segments and averaged within each segment to get the mean amplitude per participant per stimulus. The resulting mean amplitude was analysed in linear mixed-effect models (LME) in a hierarchical fashion as Experiment 1. The analyses were exceptionally computationally heavy and thus were carried out distantly using the High End Computing Cluster (HEC; http://www.lancaster.ac.uk/iss/services/hec/) of Lancaster University.

Separate LME models were built to study perceptual modality switch and linguistic dimension switch. The models were built in a hierarchical fashion as Experiment 1.

1. To examine the effect of perceptual modality switch, models were built in the following steps:

   Step 1_{P}: Models with random predictors, participant and target property crossed, modelling random intercept, and target modality (auditory coded as the reference level);

Step 2_P: Models with random predictors and target modality + perceptual switch (non-switch coded as 0, switch coded as 1);

Step 3_P: Models with random predictors and target modality + perceptual switch and their interaction.

2. To examine the effect of linguistic dimension switch, same three steps were carried out:

   Step 1_L: Models with random predictors and target modality (auditory coded as the reference level);

   Step 2_L: Models with random predictors and target modality + linguistic switch (non-switch coded as 0, switch coded as 1);

   Step 3_L: Models with random predictors and target modality + linguistic switch and their interaction.

The models in each step were compared with the model in previous step using BF. Thus, the comparison between, for example, Steps 2_P with Steps 1_P showed how much more likely the data was when the perceptual switch was included than when it was excluded. In addition, we examined whether the data was more likely under the linguistic switch or perceptual switch models by comparing the two Step 2 models above, as well as the two Step 3 models above.

Results

Data exclusion. Two participants were deleted because their accuracy to filler trials were lower than 65%. Individual trials were removed if the response to either the prime or the target or both was incorrect (42 trials, 1.52%). This left us 2713 trials, among which 1248 were auditory targets and 1465 were visual targets. For all auditory targets, 725 were both perceptual and linguistic switch (58.09%), and 523 were both perceptual and linguistic non-switch (41.91%); and for visual targets, 499 were both perceptual and linguistic switch (34.06%), 477 were perceptual switch but linguistic non-switch (32.56%), and 489 were both non-switch (33.38%).

Behavioural analysis. We analysed the RT (descriptive statistics in Table 5) in LME models using hierarchical model comparison in the same way as Experiment 1. We found that the data could be best explained by the model with target modality only. In terms of perceptual switch, the target-modality model was 1.43 times better than the model with target modality and perceptual switch without the interaction term (BF_{21} = 0.70) and 1.17 times better than the model with the interaction term (BF_{31} = 0.86). In terms of linguistic switch, the target-modality model was 2.17 times better than the model without the interaction term (BF_{21} = 0.46) and 1.16 times better than the model with the interaction term (BF_{31} = 0.86). According to the target-only model from Step 1, response to visual targets were faster than auditory targets (b = -0.60)

Table 5

Mean (SD) of RT per condition in Experiment 2

	Perceptual		Linguistic	
	Switch	Non-switch	Switch	Non-switch
Auditory	347 (195)	348 (204)	347 (195)	348 (204)
Visual	354 (216)	334 (198)	355 (220)	343 (206)

ERP analysis.

Overview.

The mean BF across the scalp for the model with linguistic switch in Step 2 (i.e., without interaction) was 0.004 (SD = 0.003), and the mean BF for the model with perceptual switch in Step 2 was 0.005 (SD = 0.004), both suggesting very strong evidence for the null hypothesis (i.e., model with target modality only). Figure 4 shows the mean BF_{21} across the scalp over time. The BFs were less than 0.33 for the whole epoch, and below 0.1 from 100ms after the target adjective onset. By comparing between perceptual and linguistic switch, it seemed that the perceptual switch had an advantage over the linguistic switch at the first 50ms of the epoch and around 300ms, whereas the linguistic switch had an advantage around 650ms. The same pattern was found for the models of interaction (i.e., data strongly favoured models with only linguistic or perceptual switch (Step 2) against models with interaction terms (Step 3; see Supplementary Materials)). Viewing across the scalp, we found overall evidence against both perceptual and linguistic switch. However, it was still possible that these switches affected ERP in a localised manner, which will be examined next.

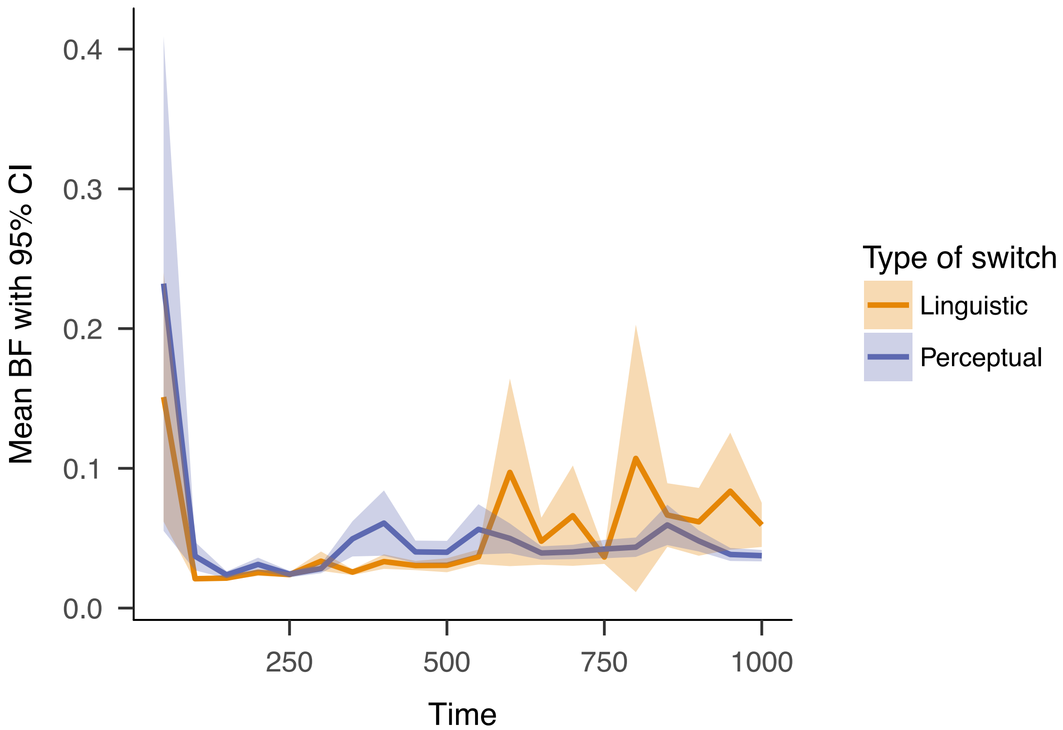


Figure 4. Line plot of mean BF_{21} with 95% credibility interval of perceptual and linguistic switch as a result of Step 2 across scalp at each 50ms slice.

Perceptual switch.

Figure **??**a showed the topoplot of the evidence in favour of (or against) models with target modality and perceptual switch (Step 2_{P}) compared with target-modality models (Step 1_{P}); Figure **??**b showed the comparison between models with the interaction term between target modality and perceptual switch (Step 3_{P}) and models without the interaction term (Step 2_{P}). There was no widespread effect of perceptual modality switching costs or its interaction with target modality, apart from the beginning of the epoch (i.e., 0ms to 50ms) in the occipital electrodes (e.g., E70, BF_{21} = 7.14), that is the data was 7.14 times more likely under the model with perceptual modality switch than without. During this time, there was evidence that perceptual switching costs made the amplitude more negative (b = -0.98). This effect was not moderated by the target modality, meaning that this early effect of perceptual switching costs had an effect on both visual and auditory targets equally.

Furthermore, the data suggested that in the following areas the interaction models were more likely than models excluding the interaction term (comparison between Step 3_P and Step 2_P models, not shown in topography). Perceptual switch interacted with the target modality between 300ms and 350ms in the frontal region (e.g., E19, BF_{32} = 19.70, perceptual switch b = 1.81, target modality b = 1.31, interaction b = -2.57). That is to say, in the frontal region, the perceptual switch was associated with more positivity than non-switch for auditory targets; but perceptual switch was associated with more negativity for visual targets. The opposite pattern was found in the right parietal region (e.g., E92, BF_{32} = 11.86, perceptual switch b = -1.26, target modality b = -1.13, interaction b = 1.88), which suggested that perceptual switch led to more negativity for auditory targets but more positivity for visual targets. Both of these effects coincided with previous findings of the early N400 effect (parietal effect with Bernabeu et al., 2017; frontal effect with Collins et al., 2011; Hald et al., 2011), which could be interpreted as a precursor of semantic processing.

Linguistic switch.

Evidence for linguistic dimension switching costs was found in similar regions as the perceptual modality switching costs (Figure 5). The data substantially favoured the linguistic modality switching costs (Step 2_L compared with Step 1_L) at the onset of the epoch in the left occipital region as well (e.g., E70, *BF21* = 3.17), such that the linguistic switch trial was associated with more negative amplitude (b = -0.92).

The data also strongly favoured the interaction between linguistic dimension switching costs and target modality between 300ms and 350ms at the left frontal (Step 3_L compared with Step 2_L; e.g., E19, BF_{32} = 238.31, linguistic switch b = 1.81, target modality b = 1.23, interaction b = -3.07) and right parietal regions (e.g., E92, BF_{32} = 167.36, linguistic switch b = -1.26, target modality b = -1.06, interaction $b = 2.27$). In other words, similar to the perceptual switching costs, the linguistic switch was associated with more positivity for auditory targets but more negativity with visual targets in the frontal region; and the reversed pattern for the parietal region. For both regions, the data was more than 150 times more likely under models with the interaction than the models without interaction.

However, it is important to note that Step 3 models with the interaction term was not better than the Step 1 models which included only target modality. Take the linguistic models for example, when models in Step 2_{L} was compared with Step 1_{L}, the data of E19 was 0.004 times more likely (i.e., 250 times less likely) under the Step 2_{L} models which considered the linguistic switching costs alone without the interaction. In other words, the Step 3_{L} model was 0.95 times more likely than the Step$_1$ model (BF_{32} x $BF_{21} = 238.31$ x $0.004 = 0.94$). In the same way, the data of E92 was 0.002 times more likely (i.e., 500 times less likely) under the Step 2_{L} model than the Step 1_{L} model, so the Step 3_{L} model was 3 times worse than the Step 1_{L} model ($BF_{31} = 0.33$).

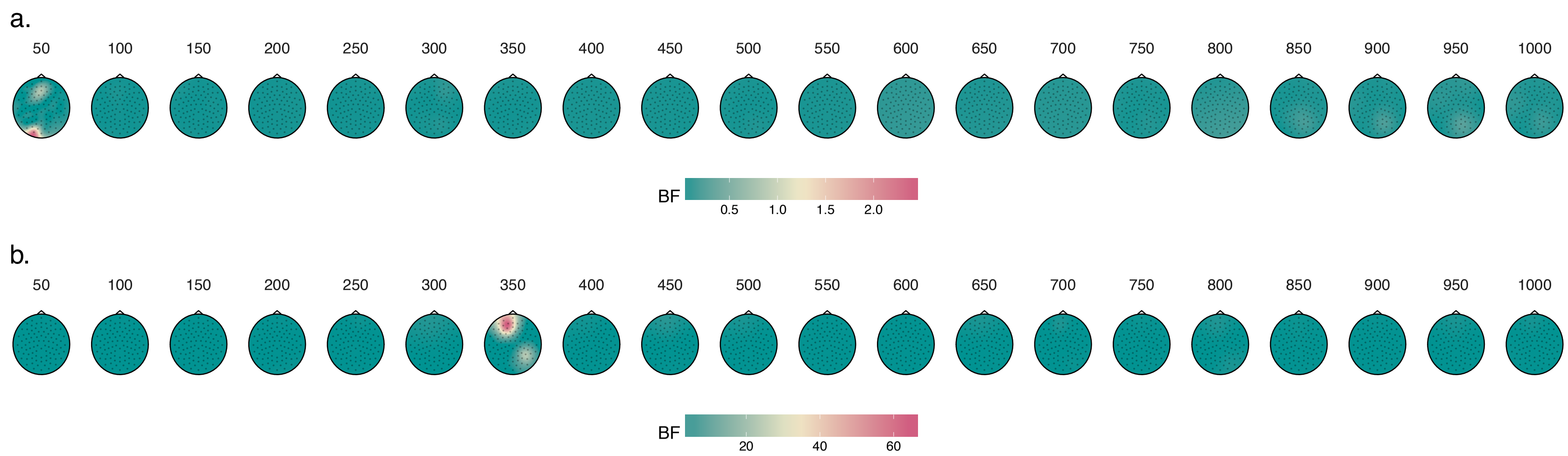


Figure 5. Effects of linguistic switch as a result of model comparisons in a. Step 2: models including linguistic switch vs Step 1: models excluding linguistic switch and b. Step 3: models including the interaction between linguistic switch and target modality vs Step 2: models excluding the interaction. BF > 3 constitutes substantial evidence in favour of linguistic switching costs; BF > 10 constitutes strong evidence in favour or linguistic switching costs; BF > 100 constitutes decisive evidence in favour of linguistic switching costs.

Comparison between perceptual and linguistic switch.

From the hierarchical model comparison, we discovered that perceptual switch and linguistic switch shared similar patterns of activities, supporting their effects at the occipital region at the onset of the epoch, and their interaction with target modality at both the left frontal region and the right parietal region between 300ms and 350ms. Therefore, it is worthwhile to examine which type of switch was supported more strongly by the data. Figure 6 showed the evidence in favour of one type of switch against the other (i.e., comparison of Step 2 models and comparison of Step 3 models). We found that perceptual switch and linguistic switch performed equally well between 0ms and 50ms in the left occipital region (e.g., E70: $BF_{PL2} = 2.91$). The data supported neither model more. However, the data supported the interaction between linguistic switch and target modality more than the interaction between perceptual switch and target modality at the left frontal region (Figure 6b; e.g., E19: $BF_{PL3} = 0.17$) and the right parietal region (E92: $BF_{PL3} = 0.11$) between 300ms and 350ms. That is to say, by comparing the interaction models alone, linguistic switching costs were more than 5.88 times more likely than perceptual switching costs. It suggested that the “early N400 effect” found in previous studies was better explained by a switch between the linguistic dimensions than the perceptual modalities. Nevertheless, the fact that the linguistic switching costs were moderated by the target modalities suggested that modality-specific resources were also recruited to construct the conceptual representations.

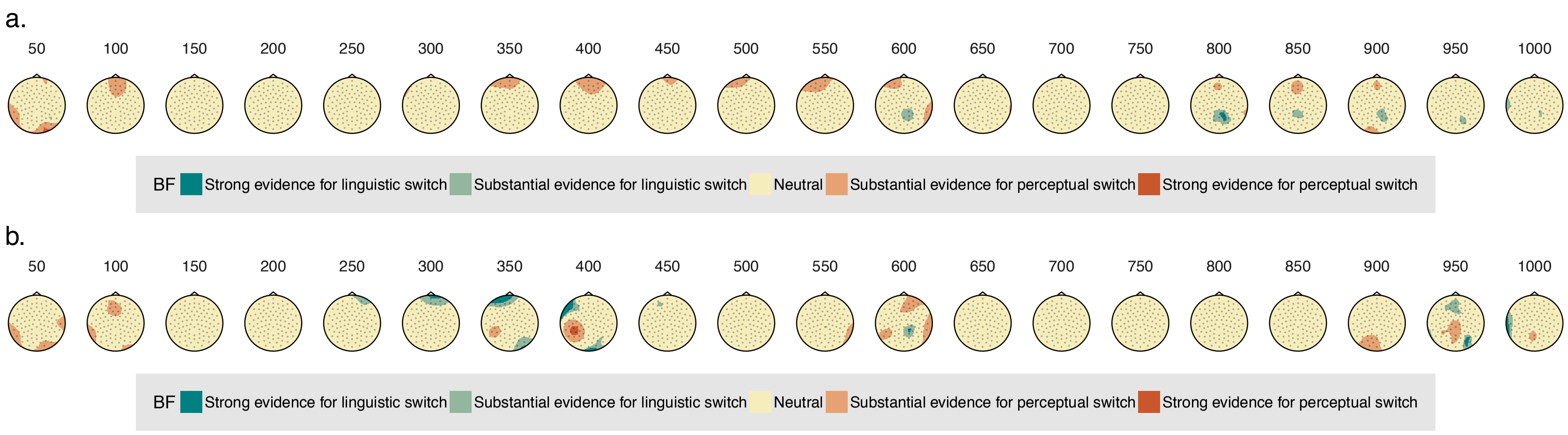


Figure 6. Evidence for perceptual switch in comparison with linguistic switch as a result of comparing models in a. Step 2 (excluding the interaction) and b. Step 3 (including the interaction). $BF_{PL} > 3$ constitutes substantial evidence in favour of perceptual switching costs; $BF_{PL} > 10$ constitutes strong evidence in favour or perceptual switching costs; $BF_{PL} < 0.33$ constitutes substantial evidence in favour of linguistic switching costs; $BF_{PL} < 0.1$ constitutes strong evidence in favour or linguistic switching costs.

Further exploration showed great distinction between the Step 2 models and Step 3 models. For example, the ERP in the frontal region around 350ms better supported the perceptual model in Step 2, but supported the interaction between linguistic switch and target modality in Step 3. This pattern suggested that it is the target modality that drove the difference between perceptual and linguistic switching costs. Therefore, we further analysed the data per target modality in the next section.

Moderation of target modality.

We performed hierarchical model comparisons again per target modality. That is,

1. For the perceptual switching costs, models were built in two steps.

   Step 1_{P}: Models with only random variables;

   Step 2_{P}: Models with random variables and perceptual switch.

2. linguistic switch, models were built in two steps.

   Step 1_{L}: Models with only random variables;

   Step 2_{L}: Models with random variables and linguistic switch.

Since for auditory targets perceptual and linguistic switching conditions were exactly the same, two sets of hierarchical steps above produced the same results.

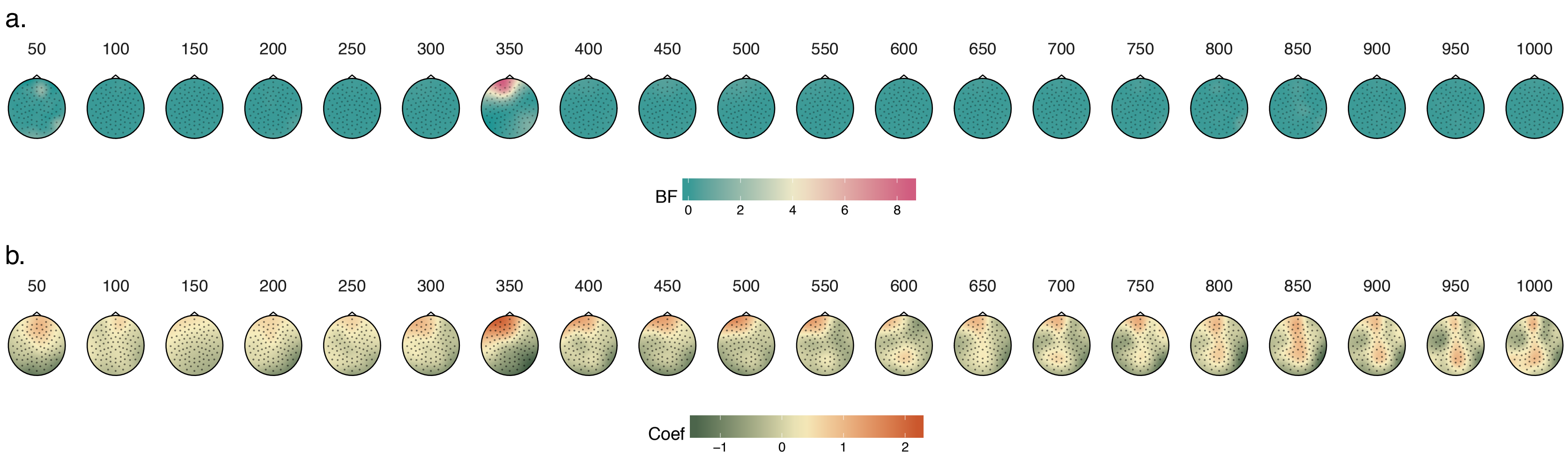


Figure 7. Switching costs for auditory targets illustrated by a. BF of models with perceptual-linguistic switch (Step 2) against models with only random variables (Step 1), and b. raw regression coefficients of perceptual-linguistic switch in Step 2 models.

Auditory targets.

Evidence supported a switching effect at the onset of the epoch (within 50ms) and between 300ms and 350ms (Figure 7a), although the pattern of regression coefficients was consistent throughout the whole epoch (Figure 7b). Switch trials evoked larger positivity in the left frontal region and larger negativity in the right parietal and occipital regions. Since linguistic switch completely overlapped with perceptual switch, it could be argued that when the linguistic and perceptual switch happened simultaneously, it could be detected between 300ms and 350ms, coinciding with the early N400 effect from previous studies.

Visual targets.

For visual targets, it was possible to distinguish between perceptual switch and linguistic switch because half of the perceptual switch trials were linguistic non-switch, while the other half were linguistic switch. From Figure 8a, we found evidence against both linguistic and perceptual switching costs between 300ms-350ms, suggesting that the interaction effect in this time window was driven by the effect for auditory targets (and the lack thereof for visual targets). Furthermore, there was strong evidence against any perceptual switching costs for visual targets over the effect of random variables. Linguistic switch (Figure 8b) had an effect at the 800ms slice in the right posterior regions (e.g., E84; b = 2.08, BF_{21Lv} = 4.32), which meant that the linguistic switch evoked a greater positivity at this later stage of processing, corresponding to LPC which was often associated with decision making and conflict resolution.

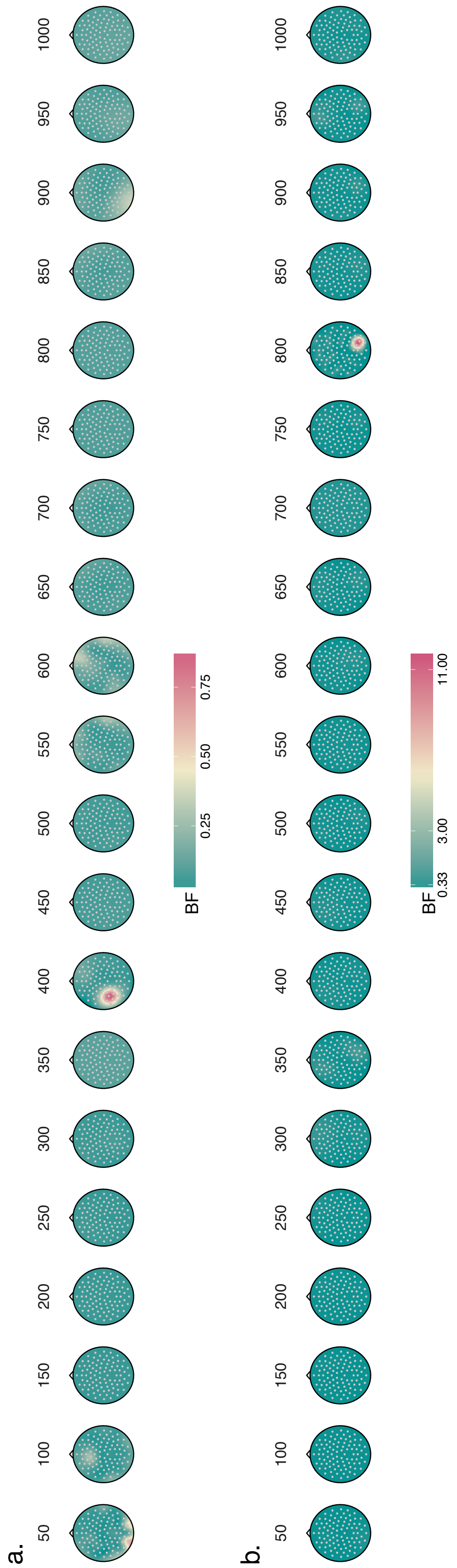


Figure 8. Effects of a. perceptual and b. linguistic switching costs for visual targets.

Again to show the relative power of perceptual and linguistic switch, Figure 9 is a topoplot of BFs by comparing the Step 2 models of visual targets, which was similar to Figure 6b. Strong evidence supported the linguistic switching costs in the frontal and occipital areas from 200ms to 400ms. The linguistic switch evoked a stronger negativity in the frontal region (e.g., at 250ms-300ms, E 18: b = -1.17, BF_{PLv} = 0.08), while it evoked a stronger positivity in the parietal region (e.g., at 300ms-350ms, E77: b = 1.20, BF_{PLv} = 0.10). This effect was consistent with the early N400 effect found for auditory targets, and the linguistic switch could be detected as early as 200ms. Again around 600ms, linguistic switch evoked a stronger positivity in the parietal region (e.g., at 600ms-650ms, E77: b = 1.51, BF_{PLv} = 0.04), corresponding with P600 or LPC.

In contrast, perceptual switch started to have a stronger effect between 300ms and 400ms in the left parietal region (e.g., E52: b = -1.10, BF_{PLv} = 14.77). The results showed that perceptual switch evoked a stronger negativity than non-switch in this left-parietal region, corresponding with the classic N400 effect. Later around 600ms, perceptual switch had greater power than linguistic switch in the right temporal region (e.g., E108: b = -1.24, BF_{PLv} = 10.95), meaning that perceptual switch elicited a stronger negativity than non-switch, which was the reverse of LPC.

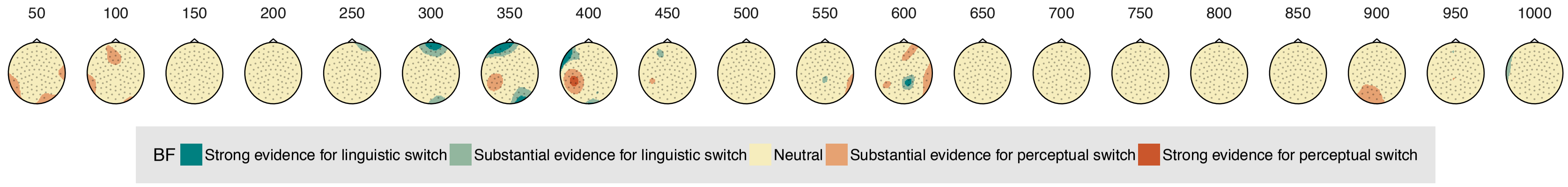


Figure 9. Evidence for perceptual switch in comparison with linguistic switch as a result of comparing models for visual targets in a. Step 2 (excluding the interaction) and b. Step 3 (including the interaction). $BF_{PL} > 3$ constitutes substantial evidence in favour of perceptual switching costs; $BF_{PL} > 10$ constitutes strong evidence in favour or perceptual switching costs; $BF_{PL} < 0.33$ constitutes substantial evidence in favour of linguistic switching costs; $BF_{PL} < 0.1$ constitutes strong evidence in favour or linguistic switching costs.

Summary of ERP results.

In summary, evidence supported both types of switch between 0ms and 50ms in the left occipital region, but the effects diverged later in the epoch. We found evidence for perceptual switching costs around 400ms in the left parietal region, and later around 600ms in the right temporal region, albeit this effect only existed for visual targets and was relative to the power of linguistic switch. We also found evidence for linguistic switching costs around 300ms in the frontal region and parietal region, prior to the effect of the perceptual switch as predicted. Furthermore, we found evidence for both linguistic and perceptual switch at the later stage of processing (after 600ms).

General Discussion

In this paper, our aim was to differentiate the contribution of linguistic dimension switch from perceptual modality switch in the modality switching costs, and thus to identify the neuroelectrical activations of simulation and linguistic information during conceptual representation. We found that what was considered to be the modality switching costs in ERP components, could be better explained by the switch between different linguistic distributional clusters, rather than the switch between perceptual modalities. Such a finding supported the idea that the linguistic component plays a crucial role, even bearing the burden of semantic representation (Connell, 2018). Because the activation of the linguistic component is more speedy than the embodied component as shown in the ERP, it was possible for the linguistic component to be a shortcut for the more computationally heavy and slower embodied simulation (Barsalou et al., 2008; Connell & Lynott, 2014; Louwerse & Jeuniaux, 2010)

An important finding in the present study was that words pertaining to different sensory modalities elicited different behavioural and neural responses. The behavioural study did not replicate previous findings, but showed that response time in the property verification task was not best explained by perceptual modality switching costs (i.e., *loud* → *bright* takes longer than *red* → *bright* because of switch from audition to vision); nor was it best explained by linguistic switch alone (i.e., switch from one linguistic dimension to another). Rather, it was best explained by the interaction between a switch between linguistic dimensions and the perceptual modality of the target word (which was slightly better than the model including the interaction between perceptual switch and the perceptual modality). In other words, the linguistic dimension switching costs occurred for words of some modalities (auditory and haptic words specifically) but not other words (visual words specifically).

The ERP results also showed the same pattern, that is: auditory words turned out to be the only category that elicited consistent perceptual/linguistic modality switching costs, while visual words did not (see also Moscoso Del Prado Martín, Hauk, & Pulvermüller, 2006; Schmidt-Snoek, Drew, Barile, & Agauas, 2015). As discussed before, it is impossible to differentiate perceptual from linguistic switch with auditory targets because the switching conditions completely overlapped. However, it is possible to propose these speculations. First, it is possible that the ERP switching costs were driven by the linguistic switch more, because it is the case for visual targets when the strength of linguistic switch was compared with perceptual switch. In other words, linguistic switch could explain the neuroelectrical activations around 350ms in the frontal region better than perceptual switch for visual target; so it is plausible to assume that the same occurred for auditory targets too. Second, the reason why the ERP results showed strong evidence at this point could be because of the overlap between perceptual and linguistic switches. Since the simulation and linguistic information agreed with each other, they elicited strong activations at this relatively early time window (i.e., before the typical N400 window).

From the timing of ERP effects, it could also be seen that the linguistic component did not play only one role, that is it did not only elicit neural activities once at the beginning of the processing as suggested by LASS (Barsalou et al., 2008). It also elicited activities later around 800ms, which was limited to visual targets (i.e., when linguistic and simulation information did not completely overlap). Therefore, it could be argued that when there was a mismatch between linguistic and embodied components, the linguistic information was utilised later during processing, possibly to reconcile the conflicting information before decision making (Polich, 2007) or integrating distinct conceptual representations (Davenport & Coulson, 2011; Kuperberg, 2007; Paczynski & Kuperberg, 2012).

In general, our findings were different from previous findings of modality switching costs (e.g., Pecher et al., 2003), which primarily stressed the contribution of modality-specific resources during conceptual representations. We found that, although modality-specifc simulation was useful and important, it was moderated and guided by the linguistic information. In fact, the evidence perceptual switch was weak, which did not have an absolute effect on either the behavioural outcome or the neuroelectrical activations, and only outperformed linguistic switch in a relative sense around 400ms after the adjective onset. This suggests that sematic representation which relies on embodied simulation (i.e., as signaled by the N400 effect; Kutas & Federmeier, 2011) had only limited effects on the processing of sensory properties. Instead, it is the linguistic distributional patterns that did the most heavy-lifting.

Limitations and future directions

Our study could not implement a full-fledged independent manipulation of the linguistic component because auditory words tend not to appear in the same linguistic cluster as other sensory properties (Louwerse & Connell, 2011). If future studies could ensure independent factorial manipulation (i.e., both perceptual and linguistic switch, perceptual switch and linguistic non-switch, perceptual non-switch and linguistic switch, both perceptual and linguistic non-switch), we would predict that the early N400 effect would appear for the two conditions where there is a linguistic switch whereas the LPC to appear for the two conditions where there was a mismatch between perceptual and linguistic switch.

Our study also found an immediate effect of both embodied and linguistic switch at the first 50ms of the epoch. Although embodied simulation and linguistic distributional patterns were found to have speedy activations, it was unlikely that they could affect semantic processing at the immediate property onset. Instead, this effect could be a result of spill-over from the concept word. Because a trial was presented with a concept (e.g., *SUN*) first for 500ms followed by a property (e.g., *bright*), it was plausible that the conceptual representation of the concept word was still ongoing by the time the epoch started. Because the properties were usually salient for the concepts, the switching costs might have already been incurred when participants were processing the concept word. Bernabeu et al. (2017), in their recent replication of previous ERP studies, reversed the order of concept and property and did not find an immediate effect from the property onset. In future studies, it is recommended to adopt the reverse order, control the concept words so that they do not automatically activate the properties before the words are shown, or analyse epochs after both the concept and property words.

Conclusion

The present study provided support for the crucial role of the linguistic component in conceptual representations. Instead of finding it to perform a peripheral role, we found that it bore the burden of conceptual representations, preceding and even overpowering the effect of the embodied component. We suggest that representing the meaning of a word is indeed embodied, but the linguistic information support conceptual representation in an important way (Connell, 2018). The timecourse of the activations of the two components satisfies the requirement for the linguistic component to be a shortcut before the slower and more detailed simulation produces a detailed representation (Barsalou et al., 2008; Connell & Lynott, 2014; Louwerse & Jeuniaux, 2010).

References

Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia*, *49*(5), 970–983. doi:10.1016/j.neuropsychologia.2011.01.003

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. doi:10.1037/a0016261

Auguie, B. (2017). *GridExtra: Miscellaneous functions for "grid" graphics*. Retrieved from https://CRAN.R-project.org/package=gridExtra

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Bache, S. M., & Wickham, H. (2014). *Magrittr: A forward-pipe operator for r*. Retrieved from https://CRAN.R-project.org/package=magrittr

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*(4), 577–609; discussion 610–60. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1693222{\&}tool=pmcentrez{

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In A. M. Glenberg & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 245–283). Oxford, England: Oxford University Press. Retrieved from http://psychology.emory.edu/cognition/barsalou/papers/Barsalou{_}et{_}al{_}chap{

Bates, D., & Maechler, M. (2017). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

doi:10.18637/jss.v067.i01

Bernabeu, P., Willems, R. M., & Louwerse, M. M. (2017). Modality switch effects emerge early and increase throughout conceptual processing: Evidence from ERPs. In *39th annual conference of cognitive science society*. PsyArXiv. Retrieved from https://osf.io/preprints/psyarxiv/a5pcz/

Brants, T., & Franz, A. (2006). Web 1T 5-gram Version 1 LDC2006T13. DVD. Philadelphia: Linguistic Data Consortium.

Chemero, A. (2009). *Radical embodied cognitive science* (p. 252). MIT Press.

Collins, J., Pecher, D., Zeelenberg, R., & Coulson, S. (2011). Modality Switching in a Property Verification Task: An ERP Study of What Happens When Candles Flicker after High Heels Click. *Frontiers in Psychology*, *2*(February), 10. doi:10.3389/fpsyg.2011.00010

Connell, L. (2018). What have labels ever done for us? The lingusitic shortcut in conceptual processing. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2011). Modality Switching Costs Emerge in Concept Creation as Well as Retrieval. *Cognitive Science*, *35*(4), 763–778. doi:10.1111/j.1551-6709.2010.01168.x

Connell, L., & Lynott, D. (2013). Flexible and fast: linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, *20*(3), 542–50. doi:10.3758/s13423-012-0368-x

Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*(3), 390–406. doi:10.1111/tops.12097

Dantzig, S. van, Cowell, R. A., Zeelenberg, R., & Pecher, D. (2011). A sharp image or a sharp knife: norms for the modality-exclusivity of 774 concept-property items. *Behavior Research Methods*, *43*(1), 145–54. doi:10.3758/s13428-010-0038-8

Davenport, T., & Coulson, S. (2011). Predictability and novelty in literal language comprehension: an ERP study. *Brain Research*, *1418*, 70–82. doi:10.1016/j.brainres.2011.07.039

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., … Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597–600. doi:10.1038/26967

Dowle, M., & Srinivasan, A. (2017). *Data.table: Extension of 'data.frame'.* Retrieved from https://CRAN.R-project.org/package=data.table

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. doi:10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. doi:10.3758/BF03193146

Firth, J. R. (1957). A synopsis of linguistic theory 1930-55. In *Studies in linguistic analysis (special volume of the philological society)* (Vols. 1952-59, pp. 1–32). Oxford, England: Blackwell Publishers. Retrieved from https://books.google.co.uk/books/about/A{_}Synopsis{_}of{_}Linguistic{_}Theory

Gelman, A., & Su, Y.-S. (2018). *Arm: Data analysis using regression and multilevel/hierarchical models.* Retrieved from https://CRAN.R-project.org/package=arm

Genz, A., & Bretz, F. (2009). *Computation of multivariate normal and t* *probabilities*. Heidelberg: Springer-Verlag.

Hald, L. A., Hocking, I., Vernon, D., Marshall, J.-A., & Garnham, A. (2013). Exploring modality switching effects in negated sentences: further evidence for grounded representations. *Frontiers in Psychology*, *4*(February), 93. doi:10.3389/fpsyg.2013.00093

Hald, L. A., Marshall, J.-A., Janssen, D. P., & Garnham, A. (2011). Switching Modalities in A Sentence Verification Task: ERP Evidence for Embodied Language Processing. *Frontiers in Psychology*, *2*(March), 45. doi:10.3389/fpsyg.2011.00045

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1-3), 335–346. doi:10.1016/0167-2789(90)90087-6

Harrell Jr, F. E. (2018). *Rms: Regression modeling strategies*. Retrieved from https://CRAN.R-project.org/package=rms

Harrell Jr, F. E., Charles Dupont, & others. (2018). *Hmisc: Harrell miscellaneous*. Retrieved from https://CRAN.R-project.org/package=Hmisc

Hauk, O. (2016). Only time will tell – why temporal information is essential for our neuroscientific understanding of semantics. *Psychonomic Bulletin & Review*, *23*(4), 1072–1079. doi:10.3758/s13423-015-0873-9

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, *41*(2), 301–307. doi:10.1016/S0896-6273(03)00838-9

Hauk, O., Pulvermüller, F., Ford, M., Marslen-Wilson, W. D., & Davis, M. H. (2009). Can I have a quick word? Early electrophysiological manifestations of psycholinguistic processes revealed by event-related regression analysis of the EEG. *Biological Psychology*, *80*(1), 64–74.

doi:10.1016/j.biopsycho.2008.04.015

Hendrix, P., Bolger, P., & Baayen, R. H. (2017). Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 128–149. doi:10.1037/a0040332

Henry, L., & Wickham, H. (2018). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr

Hothorn, T. (2017). *TH.data: TH's data archive*. Retrieved from https://CRAN.R-project.org/package=TH.data

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Højsgaard, S., & Halekoh, U. (2016). *DoBy: Groupwise statistics, lsmeans, linear contrasts, utilities*. Retrieved from https://CRAN.R-project.org/package=doBy

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes Factors. *The Journal of Problem Solving*, *7*, 2–9. doi:10.7771/1932-6246.1167

Jeffreys, H. (1998). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press. Retrieved from https://global.oup.com/academic/product/the-theory-of-probability-9780198503682?cc=gb{\&}lang=en{\&}

Kintsch, W., & Dijk, T. V. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. Retrieved from http://psycnet.apa.org/journals/rev/85/5/363/

Koenker, R., & Ng, P. (2017). *SparseM: Sparse linear algebra*. Retrieved from https://CRAN.R-project.org/package=SparseM

Kuhn, M., Steve Weston, Wing, J., Forester, J., & Thaler, T. (2016). *Contrast: A collection of contrast methods*. Retrieved from https://CRAN.R-project.org/package=contrast

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*(1), 23–49. doi:10.1016/j.brainres.2006.12.063

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–47. doi:10.1146/annurev.psych.093008.131123

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–40. doi:10.1037/0033-295X.104.2.211

Lebois, L. A. M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are Automatic Conceptual Cores the Gold Standard of Semantic Processing? The Context-Dependence of Spatial Meaning in Grounded Congruency Effects. *Cognitive Science*, *39*(8), 1764–1801. doi:10.1111/cogs.12174

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*(2), 273–302. doi:10.1111/j.1756-8765.2010.01106.x

Louwerse, M. M., & Connell, L. (2011). A taste of words: linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, *35* (2), 381–98. doi:10.1111/j.1551-6709.2010.01157.x

Louwerse, M. M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. In A. Glenberg & A. C. Graesser (Eds.), *Embodiment and meaning: A debate* (pp. 309–326). Oxford, England: Oxford University Press. Retrieved from http://www.madresearchlab.org/Selected{__}Publications{__}files/LouwerseJeuniauxBoo

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114* (1), 96–104. doi:10.1016/j.cognition.2009.09.002

Luck, S. J. (2005). Ten Simple Rules for Deisgning ERP Experiments. In *Event-related potentials: A methods handbook* (pp. 17–32). doi:10.1371/journal.pcbi.0020012

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41* (2), 558–64. doi:10.3758/BRM.41.2.558

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of bayes factors for common designs*. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Moscoso Del Prado Martín, F., Hauk, O., & Pulvermüller, F. (2006). Category specificity in the processing of color-related and form-related words: An ERP study. *NeuroImage*, *29* (1), 29–37. doi:10.1016/j.neuroimage.2005.07.055

Nowosad, J. (2017). *Rcartocolor: 'CARTOColors' palettes*. Retrieved from https://CRAN.R-project.org/package=rcartocolor

Otten, L. J., & Rugg, M. D. (2005). Interpreting event-related brain potentials. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 3–16).

CHAPTER 7

ERP Markers of Embodied and Linguistic Components during Metaphor Processing

In Chapters 4 and 5, we studied conceptual representations during metaphor processing, specifically whether the conceptual representation of metaphor comprehension included both embodied and linguistic component, and whether the linguistic component could become a shortcut to conserve resources. Our results seemed to confirm the former, but remained open on the latter. On one hand, we found that both ease of simulation (i.e., how easy it is to simulate a concept) and linguistic distributional frequency (i.e., how often two component words co-occur in a linguistic context) affected performances during metaphor processing uniquely and independently. However, on the other hand, we did not find evidence that showed an increase in the reliance on LDF even though we had created conditions that encouraged the use of the linguistic shortcut, such as shallow processing and stringent time limits. In fact, the effect of the linguistic component seemed to be wiped out if people were under great time pressure.

In Chapter 6, we explored the neuroelectric markers of embodied and linguistic components in literal language processing. Although the analyses were performed in an exploratory fashion and the behavioural effect of modality switching costs was not replicated, the results confirmed the overarching hypothesis in this thesis, that is: embodied and linguistic components both exist and account for the neurophysiological activities deferentially. We found the activities in the perceptual modality switch to be strongly associated with the ERP activities in the left parietal region around 400ms; while the linguistic dimension switch was associated with the activities in the occipital and frontal regions around 350ms, which was before the activations of the embodied component. The early effects of linguistic dimension switch provided support for the linguistic shortcut hypothesis.

In this chapter, I will return to the theme of the thesis (i.e., metaphor processing), with the findings of the previous three chapters in mind. The aims are two-fold, as reflected in the two experiments in the paper presented in this chapter. First, I will examine again how depth of processing affects conceptual representation during metaphor processing. The prediction remains that the linguistic shortcut will play a larger role in shallow processing, and the embodied component will in turn play a smaller role in shallow processing. Although this question has been touched upon in Chapters 4 and 5, it was not the central question in those chapters, and the design of the experiments did not primarily test this question. In this chapter, the experiments will adopt a within-participant design, so that the hypothesis of concern will be that tasks can moderate the effects of EoS or LDF.

The second aim of this chapter was to study the timecourse of embodied and linguistic components during metaphor processing. Since ERP more sensitive than the behavioural measures in previous chapters, it is possible to detect the differences in the activation of the linguistic component using ERP even though we did not find any evidence for the linguistic shortcut on response decisions and RT in Chapters 4 and 5 (which may well be the case in this chapter). The ERP findings concerning literal language processing in Chapter 6 will be used as hypotheses for the study in this chapter; but the analyses in this chapter will still be component free and cover the whole scalp and epoch.

Paper 5: Continuous neural activities of grounded representations in metaphor processing

Page numbers: 275 - 334

Continuous neural activations of simulation and linguistic information during metaphor processing

Pei Q. Liu[1] & Louise Connell[1] & Dermot Lynott[1]

[1] Department of Psychology, Lancaster University

Author Note

Department of Psychology, Lancaster University, UK.

Correspondence concerning this article should be addressed to Pei Q. Liu, Department of Psychology, Fylde College, Lancaster University, Bailrigg, LA1 4HX, UK. E-mail: p.liu1@lancaster.ac.uk

Abstract

Language processing relies on conceptual representations which are composed of two crucial components, embodied simulation and linguistic distributional pattern. The embodied component refers to the reactivation of previous sensorimotor experiences related to the concept (e.g., experiences with a clever student when reading "bright student"); the linguistic component refers to the co-occurrence pattern of the constituent words (i.e., how often "bright" and "student" appear in the same context). In this study, we examined the existence and roles of these components in metaphor processing. Using both a behavioural study and EEG, we studied how these components affected the speed, success rate and neurophysiological activations of metaphor comprehension. We found that, while performance of metaphor comprehension was mainly influenced by the embodied component, the linguistic component was activated before the embodied component reached its peak and could act as a shortcut to construct good-enough representation, such that people found it easier to accept and hard to reject a metaphor when the distributional frequency of constituent words was high. In other words, the linguistic distributional pattern could provide a guide for conceptual representations before the embodied component was fully engaged.

Keywords: Metaphor processing, embodied simulation, linguistic distributional pattern, EEG

Word count: 7212 words

Continuous neural activations of simulation and linguistic information during metaphor processing

In the past two decades, researchers have gradually come to the consensus that conceptual representation has at least two components, an embodied component and a linguistic component (Barsalou, Santos, Simmons, & Wilson, 2008; Connell & Lynott, 2014; Louwerse & Jeuniaux, 2010; Vigliocco, Meteyard, Andrews, & Kousta, 2009). As new concepts are acquired, neural networks are activated in response to the physical environment as well as language. The activation of different neural regions leaves traces of conceptual representation; and during conceptual processing (e.g., language comprehension), these traces are reactivated to simulate the sensory, motor, affective, introceptive and other bodily experiences, thus forming the embodied component, as well as to activate the distributional pattern of linguistic symbols, thus forming the linguistic component.

The embodied and linguistic components are highly intercorrelated, because the distributional pattern of words in language often reflects the associations of objects, events, and ideas in the real world (Louwerse, 2011). The fact that *bright* appears with *sun* frequently in close proximity reflects the physical reality that the sun is usually bright. Thus, the linguistic information could approximate the embodied component, thereby informing a response on its own; and the linguistic information could in turn activate related simulation information (Johns & Jones, 2012).

Andrews, Vigliocco, and Vinson (2009) modeled conceptual representation with both simulation and linguistic information and found that word meaning is best represented by the combination of these two types of data. In other words, both embodied and linguistic components are central to the conceptual system. The Language and Situated Simulation theory (LASS: Barsalou et al., 2008) as well as the Symbol Interdependency Hypothesis (SIH: Louwerse, 2011) propose that both components are activated simultaneously during language comprehension. The linguistic component reaches the peak of its activation through the distributional patterns, which is relatively faster than the activation of the embodied component and with less cost. The latter, on the other hand, requires deeper processing of situated simulation. As a result, this temporal advantage of the linguistic component could be utilised during conceptual representation depending on the current context or task demands. In other words, conceptual representation is flexible enough to adjust the reliance on the embodied versus the linguistic component in order to satisfy specific task demands (Connell & Lynott, 2014; Lynott & Connell, 2010).

Taken together the fact that the linguistic component can provide a fuzzy approximation of the embodied component, and that the former is activated faster than the latter, it was hypothesised that the linguistic component could act as a shortcut to guide conceptual representation before the more costly embodied component is fully engaged, particularly when superficial processing could suffice (Connell, 2018; Connell & Lynott, 2013). Support for this linguistic shortcut hypothesis should contain two parts. First, the linguistic component should enjoy greater importance during language processing when the task allows relatively shallow and speedy processing. Second, the linguistic component should reach the peak of its activation prior to the embodied component.

Evidence for the former requirement was found with the property verification task and the conceptual combination task. The linguistic distributional pattern among sensory properties (e.g., co-occurrence frequencies between *loud* and *bright*) was powerful enough to account for faster responses to these properties, but not for the slower responses (Louwerse & Connell, 2011). The distributional pattern could also account for responses to novel concepts (e.g., *cactus beetle*) when people performed a simple yes/no sensibility judgement task instead of an interpretation generation task in which they needed to provide a verbal interpretation (Connell & Lynott, 2013). To support the temporal advantage of the linguistic component, Louwerse and Hutchinson (2012) found in an EEG study that neural regions involved in linguistic processing (e.g., left inferior frontal gyrus) was activated to a larger degree than the regions involved in perceptual processing (near the lingual gyrus) at the early stage of conceptual processing, while the pattern was reversed at the later stage.

In summary, conceptual representation involves both simulation and linguistic information, with the linguistic component acting as a shortcut depending on the demands of the task. As conceptual representation should underlie all types of language processing, we expect the same pattern to be found during metaphor processing as well. In this study, we investigate the interplay between embodied and linguistic components in metaphor processing. Furthermore, we trace their roles during real-time processing using event-related EEG potentials (ERP).

Simulation and Linguistic Representations of Metaphors

Recent studies have viewed metaphor processing as an aspect of language processing in general, and placed it under the same lens of conceptual representations (Liu, Connell, & Lynott, 2017, 2018c). In those studies, participants read adjective metaphors (e.g., \emph{Student can be bright, *Supporters can be lukewarm*) that are controlled on both embodied and linguistic components. That is, these metaphors independently varied on two dimensions, which were the ease of simulation (EoS) and the linguistic distributional frequencies (LDF). EoS was a novel norming measure of how much effort it took for people to successfully arrive at a conceptual representation (Liu, Connell, & Lynott, 2018b), which combined the measures of three scales, that is how easy it was to imagine, to use, and to make sense of the metaphors. The combined EoS measure could outperform any of the individual scale in predicting the performance to a sensibility judgement task. This advantage suggested that EoS measured the underlying mechanism of metaphor comprehension, which was assumed to be simulation. On the other hand, LDF was a minimal measure of the linguistic distributional patterns, quantified the bi- to 5-gram frequencies of the constituent words (e.g., *student* and *bright* with zero, one, two or three intervening words).

Just as what was found for language processing, both variables were found to contribute to a unique portion of variance of performance during metaphor processing tasks. People primarily based their performance on the embodied component, as they attempted to form a coherent simulation. They found it easy to accept a metaphor when it was considered easy to simulate, and easy to reject a metaphor when it was considered hard to simulate. The effect of EoS varied based on the depth of processing and the time available for responses. When participants needed to provide an interpretation to the metaphors (i.e., during deeper processing) or when they were allowed more time, EoS showed a larger effect than when they performed a yes/no sensibility judgement (i.e., during relatively shallower processing) or when they were under limited time constraint.

On the other hand, the linguistic component also affected people's judgments. When the constituent words often co-occurred (e.g., *bright* and *student*), people could reach a decision to accept the metaphor straightaway, without further expending much effort on simulation. Moreover, information of the linguistic component could guide the activities of embodied simulation. LDF moderated the effect of EoS when people rejected a metaphor as nonsensical (Liu et al., 2018c). High LDF could signal that the processing was likely to succeed, thus making people spending more effort on simulation, even when processing turned out to be difficult. However, the linguistic shortcut hypothesis was not fully supported because, opposite to the predictions, people relied on the linguistic information even more during deep processing and a longer response interval. Nevertheless, these studies showed that the embodied and linguistic components were variably engaged based on required depth of processing and the length of response window. These results highlighted the fact that conceptual representation could be highly task-dependent (Connell & Lynott, 2014; Lebois, Wilson-Mendenhall, & Barsalou, 2015), and subject to the strategies chosen by participants.

As previous studies did not support the linguistic shortcut hypothesis fully, it is necessary to replicate the results and to investigate the actual role of the linguistic component during metaphor processing, which is the aim of the current study. One of the ways to paint a fuller picture of the role of the linguistic component is to study the timecourse of its activation during online processing using EEG.

ERP Markers of Metaphor Processing

EEG could help us in the study of the timecourse of linguistic and embodied components during online processing. To support (or reject) the linguistic shortcut hypothesis, it is important to show that the linguistic component in fact reaches the peak of its activations before the embodied component. EEG proves to be a useful tool in this respect because it reveals the real-time activities of the brain during conceptual processing (Amsel, 2011). Previous studies have examined the neural activities during metaphor processing (e.g., Arzouan, Goldstein, & Faust, 2007; De Grauwe, Swain, & Holcomb, 2010; Pynte, Besson, Robichon, & Poli, 1996), but have rarely differentiated the effect of the linguistic component from the effect of the embodied component.

ERP Markers of Simulation versus Linguistic Information

Liu, Connell, and Lynott (2018a) explored the ERP manifestations of embodied and linguistic components in literal language processing. They found that the neuroelectrical activities associated with the embodied component were different from that of the linguistic component. The former were associated mainly with greater activities around 400ms after the stimulus onset in the left parietal region. These activations coincided with the classical N400 effect, which was a marker of greater effort in semantic representation (Kutas & Federmeier, 2011). In contrast, the linguistic component was activated at first around 300ms from the stimulus onset at the frontal region, and again after 600ms in the posterior (i.e., parietal and occipital) regions, which might suggest that the linguistic component was associated with the early stage of semantic processing (Collins, Pecher, Zeelenberg, & Coulson, 2011) as well as conflict resolution (Davenport & Coulson, 2011; Paczynski & Kuperberg, 2012) respectively.

Present Study

In this study, our aim was to replicate and extend previous findings on the linguistic-simulation representations of metaphors. Since we are particularly interested in testing the efficacy of linguistic information, we kept the measure minimal as in previous studies (Liu et al., 2017, Liu et al. (2018c)), using co-occurrence frequencies within a 5-word window only. In the behavioural study (Experiment 1), we will replicate the same tasks as previous studies (i.e., the shallow sensibility judgement task, and deep interpretation generation task). Furthermore, to address the possibility that the effect of the tasks in previous studies could be confounded with the differences between participants, we conducted a within-participant design. In this pre-registered experiment (https://aspredicted.org/ff8ms.pdf), we expected to fully replicate the effects of EoS from previous studies. Also we hypothesised that if the linguistic shortcut hypothesis was true, LDF would affect the speed of acceptance on its own. EoS and LDF would be moderated by tasks, such that EoS would have a larger effect in the deep task than the shallow task, and vice versa for LDF. Furthermore, linguistic information should moderate the activities of embodied simulation, especially when a metaphor was rejected.

Then in the EEG study (Experiment 2), we will investigate the continuous neuroelectrical activities of embodied simulation and the activation of linguistic distributional patterns. We expect EoS to reach the greatest effect around 400ms in the left parietal region, replicating the ERP marker of the embodied component in the behavioural study. In contrast, LDF should have a greater effect early in the trial (around 300ms), thus supporting the linguistic shortcut hypothesis; as well as later after 600ms in a similar way as during literal language processing, because metaphor processing should involve the resolution of conflicting representations (Coulson & Van Petten, 2002). Since these hypotheses were proposed based on the previous study on literal language processing, we expect the results the exceed the current hypothesised regions. As a result, these hypotheses are better considered as tentative, and we will explore the neuroelectrical activities beyond these areas of interest.

Experiment 1: Behavioural Study

A pre-registered behavioural study here (https://aspredicted.org/ff8ms.pdf) tests the existence and the role of embodied and linguistic information during metaphor processing. In this study, participants performed both a shallow sensibility judgement task and a deep interpretation generation task, which was hypothesised to alter the reliance on the two components. Specifically, the linguistic shortcut hypothesis suggested that shallow processing would encourage the use of the linguistic component. In this section, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Method

Participants. Forty-eight students from Lancaster University participated in the experiment (age: 20.96 (SD = 3.51); female: 31; right-handed: 37). All were native speakers of English, speaking English as their first language since infancy. The size of sample was pre-determined to match the previous between-participant study (Liu et al., 2018c).

Materials. 452 metaphors in the form of "NOUN can be ADJECTIVE" (e.g., *Student can be bright*) were used as stimuli. All adjectives were perceptual adjectives pertaining to vision, audition or haptics (Dantzig, Cowell, Zeelenberg, & Pecher, 2011; Lynott & Connell, 2009). Each adjective was paired with four nouns so that the metaphor varied on two dimensions, ease of simulation (EoS) and linguistic distributional frequency (LDF). EoS was operationalised in a norming study (Liu et al., 2018b) which extracted the principle component of three different ratings of the metaphors: sensibility rating (How much sense does the sentence make when you read it in text or hear it in conversation?), usability rating (How easy is it for you to use the sentence in text or conversation?), and imaginability rating (How easy is it for you to imagine the concept described in the sentence?). These three ratings measured the effort to arrive at a mental representation, which, in line with the embodied account of language processing (Barsalou et al., 2008; e.g., Zwaan, 2004, 2014), is constituted of the simulation of sensorimotor, affective, and other information. Hence, we assumed that the composite measure of these ratings measured the underlying embodied simulation during language comprehension, and thus labeled it ease of simulation (EoS). For instance, *Student can be bright* (EoS = 1.84) and *Solution can be bright* (EoS = 1.41) were considered to be easy to simulate; while *Illness can be bright* (EoS = -1.32) and *Supply can be bright* (EoS = -1.02) were considered to be hard to simulate.

LDF, on the other hand, was defined by the bi- to 5-gram frequencies of each pair of noun and adjective in the Google Web1T Corpus, which had 1 trillion words (Brants & Franz, 2006). The raw frequency was logarithmically transformed as lg(N +41), where 41 is the lowest non-zero frequency in the corpus. For example, *Student can be bright* (LDF = 4.08) and *Supply can be bright* (LDF = 3.72) had higher distributional frequencies than *Solution can be bright* (LDF = 3.11) and *Illness can be bright* (LDF = 2.94). The two dimensions (i.e., EoS and LDF) had a correlation of $r = .27$, sharing approximately 7.29% of variance, and they were separable using statistical measure which will be discussed later in the Results section. The full set of norms can be accessed at https://osf.io/xgysz.

The 452 sentences could be divided into four lists, in which each adjective appeared only once. The EoS and LDF matched across the lists (EoS: $F_{(3,448)} = 0.03$, $p = .99$; LDF: $F_{(3,448)} = 0.09$, $p = .96$). Thus, two lists of sentences were used for the whole experiment. As the whole experiment was composed of two tasks, participants saw two lists of sentences, one for each task (i.e., each adjective appeared twice throughout the experiment, paired with different nouns).

Procedure. All participants performed two tasks in a counterbalanced order, a sensibility judgement task (or shallow processing condition) and an interpretation generation task (or deep processing condition). Sentences in each task were presented in a random order. Ten extra items were used before each task as practice. In both the practice and the test sessions, participants read the sentences one word/phrase at a time (Figure 1). In the shallow sensibility judgement task, participants were asked to decide whether the sentence was sensible or not. They started by seeing a fixation cross at the beginning of the trial, followed by the noun, “can be”, and then the adjective. The adjective stayed on the screen for 200ms followed by a blank screen for 2800ms during which a response should be made. If they judged the sentence as sensible, participants were asked to press the comma (,) key; and if they judged the sentence as nonsensical, the full stop key (.). Once a response was made, the fixation cross would appear on the screen to signal the beginning of the next trial. However, if no response was made within the length of the response interval, a feedback warning would appear on the screen saying ***TOO SLOW***.

The procedure of the deep interpretation generation task was the same as the shallow sensibility judgement task from the fixation cross until the adjective. After the adjective disappeared, a blank screen remained for 7800ms as a response interval, during which participants were asked to think of an interpretation for the sentence and make a response once they had thought of an interpretation or decided they could not think of an interpretation. The response key was the same as the shallow task. If participants indicated that they had thought of an interpretation for the sentence, after pressing the comma key, a blank screen would appear for them to type down their interpretation of the sentence. They were required not to simply replace the adjective with a synonym. After an interpretation was provided, they would press Enter to proceed to the next trial; and if they indicated that the sentence was not interpretable, they would automatically proceed to the next trial. Like in the shallow task, if no response was made within the length of the response interval, participants would receive a feedback warning. In both tasks, their response decision (accept/reject the metaphor) and the response time (RT) were recorded for analyses. There were two self-administered breaks within each task and one break between tasks.

Design and analysis. The response decisions and RT were analysed as pre-registered. Response decisions (coded 0 for “no” and 1 for “yes”) were analysed in a mixed-effect logistic regression.[1] Task (coded 0 for sensibility judgement and 1

---

[1] We used R (Version 3.5.0; R Core Team, 2017) and the R-packages *abind* (Version 1.4.5; Plate & Heiberger, 2016), *arm* (Version 1.10.1; Gelman & Su, 2018), *BayesFactor* (Version 0.9.12.4.2;

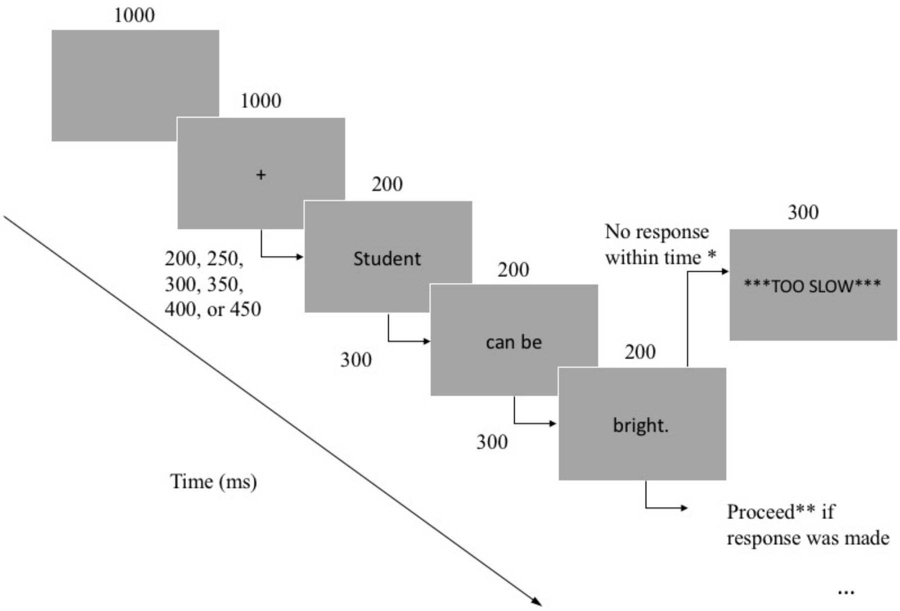


Figure 1. Presentation sequence in Experiment 1. * Response time window depended on task. Shallow sensibility judgment task allowed 3s, and deep interpretation generation task allowed 8s. ** After response, proceed to next trial in shallow sensibility judgment task; and in deep interpretation generation task, proceed to type down the interpretation if answered “yes”, and to next trial if answered “no”.

for interpretation generation), EoS and LDF, with their interactions, were used as

Morey & Rouder, 2015), *bookdown* (Version 0.7; Xie, 2016), *broom* (Version 0.4.4; Robinson, 2018), *coda* (Version 0.19.1; Plummer, Best, Cowles, & Vines, 2006), *contrast* (Version 0.21; Kuhn, Steve Weston, Wing, Forester, & Thaler, 2016), *cowplot* (Version 0.9.2; Wilke, 2017), *data.table* (Version 1.11.4; Dowle & Srinivasan, 2017), *doBy* (Version 4.6.1; Højsgaard & Halekoh, 2018), *dplyr* (Version 0.7.5; Wickham, François, Henry, & Müller, 2018), *Formula* (Version 1.2.3; Zeileis & Croissant, 2010), *ggplot2* (Version 2.2.1; Wickham, 2009), *gridExtra* (Version 2.3; Auguie, 2017), *Hmisc* (Version 4.1.1; Harrell Jr, Charles Dupont, & others., 2018), *interplot* (Version 0.1.5; Solt & Hu, 2015), *knitr* (Version 1.20; Xie, 2015), *lattice* (Version 0.20.35; Sarkar, 2008), *lme4* (Version 1.1.17; D. Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.0.1; Kuznetsova, Brockhoff, & Christensen, 2017), *magrittr* (Version 1.5; Bache & Wickham, 2014), *MASS* (Version 7.3.50; Venables & Ripley, 2002), *Matrix* (Version 1.2.14; D. Bates & Maechler, 2017), *mgcv* (S. N. Wood, 2003, 2004, Version 1.8.23; 2011; S. Wood, N., Pya, & S"afken, 2016), *multcomp* (Version fixed variables; participant and item were used as crossed random variables modelling random intercepts.[2] RT was analysed first in an omnibus mixed-effect linear regression, with task, response decision, EoS, LDF and their interactions as fixed predictors; participant and item as crossed random variables modelling random intercepts.[3]

Then RT was separated by the response decision to test the specific hypotheses for different responses, and each analysed in a mixed-effect linear regression. Further analyses of response decisions and RT were conducted in mixed-effect models which modeled random intercepts and slopes (not pre-registered). These analyses were not pre-registered but deemed appropriate because mixed-effect models with random intercepts and slopes would generalise better than a random-intercept-only model (Barr, Levy, Scheepers, & Tily, 2013). However, a full maximal model did not always

---

1.4.8; Hothorn, Bretz, & Westfall, 2008), *mvtnorm* (Version 1.0.8; Genz & Bretz, 2009), *nlme* (Version 3.1.137; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017), *papaja* (Version 0.1.0.9735; Aust & Barth, 2017), *plyr* (Wickham, 2011; Version 1.8.4; Wickham et al., 2018), *psych* (Version 1.8.4; Revelle, 2017), *purrr* (Version 0.2.5; Henry & Wickham, 2018), *rcartocolor* (Version 0.0.22; Nowosad, 2017), *rms* (Version 5.1.2; Harrell Jr, 2018), *SparseM* (Version 1.77; Koenker & Ng, 2017), *survival* (Version 2.42.3; Terry M. Therneau & Patricia M. Grambsch, 2000), *TH.data* (Version 1.0.8; Hothorn, 2017), and *tidyr* (Version 0.8.1; Wickham & Henry, 2017) for all our analyses and the writing up of this manuscript.

[2] Model including participants as random intercepts explained the data better than the empty model by 4.88 x 10175 times (BF_{10} = 4.88 x 10^{175}); model with participants and item as crossed random intercepts further better explained the data than the model with only participants as random intercept (BF_{10} > 1.80 x 10^{308}); inclusion of fixed predictors further better explained the data by 1.80 x 10^{132} times (BF_{10} = 1.80 x 10^{132}).

[3] Model including participants as random intercepts explained the data better than the empty model (BF_{10} > 1.80 x 10^{308}); model with participants and item as crossed random intercepts in fact worsen the model (BF_{10} = 0.01) but was still used for consistency; inclusion of fixed predictors further better explained the data (BF_{10} > 1.80 x 10^{308}).

converge. Therefore, we looked for the maximal random effect structure justified by the sample for both response decisions and RT data, and reported it here, if they provided a better fit than the random-intercept models. The random-intercept models can be found in Supplementary Materials (https://goo.gl/sRkXun).

In both the present experiment and Experiment 2, we used the Bayes Factor (BF), instead of the Chi-squared test, to decide whether the addition of new random or fixed variables improved a model. BF uses the Bayesian approach to calculate the amount of evidence the data lends to the alternative hypothesis in relations to the amount of evidence for the null hypothesis. BF_{10} denotes the ratio of the likelihood of the data under the alternative hypothesis against the null hypothesis. If BF_{10} is 5, it means that the data is 5 times more likely to occur under H_1 than H_0. Therefore, BF can inform us about quality of the hypotheses. We can be more confident with the alternative hypothesis if $BF_{10} = 100$ than if $BF_{10} = 10$. Inversely, if BF_{10} is less than 1, we would be more confident about the null hypothesis than the alternative hypothesis. As a result, the BF is superior to *p*-values produced by null-hypothesis significance testing (NHST), which only tells us the probability of Type I error but not how good the alternative hypothesis is. See Jarosz and Wiley (2014) and Wagenmakers (2007) for detailed discussions.

The BF of regression models can be easily obtained in R by extracting the Bayesian information criteria (BIC). BF_{01} (comparing H_0 against H_1) is calculated as:

$$BF_{01} = e^{\frac{(BIC(H_1) - BIC(H_0))}{2}}$$

BF_{10}, which denotes the likelihood of H_1 against H_0 is simply the inverse of BF_{01}. While the continuous scale of BF is useful enough, we could conveniently describe the results, using Jeffreys (1998)'s guideline to characterise BF (Table 1).

Table 1

Interpretation of BF as evidence for H 1

BF_{01}	BF_{10}	Support for H_1
1-.33	1-3	Anecdotal
.33-.10	3-10	Substantial
.10-.03	10-30	Strong
.03-.01	30-100	Very strong
<.01	>100	Decisive

Results and Discussion

No participants were deleted for slow response (> 3SD from grand mean). 1.35% of trials were missed (i.e., no response within the time limit), which were deleted. Eight trials were deleted for motor error (RT < 200ms). Beyond that, 39 trials in the deep processing task were marked out as invalid interpretations (blank, "0", "wrong button", etc.). Finally, we removed individual trials with outlying RT (> 3SD from individual means per response per task) that was 0.80% and 1.34% of "yes" and "no" responses respectively in the shallow task; 1.05% and 1.69% of "yes" and "no" response respectively in the deep task. Consequently, there remained 5294 data points for the shallow task and 5199 for the deep task.

Response decision. The general acceptance rate was 46.07%. That was 47.02% of the shallow task and 45.10% of the deep task. Mixed-effect logistic regressions were conducted. The random-intercepts model as per pre-registration showed different results from the maximal random-slopes model justified by the data, but the latter explained the data better than the former (BF_{10} = 8.57 x 10^{32};

Table 2

Logistic mixed-effect regression of response decision

	b	95% CI
Intercept	-0.28	-0.58 - 0.02
Task	-0.15	-0.39 - 0.1
LDF	0.07	-0.01 - 0.15
EoS	1.32	1.16 - 1.48
Task : LDF	0.05	-0.03 - 0.13
Task : EoS	0.01	-0.08 - 0.1
LDF : EoS	0.05	-0.02 - 0.13
Task : LDF : EoS	-0.01	-0.1 - 0.07

Note. b is non-standardised regression coefficient.

detailed explanation of BF in Experiment 2). Therefore, we reported the maximal model here (Table 2), that is participant as a random predictor to model the intercept and slope of task, LDF and EoS without interactions; and item as a random predictor to model the intercept and slope of task LDF and EoS with interactions). The results showed that only EoS, among all fixed predictors, had a significant effect on response decisions. As EoS increased by one unit, the odds to accept the metaphors increased 3.75 times. The random-intercepts model showed that beyond the effect of EoS, EoS interacted with task and LDF also had a positive effect on response decision (see Supplementary Materials).

Response time. The grand mean of RT was 1391ms (SD = 885ms), which was 1114ms (SD = 455ms) for shallow task and 1674ms (SD = 1102ms) for deep

Table 3

Mean (SD) RT in ms per task per response decision

Task	Shallow	Deep
Yes	1118 (457)	1815 (1085)
No	1110 (454)	1557 (1102)

task. The average RT per response decision per task was shown in Table 3.

Analyses of RT found evidence of net suppression (J. Cohen, Cohen, West, & Aiken, 2003), which means that the shared variance of LDF and EoS was masking the true relationship between the variables and RT. Because LDF and EoS correlated to some degree, LDF contributed to the fit of the model by explaining the error variance of EoS. As a result, the effect of EoS was enhanced while the effect of LDF was diminished. This could be seen by the fact that the effect of LDF in the LME model was the opposite of its zero-order correlation with RT. As seen in Supplementary Materials, LDF had a negative effect on the speed to reject a metaphor (i.e., "no" RT) in the mixed-effect model, but a positive correlation with RT according to the zero-order correlations. Therefore, as pre-registered, we orthogonalised EoS and LDF using a principle components analysis (PCA) with varimax rotation and Kaissar covariance normalisation. PCA managed to obtain two perpendicular variables each corresponding to one of the original variables. This procedure solved the problem of net suppression because the orthogonal variables did not have any shared variance. The orthogonal variables obtained by PCA did not lose any information from the original variables, and they correlated with each variable (r = .99). The orthogonalisation procedure was done both for the omnibus regression and separately with the RT to "yes" and "no" responses for the analyses per response. The results reported here used only orthogonal variables.

The omnibus regression was shown in Table 4. Again, the maximal model justified by the data provided a massively better fit than the random-intercept model[4] ($BF_{10} > 1.80 \times 10^{308}$; participant as a random predictor to model the intercept and slope of task; and item as a random predictor to model the intercept only)), so the former was reported here and the latter in Supplementary Materials. EoS had a positive effect when LDF was at its mean, which was moderated by task and response. It showed that EoS had a positive effect for both tasks, which was larger for the deep task than the shallow task. The effect was positive for “no” response but negative for “yes” response. In contrast, LDF did not have an effect when EoS was at its mean regardless of task and response. However, LDF did interact with EoS when response was “no” in the sensibility judgement task, which was attenuated by “yes” response and the deep interpretation generation task, as will be discussed in separate models below.

Because of the interaction between response decision and other variables, and our hypotheses that opposite effects would exist for “yes” and “no” responses, we separated the data into two sets and analysed the RT per response with re-orthogonalised variables (Table 5). For “yes” response (i.e. when participants accepted the metaphors), both EoS and LDF had a negative effect on both tasks. That is as EoS or LDF increased, people were faster at accepting a metaphor. Only EoS interacted with task, which means that it had a larger (more negative) effect when people needed to generate an interpretation for the metaphor than when they needed to make a simple “yes/no” judgement concerning sensibility; whereas the interaction between LDF and task was not significant meaning that the effect of LDF did not change between tasks.

[4] The largest number calculable on a computer is 2^{1024}, which is approximately 1.80×10^{308}.

Table 4

Omnibus linear mixed-effect regression of RT

	b	95% CI
Intercept	1424.55	1293.91 - 1555.2
Task	411.02	283.14 - 538.89
Response	28.53	-1.2 - 58.25
LDF	2.4	-17.76 - 22.56
EoS	58.51	37.49 - 79.53
Task * Response	19.4	-22.48 - 61.28
Task * LDF	15.82	-11.64 - 43.29
Response * LDF	-10	-37.86 - 17.86
Task * EoS	49.02	20.53 - 77.51
Response * EoS	-189.33	-217.94 - -160.71
LDF * EoS	-31.84	-50.88 - -12.81
Task * Response * LDF	-16.08	-55.36 - 23.21
Task * Response * EoS	-101.39	-141.15 - -61.64
Task * LDF * EoS	-6.27	-32.04 - 19.51
Response * LDF * EoS	22.49	-3.91 - 48.89
Task * Response * LDF * EoS	4.6	-32.04 - 41.24

Note. *b* is non-standardised regression coefficient.

Table 5

Mixed-effect linear regression model of RT per response decision

	Yes		No	
	b	95% CI	*b*	95%CI
Intercept	1465.66	1336.17 - 1595.14	1457.7	1304.26 - 1611.13
Task	476.63	345.85 - 607.42	431.55	280.32 - 582.78
LDF	-25.03	-45.81 - -4.25	14.8	-5.35 - 34.95
EoS	-158.14	-191.16 - -125.13	67.32	35.1 - 99.54
Task * LDF	-9.64	-39.56 - 20.29	17.48	-8.74 - 43.69
Task * EoS	-71.75	-102.57 - -40.93	41.21	14.55 - 67.87
LDF * EoS	-8.14	-27.8 - 11.53	-27.29	-44.89 - -9.69
Task * LDF * EoS	-3.78	-32.27 - 24.71	-5.48	-29.18 - 18.22

Note. *b* is non-standardised regression coefficient.

'Yes' RT maximal model over random-intercepts model: $BF_{10} = 2.80x10^{194}$.

'No' RT maximal model over random-intercepts model: $BF_{10} > 1.80x10^{308}$

For "no" response (i.e. when people rejected the metaphors), EoS had a significant, positive effect when LDF was at its mean, that is as ease of simulation increased, it took people longer to reject a metaphor. This effect of EoS interacted with task as well, such that EoS had a larger (more positive) effect for the deep task than the shallow task. LDF further interacted with EoS with a negative term, such that as LDF increased, the effect of EoS decreased. This interaction remained the same for both shallow and deep processing (Figure 2).

Summary. EoS strongly affected metaphor processing as predicted. As metaphors become easier to simulate, the likelihood and speed of accepting them increased regardless of LDF; whereas the speed of rejecting them decreased, which

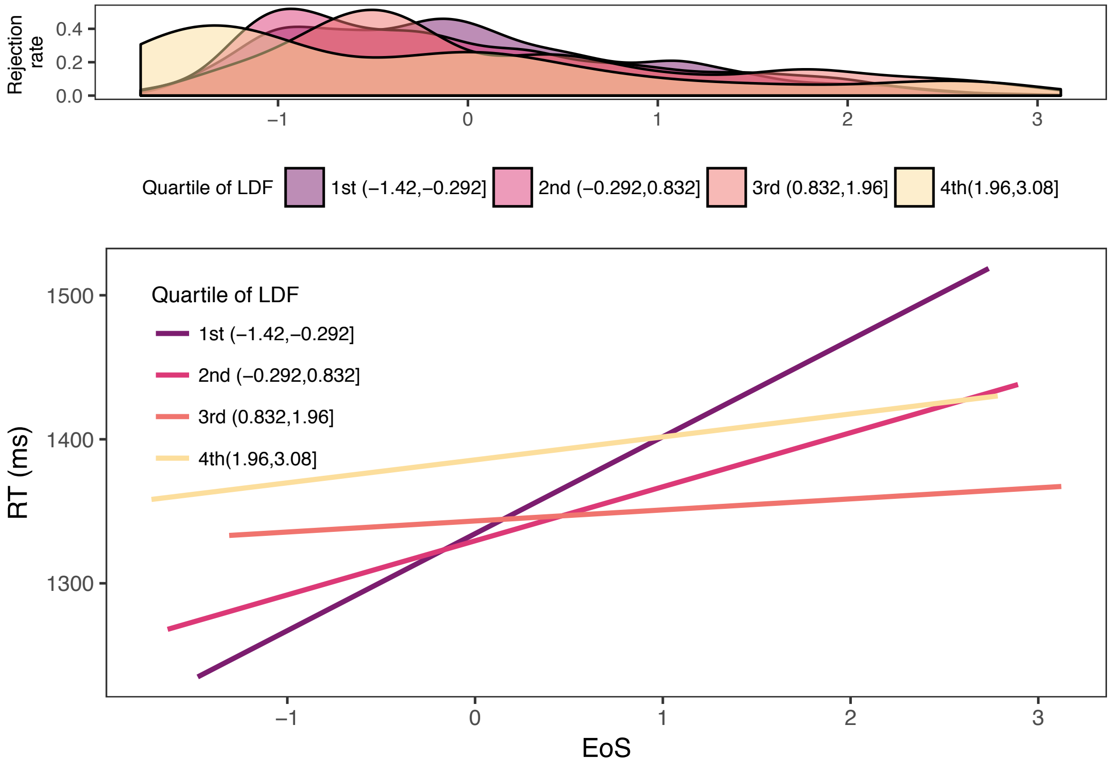


Figure 2. Effects of EoS on the “no” RT at four quartiles of LDF. As LDF increased from the 1st quartile (violet) to the highest 4th (yellow), the effect of EoS flatlined. The density plot on the top shows the proportion of “no” response along EoS per quartile of EoS.

was attenuated when LDF increased. That is to say, when the constituent words did not often co-occur, the easier it was to simulate a metaphor, the harder it was to reject it; but when constituent words often co-occurred, the effect of EoS decreased and flatlined.

LDF, though not affecting response decisions, had a significant effect on RT as expected, such that as constituent words co-occurred more often, the speed of accepting metaphors became faster. LDF also had an effect on the rejection speed as expected, but the effect was only limited to low EoS. In other words, when a metaphor was hard to simulate, the more often the constituent words co-occurred, the slower it was for people to reject it; but when a metaphor was easy to simulate, the effect disappeared. As a result, it was fastest to reject a metaphor when both EoS and LDF were low, and slowest when LDF was low but EoS was high.

Task was found to moderate EoS as expected. The effect of EoS was larger for the deep interpretation generation task than for the shallow sensibility judgement task. In contrast, the effect of LDF was not affected by task, contrary to the linguistich shortcut hypothesis. The negative effect of LDF on the acceptance speed remained the same for both sensibility judgement and interpretation generation task, and the interaction between EoS and LDF on the rejection speed did not vary between tasks either.

Experiment 2: EEG Study

In this experiment, we will examine the neuroelectrical activations of embodied and linguistic components during metaphor processing. As this study was the first to the authors knowledge to study the simulation-linguistic representation of metaphor processing, we did not have any confirmatory hypotheses concerning the results. However, we proposed some tentative hypotheses concerning the timecourse and scalp regions of the activations. The activations of the embodied component will be manifested around 400ms after the adjective onset at the left parietal region. The activations of the linguistic component will be prior to those of the embodied component, possibly around 300ms in the frontal region; and they will re-emerge after 600ms in the parietal region. Although the experiment was not pre-registered, we will report here how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Method

Participant. Sixteen native speakers of English were needed to match previous studies on metaphor processing (e.g., Coulson & Van Petten, 2002) and the property verification task of (Liu et al., 2018a). One participant was excluded because the data was not recorded properly due to technical trouble, so one more was recruited to make up for the loss. All resulting usable participants were students of Lancaster University (age: 22.06 (*SD* = 6.13); female: 11; right-handed: 15).

Materials. The stimuli derived from the same set of metaphoric sentences as ones in Experiment 1. The only difference was that the sentences in this experiment did not end after the adjectives, but all had the word "*sometimes.*" as the end, in order to prevent the end of sentence wrap-up effect (Brouwer, Fitz, & Hoeks, 2012; Kutas, Federmeier, & Sereno, 1999), which could be more noisy because it took longer time and involved the integration of phrases and clauses and the processing of phrase boundaries. The sentences were distributed into six lists of 169 or 170 items, each composted of one and a half lists of items from Experiment 1. In each of these six lists in this experiment, half of the adjectives appeared twice pairing with two different nouns and half appeared only once.

Procedure. The items lists were used both in the sensibility judgement task and in the interpretation generation task. Participants took both tasks each with a different list, such that they saw different sentences (different nouns, same adjectives) between the two tasks. The order of the tasks were counterbalanced and the order of sentences were fully randomised. Participants were given four self-administered break in each task as well as one long break between tasks, resulting in five blocks of approximately 34 trials in each task.

At the beginning of each block, participants were given a four-second waiting period for the EEG recording to initiate. The trial structure was similar to Experiment 1 with few changes (Figure 3). Before the fixation cross, we added 1 second of blinking time during which a cue to blink remained on the screen. A

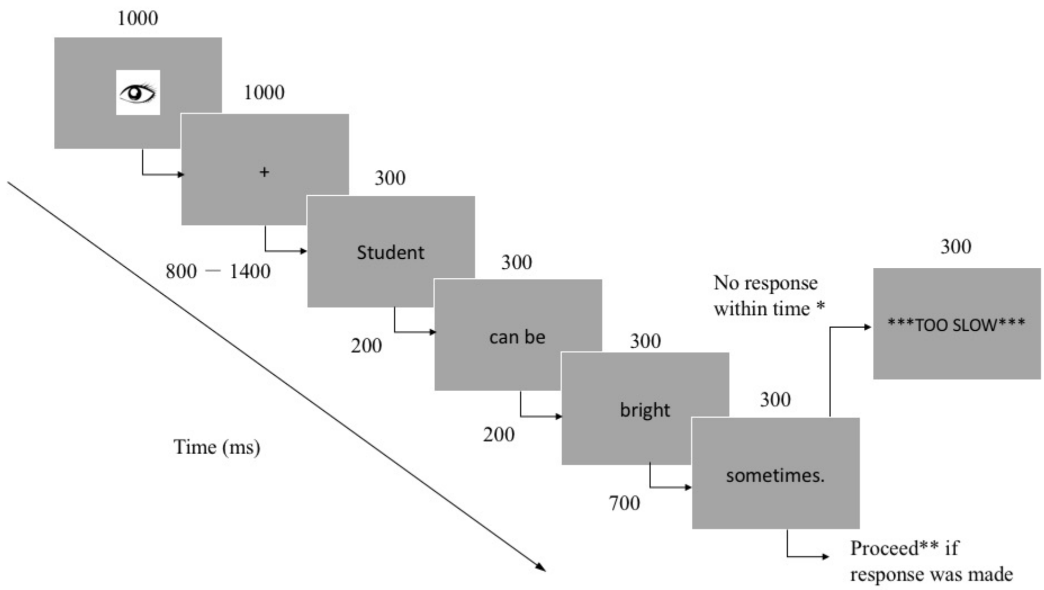


Figure 3. Presentation sequence in Experiment 2. * Response time window depended on task. Shallow sensibility judgment task allowed 3s, and deep interpretation generation task allowed 8s (including the word). ** After response, proceed to next trial in shallow sensibility judgment task; and in deep interpretation generation task, proceed to say out loud the interpretation if answered "yes", and to next trial if answered "no".

fixation cross followed for a variable amount of time between 800ms to 1400ms. Then, the nouns, "can be", and the adjectives word were presented sequentially. After the adjectives disappeared, a blank screen remained for 700ms during which participants did nothing but focusing on the screen, in order to eliminate the readiness potential which would occur if participants needed to make a response immediately after the adjectives (Dehaene et al., 1998). Following the blank screen after the adjective, the word "sometimes" would follow for another 300ms as the end of the sentence followed by a blank screen as the response interval (2700s for the shallow task and 7700ms for the deep task), during which participants were required to perform either the sensibility judgement task or the interpretation generation task as Experiment 1.

To make a response, participants were asked to press a red button at the position of “F” on the keyboard for “yes”, or a white button at the position of “J” for “no”. In the sensibility judgement task, the experiment will proceed to the next trial after the button press, but in the interpretation generation task, if participants pressed “yes”, they would see a sign of a microphone on the screen and they would say their interpretation of the sentence out loud. Their interpretations were recorded by an external Sony recorder. This was done differently from Experiment 1 to avoid contamination of the data by unwanted head movements during typing. After the interpretation was recorded, they pressed the space bar to move on to the next trial. Like in Experiment 1, if no response was made when the response interval had elapsed, a feedback warning was shown on the screen.

Apparatus. Words were presented on a 19-inch CRT monitor operating at 100Hz refresh rate using EPrime 1.0, and behavioural responses were recorded using a QWERTY keyboard. High-density EEG was recorded continuously using Hydrocel Geodesic Sensor Nets (Electrical Geodesics Inc., Eugene, OR, USA) at 128 locations referenced to the vertex (Cz) online. The ground electrode was at the rear of the head (between Cz and Pz). Electrophysiological signals were acquired at the sampling rate of 1000Hz by an Electrical Geodesics Inc. amplifier with a band-pass filter of 0.1-100Hz. The impedance was checked prior to the beginning of the recording and they were considered acceptable if lower than 50kΩ.

EEG analysis. The digitized EEG was further processed offline using NetStation v 4.5.4. It was band-pass filtered between 0.1- 30Hz and was segmented into epochs including 200ms before adjective onset and 1000ms after adjective onset. EEG epochs were automatically rejected for body and eye movements the signal exceeded ± 200 μV at any electrode. Data were further checked through visual inspection for eye-movements, eye-blinks and other body movement artifacts not detected by the automated algorithm. The artifact free segments were baseline-corrected with respect to the average amplitude in the 200 ms window proceeding stimulus onset, and were re-referenced to the average reference offline.

These segments were then exported to a comma-separated values (csv) file and analysed in R. In order to explore the continuous activations of embodied and linguistic components, we analysed the EEG output throughout the whole 1-second epoch after the adjective onset, and across the whole scalp (Amsel, 2011; Hauk, Pulvermüller, Ford, Marslen-Wilson, & Davis, 2009; Hendrix, Bolger, & Baayen, 2017; Liu et al., 2018a). The EEG output per participant per stimulus were sliced into 20 segments of 50ms and were averaged. The electrodes analysed were the 90 electrodes in the area on the scalp covered by the 20-20 EEG net (Appendix B). The resulting datapoints were analysed in linear mixed-effect models (LME) with participants and items as crossed random variables to predict random intercepts. The analyses were exceptionally computationally heavy and thus were carried out distantly using the High End Computing Cluster (HEC; http://www.lancaster.ac.uk/iss/services/hec/) of Lancaster University.

Using LME offers two advantages over the conventional method of analysis using ANOVA. First, it solves the language-as-fixed-effect fallacy (Clark, 1973) by accounting for the random variance from both participants and stimulus at the same time. Second, LME offers the possibility to take on continuous variables. In the present study, both critical variables (i.e., EoS and LDF) are continuous. Forcing them into distinct categories (e.g., high frequency vs low frequency) causes the loss of data, and may mask the true relationship between the predictor and the dependent variable (Amsel, 2011; Baayen, Davidson, & Bates, 2008). In a word, using LME to treat EEG data can avoid the inflation of Type I error rate and provide greater power.

Because of the behavioural results from Experiment 1, we have reason to believe that EoS and LDF will have different effects on "yes" and "no" response.

Therefore, in the analysis of ERP, we treated the key variables in separate models per response decision. We created LME with task (deep interpretation generation coded as 0; shallow sensibility judgement coded as 1) and the critical variables. Following Experiment 1, we also orthogonalised EoS and LDF per response. In order to ascertain evidence in favour of (or against) each key variable, we conducted model comparison in a hierarchical fashion. The baseline model was composed of participants and items as crossed random variables modelling random intercepts, and no fixed effects. Then in each hierarchical step, new fixed effects are added in and the new model was compared to the model in the previous step.

For the effects of EoS, the models were built in the following three steps.

Step 1: Baseline model + task as a fixed effect;

Step 2: Step 1 model + EoS as a fixed effect;

Step 3: Step 2 model + Interaction between task and EoS as fixed effect.

For the effects of LDF, the same three steps as above were repeated.

Two further model comparisons were conducted to extract the relative advantages of EoS and LDF:

1. Comparison between the EoS model obtained from Step 2 (i.e., EoS + task as fixed effects) with the LDF model obtained from Step 2;
2. Comparison between the EoS model obtained from Step 3 with the LDF model obtained from Step 3.

We again used BF to conduct these model comparisons. Apart from the superiority over p-values as discussed in Experiment 1, BF further offered a practical advantage specific to this experiment, that is BF does not require post hoc corrections of multiple comparisons. In this experiment, we need to conduct multiple analyses to the same set of data, which dramatically inflates the Type I error rate. Five models need to be built per time slice per electrode, resulting in 9000 models in total. Although a post hoc correction can be conducted (e.g., Bonferroni correction), it will render the test unnecessarily conservative. To keep the family-wise critical level as .05, a critical level of 5.56×10^{-6} is needed per model. This stringent criterion would render any *p*-values we can obtain meaningless. As a result, we chose to report only BF in this experiment together with the regression coefficients from LME models for effect size (results of NHST can be found in Supplementary Materials).

Results

Data cleaning. The data was cleaned in a similar way as Experiment 1. No participants were removed for slow responses. 0.98% of trials were missed and 3 trials in the interpretation generation task were removed due to invalid interpretations. No trials were removed because of motor error because by the time a response was made after the last word of the sentence, participants already had 1s to process the sentence. 5537 trials remained for the analysis, which included 2753 trials for the deep task and 2784 for the shallow task.

Behavioural results. Analyses of the behavioural data were conducted with participants and items as crossed random factors modelling random intercepts and slopes in a maximal model justified by the data. The response decision was analysed in a logistic mixed-effect regression[5] and RT was analysed in linear

[5] Model including participants as random intercepts explained the data better than the empty model by 7.87×10^{123} times ($BF_{10} = 7.87 \times 10^{123}$); model with participants and item as crossed random intercepts further better explained the data than the model with only participants as random intercept by 7.68×10^{174} times ($BF_{10} = 7.68 \times 10^{174}$); inclusion of fixed predictors further better explained the data by 3.21×10^{131} times ($BF_{10} = 3.21 \times 10^{131}$).

mixed-effect regressions[6] as Experiment 1. The results were largely similar to those of Experiment 1 (see Supplementary Materials), with differences in the effect of LDF on rejection speed. First as predicted, LDF had a positive effect on rejection speed when EoS was at the mean in this experiment (b = 26.54, 95% CI = [10.96, 42.12], df = 2817.30, t = 3.34, $p < .001$). Second, it also interacted with EoS, such that as LDF increased, the effect of EoS also increased (b = 19.32, 95% CI = [4.63, 34.00], df = 2817.60, t = 2.579, $p = .010$). Again, the effect of EoS increased in deep processing task, while there was no evidence that the effect of LDF increased in shallow processing task.

ERP results.

Overview.

The timecourse of the critical variables was illustrated by the mean BF by comparing EoS models with LDF models in Step 2 and Step 3 (Figure 4; all BFs are in the Supplementary Materials). Strong evidence supported the activations of EoS from 400ms and persisted until the end of the epoch for “yes” response. Two peaks of activations were first around 500ms, and then after 750ms. No strong evidence for LDF was found overall, that is no mean BF was below 0.1. At two points LDF seemed to have slightly outperformed EoS (i.e., $BF_{EL} < 1$), which were at 200ms when comparing the Step 2 models and at 600ms when comparing the Step 3 models. Further investigations into the topography of the key variables will show localised activations of the key variables in more details.

Task effects.

[6] Model including participants as random intercepts explained the data better than the empty model by 1.40 x 10^{117} times (BF_{10} = 1.40 x 10^{117}); model with participants and item as crossed random intercepts in fact worsen the model (BF_{10} = 0.01) but was still used for consistency; inclusion of fixed predictors further better explained the data by 1.01 x 10^{72} times (BF_{10} = 1.01 x 10^{72}).

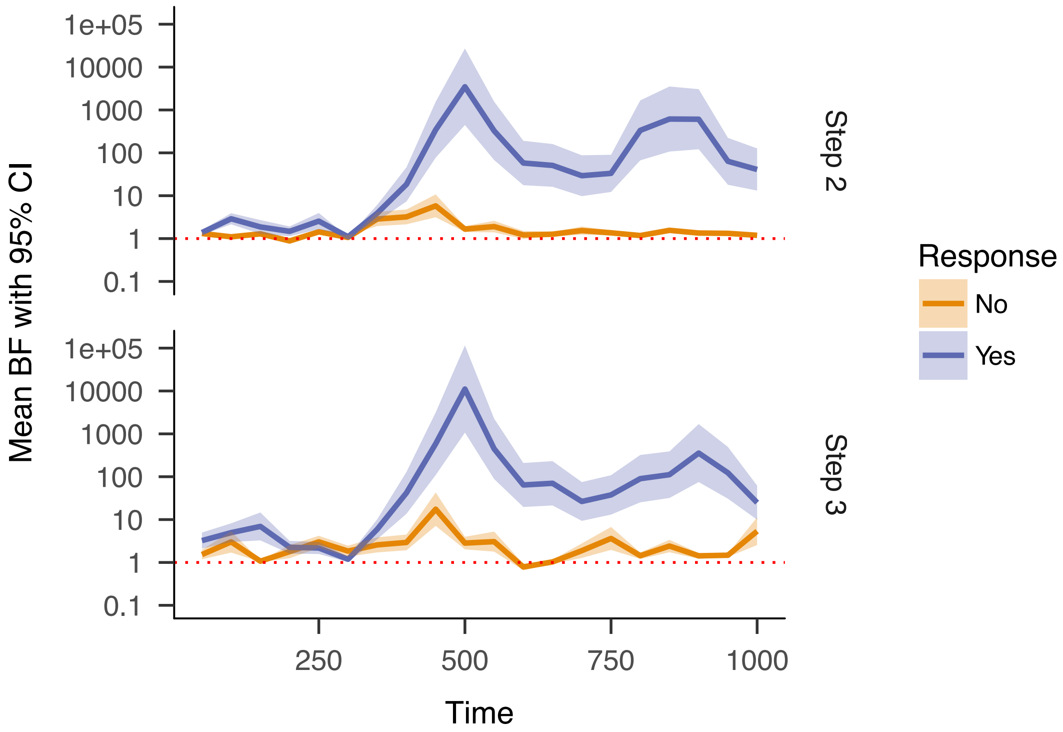


Figure 4. Logarithmically transformed mean BF_{EL} of models of EoS compared with models of LDF in Step 2 (which excluded the interaction terms) and Step 3 (which included the interaction term) with 95% CI. The red dotted line is the cutoff when $BF_{EL} = 1$ (i.e., EoS outperformed LDF when above the line, and LDF outperformed EoS below the line).

Step 1 (Figure 5) of model comparison showed strong evidence that task had an effect on EEG amplitude. For "yes" response, the effect started from 400ms after onset and lasted until the end of the epoch, peaking around 700ms; for "no" response, the effect only appeared around 650ms. In the following steps, we obtained evidence for EoS and LDF separately, above and beyond the effect of task.

EoS effects.

"Yes" response.

Our hypothesis for EoS stated that we should find strong effect of EoS in the left parietal area round 400ms (similar to a typical N400 effect). This was confirmed by Step 2 of model comparison for "yes" response, which showed evidence for EoS from 400ms to 600ms, when people accepted the metaphors (Figure 6 a and c). The data were on average 6.41 times (i.e., mean $BF_{21} = 6.41$) and 148.54 times maximally (max $BF_{21} = 148.54$) more likely to occur in models including EoS than the task-only models excluding EoS. EoS had a positive effect on amplitude, that is to say: metaphors that were harder to simulate elicited a more negative amplitude. From Panel c, it could be seen that in the left parietal region the effect was most prominent near the central line (e.g., E37: mean b between 400ms and 600ms was 0.56, mean $BF_{21} = 66.75$).

Further exploration of EoS's effects found them to go beyond the regions and time windows in the original tentative hypotheses. In fact, EoS had a strong and prolonging main effect across the centre of the scalp when people accepted the metaphors (Figure 6a). The strongest evidence for an EoS effect appeared at E7 in the central area, peaking at the 500ms slice ($BF_{21} = 1925.13$; $b = 0.78$). Later in the epoch, EoS again showed a positive effect supported by strong evidence in the parietal area (e.g., E67 at 850ms-900ms, $BF_{21} = 135.63$; $b = 0.79$).

"No" response.

In contrast, the data did not support the effect of EoS on "no" response. The ERP of "no" response was actually better explained by the Step 1 model with task only than the model with task and EoS together (Mean $BF_{21} = 0.02$; SD = 0.05). That is, on average, the EEG amplitude of rejected trials was 50 times more likely in the task-only model than the model with both task and EoS as predictors. This means that for those metaphors that were eventually rejected as nonsensical or uninterpretable, EoS did not affect the processing within the first 1000ms.

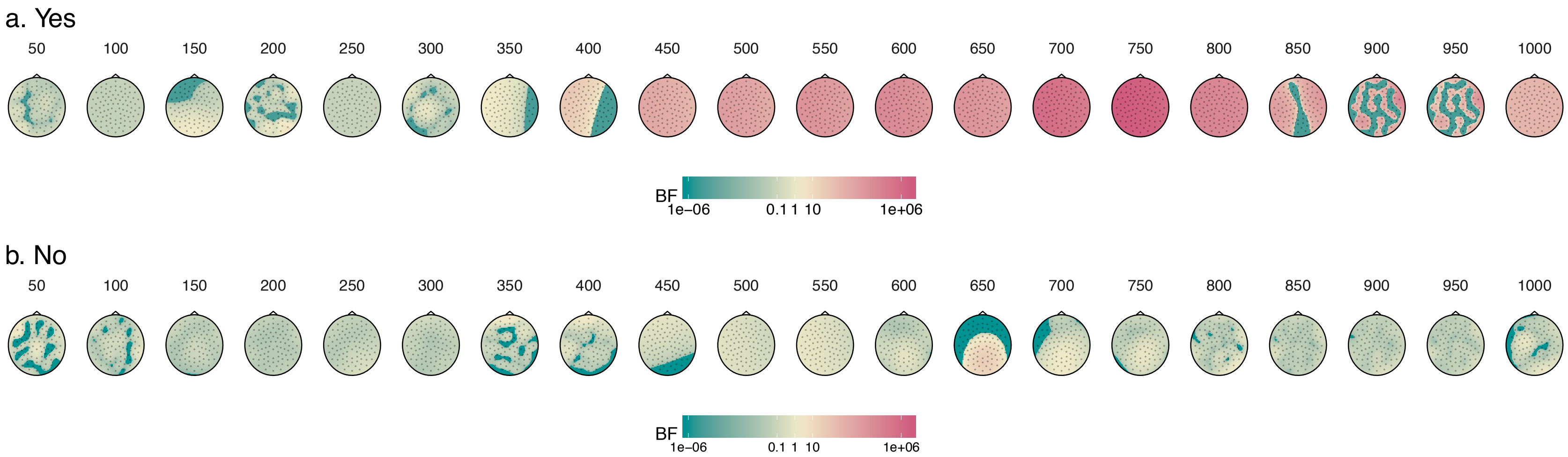


Figure 5. Task effect for a. "yes" response, and b. "no" response. When $BF_{10} > 10$, the data strongly supported the effect of task over the baseline model; whereas when BF_{10}, the data strong supported the baseline model over the effect of task. BF was logarithmically transformed for plotting.

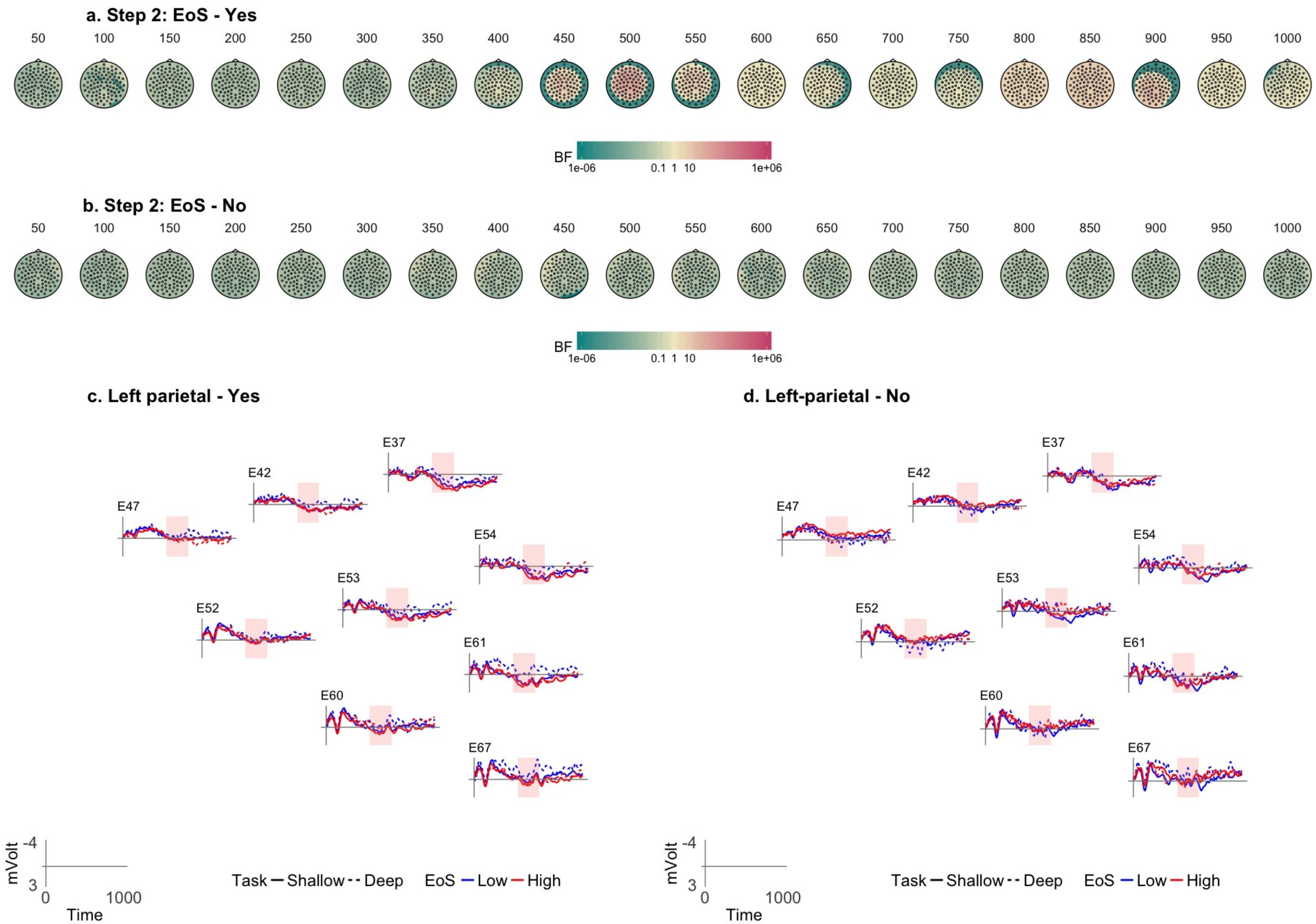


Figure 6. Evidence for the main effects of EoS measured by BF for a. "yes", and b. "no" response. When $BF_{21} > 10$, the data strongly supported the effect of task over the baseline model; whereas when $BF_{21} < 0.1$, the reverse is true. BF was logarithmically transformed for plotting. The waveforms at the left-parietal region are for "yes" (c) and "no" (d) responses respectively.

LDF effects.

For LDF, our hypotheses stated that we should find evidence for its effect in two areas. In temporal order, first we should find an effect before 350ms in the frontal and parietal area, prior to the window when the effect of EoS reached its peak. Second, we should find an effect after 600ms to the end of the epoch in the right posterior (i.e., parietal, occipital) regions. We did not find any evidence in Step 2 for LDF's effect, either in these hypothesised regions or on the whole scalp (Figure 7). In fact, we found evidence against the effect of LDF above and beyond the effect of task ("Yes": Mean $BF_{21} = 0.02$ (0.08); "No": Mean $BF_{21} = 0.03$ (0.10). In other words, the EEG amplitude of the "yes" trials were 50 times more likely, and that of the "no" trials were 33 times more likely, in the task-only models than in models including both LDF and task.

Moderation effect of the task.

Step 3 of the hierarchical model comparison dealt specifically with the interaction between task and the key variables. According to our hypotheses, task should interact with both EoS and LDF, though we do not have specific hypotheses concerning the timecourse or the scalp region. It should enhance the effect of EoS in the deep processing task specifically, but enhance the effect of LDF in the shallow processing task instead. To illustrate the differences between shallow and deep tasks, we further analysed the effect of EoS and LDF in separate models per task, which produced event-related regression coefficients (ERRC) which provided the direction and magnitude of effects (Hauk et al., 2009). For example, if the ERRC of EoS was 2.50, it meant that as EoS increased one unit, the EEG amplitude increased for 2.50 μV.

The interaction effect between EoS and Task was not found in the hypothesised area for "yes" response (i.e., the left parietal area; Figure 8a).

Evidence was only found for "no" response (i.e., when people rejected the metaphors), at the end of the epoch in E108 in the right temporal area (BF_{32} = 14.03; Figure 8b), which means that the EEG amplitude of E108 was explained by the model with the interaction between EoS and task 14.03 times better than the model without the interaction. This interaction was driven by a difference between tasks starting from 400ms after the adjective onset (Figure 8c). EoS had a sustaining negative effect in the shallow sensibility judgement task. In other words, at the later stage of processing for sensibility judgement, metaphors that were hard to simulate elicited a greater positivity than those that were easy to simulate; which effect was absent in the deep interpretation generation task (Step 3 model, EoS effect in deep task: b = 0.37; interaction term: b = -1.75). This difference between tasks culminated at the end of the epoch, thus showing the strong evidence for the interaction model at the final time slice.

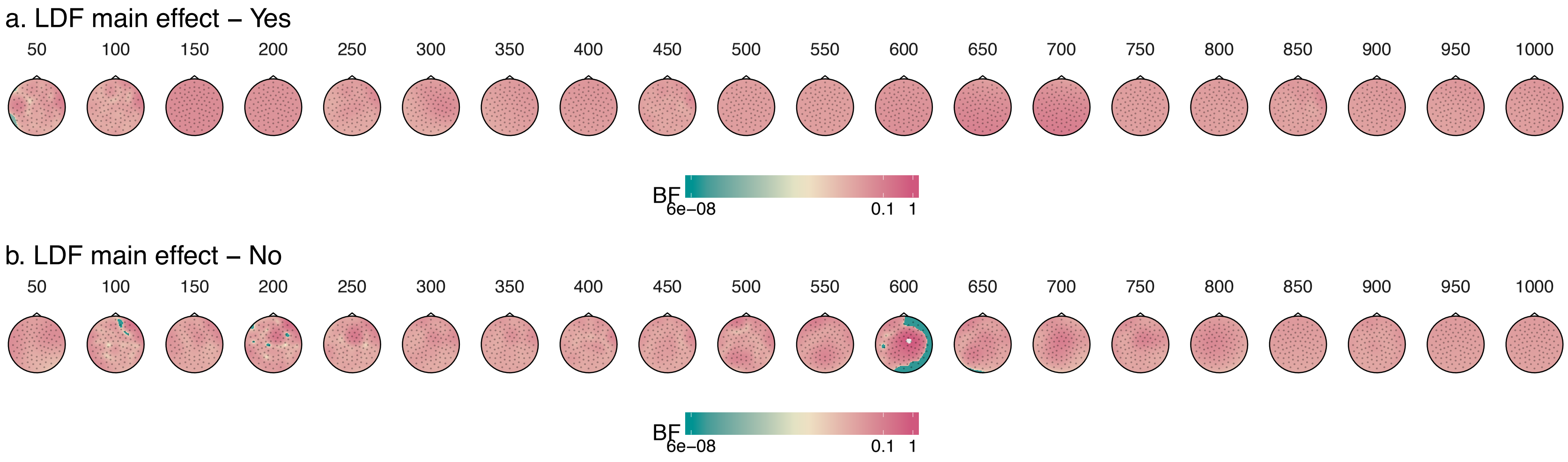


Figure 7. Evidence for main effects of LDF for a. "yes", and b. "no" responses measured by BF. When BF < 0.1, the data was strongly against the effect of LDF. BF was logarithmically transformed for plotting.

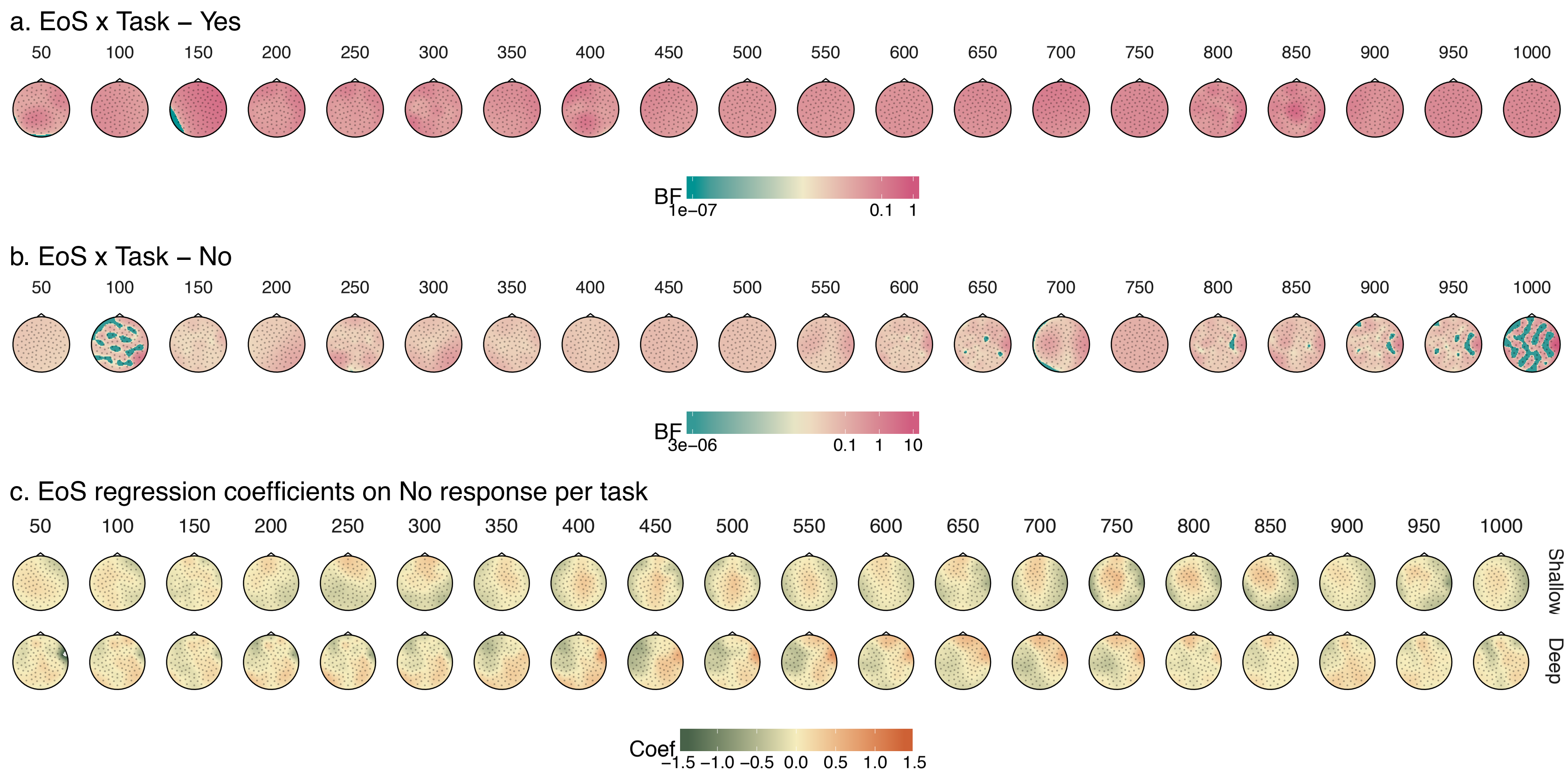


Figure 8. Evidence for the interaction between EoS and task for a. “yes”, and b. “no” responses measured by Bayes factors (Step 3 comparison). When $BF_{32} > 10$, the data strongly supported the interaction between EoS and task; whereas when BF_{32} < 0.1, the data strong was against the interaction between EoS task. BF was logarithmically transformed for plotting. Panel c shows the regression coefficients of EoS for “no” response in shallow vs deep processing tasks.

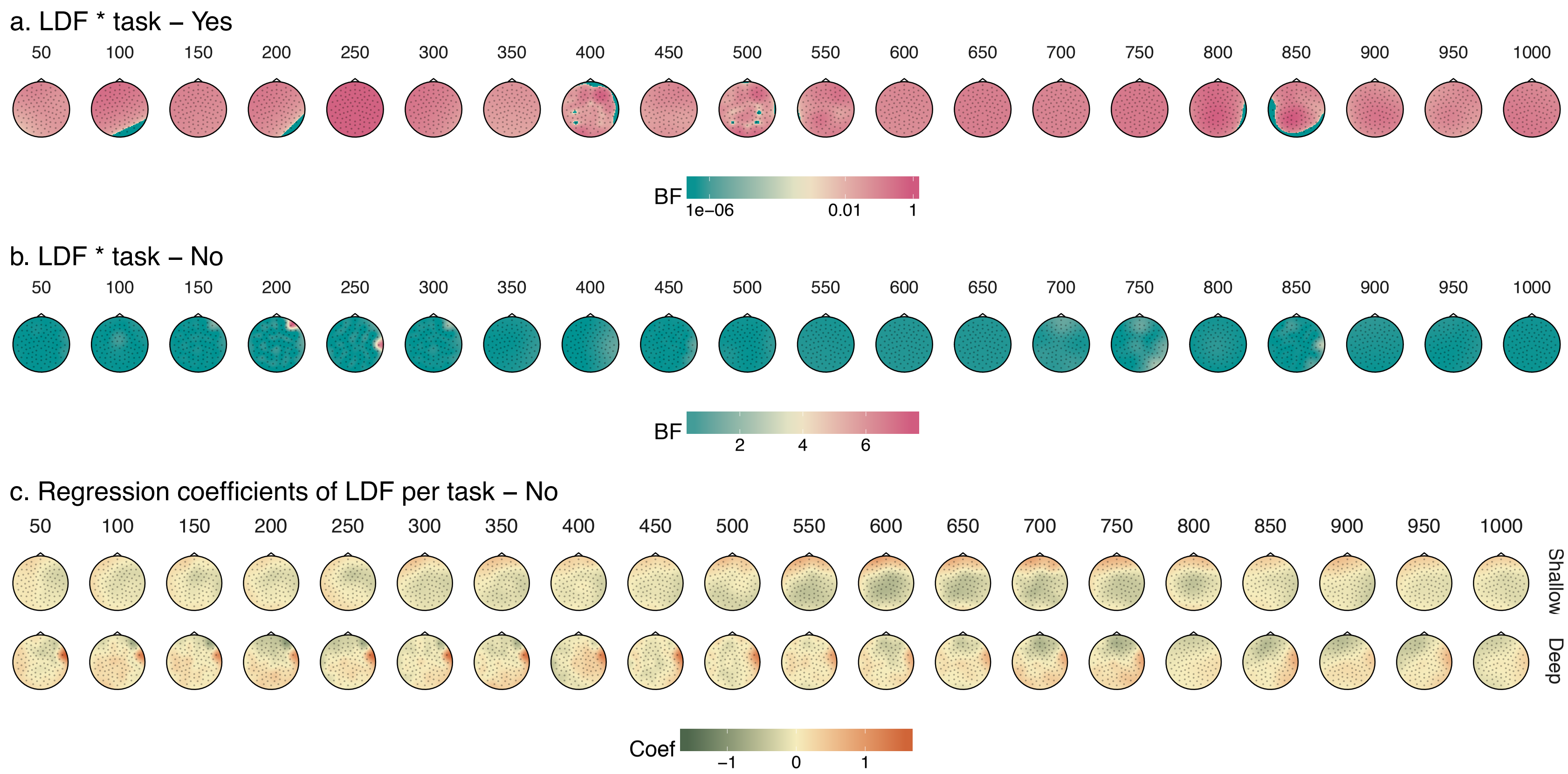


Figure 9. Evidence for the interaction between LDF and task for a. "yes", and b. "no" responses measured by Bayes factors (Step 3 comparison). When $BF_{32} > 10$, the data strongly supported the interaction between LDF and task; whereas when $BF_{32} < 0.1$, the data strong was against the interaction between LDF task. BF was logarithmically transformed for plotting. Panel c shows the regression coefficients of LDF for "no" response in shallow vs deep processing tasks.

Interaction between LDF and task was also only found for “no” response. At an early window, evidence was found around 200ms at the right frontal (E2: BF_{32} = 8.49; Figure 9b) and temporal electrodes (E108: BF_{32} = 7.79), which means that the data from these two electrodes were about 8 times more likely under models with the interaction between LDF and task than models without the interaction. From Figure 9c, it could be seen that the effect of LDF flipped directions for shallow and deep processing tasks. In E2, LDF had a negative effect on amplitude in the deep processing task (i.e., metaphors whose constituent words rarely co-occurred elicited a greater positivity; LDF in deep task: b = -1.33) which was smaller and in the opposite direction in the shallow processing task (interaction term: b = 1.44); and in E108, LDF had a positive effect in deep processing (LDF effect in deep task: b = 0.37; i.e., metaphors whose constituent words rarely co-occurred elicited a greater negativity) but a larger and opposite effect in shallow processing (interaction term: b = -0.77; low LDF elicited a greater positivity). This pattern sustained throughout the whole epoch as seen in Figure 9c, but was only supported by Bayes factors between 150ms and 250ms. This early effect in the right frontal region partially supported the hypothesis concerning early effect LDF, although it appeared earlier than the hypothesised time of 350ms.

Furthermore, we also found some support for the hypothesis that LDF should have a late effect in the right posterior region. At the 750ms slice in the right temporal and parietal regions, LDF interacted with task such that the effect of LDF flipped its direction. During deep processing, LDF had a negative effect at E90 in the right parietal region (LDF in deep task: b = 0.51), but a slightly larger, positive effect during shallow processing (interaction term: b = -1.01). Similar effect was found at E102 in the right temporal electrode (see Supplementary Materials). In other words, metaphors whose constituent words rarely co-occurred elicited a greater positivity in the shallow processing task, but a greater negativity in the deep processing task, with the effect on shallow task slightly larger.

Comparison between EoS and LDF.

Further model comparisons were carried out to extract information about the relative advantage of EoS and LDF. As we did not find widespread effects of LDF either in Step 2 or in Step 3, here we compared the LDF models obtained in Steps 2 and 3 with the EoS models obtained in Steps 2 and 3 respectively. Such information about the relative strength of EoS and LDF, though both weak at times, was useful for us to pinpoint the timecourse and topography of these variables to see when they differed most strongly.

Figure 10 demonstrated a clear pattern of differences between EoS and LDF. Consistent differences started to emerge after 350ms from the adjective onset. While EoS mainly accounted for the EEG amplitude in the central region, LDF accounted for the peripheral region. For “yes” response, the largest advantage of LDF over EoS was in the left occipital area (e.g., E71) after 600ms, similar to our hypothesis that LDF would have a greater effect at a later stage of processing in the posterior region. At 650ms, the model with LDF and task was 143 times better than the model with EoS and task at explaining the EEG data ($BF_{EL} = 0.007$), and the interaction model displayed the same advantage. LDF had a negative effect on the amplitude, which meant that metaphors with less frequently co-occurring words elicited greater positivity than metaphors with more frequently co-occurring words (Step 3 model, LDF in deep task: b = -0.35), and the effect was larger for shallow processing than deep processing, as predicted by the linguistic shortcut hypothesis. (Step 3 model, interaction term: b = -0.33).

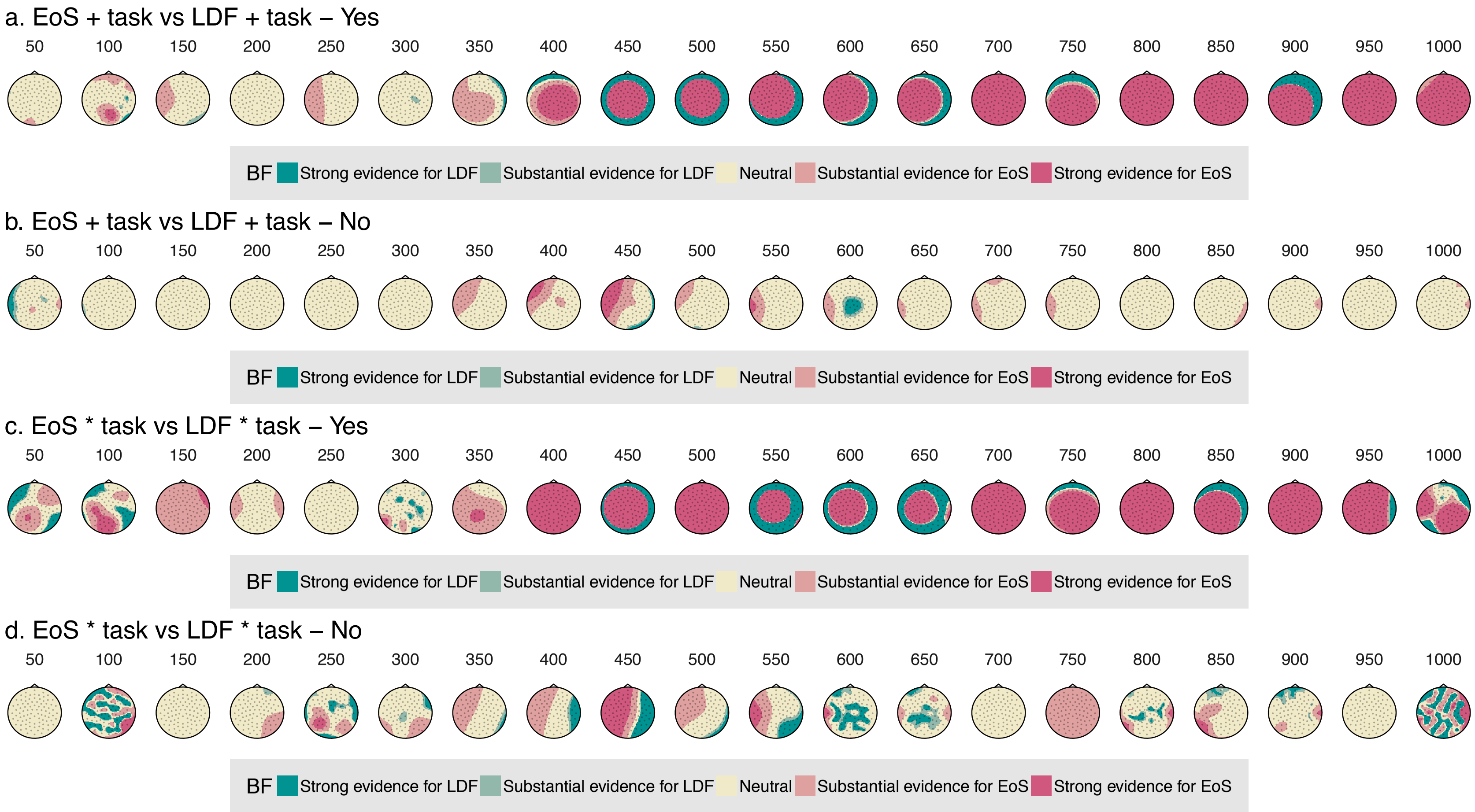


Figure 10. Evidence for EoS compared with evidence for LDF as a result of comparisons of the models obtained from Step 2 (i.e., EoS or LDF and task as fixed predictors; Panels a and b) and Step 3 (i.e., interactions between EoS or LDF and task as fixed predictors; Panels c and d). Strong evidence means that $BF_{EL} > 10$ or $BF_{EL} < 0.1$; substantial evidence means that $BF_{EL} > 3$ or $BF_{EL} < 0.33$.

The largest advantage of LDF over EoS for “no” response is in the right central area at 600ms (see Panels b and d). For example, the data in E106 was 250 times more likely in the model with LDF and task (without interaction) than the model with EoS and task (BF_{EL} = 0.004), and the same advantage was shown when comparing the interaction models too. LDF had, again, a negative effect on the amplitude (i.e, low LDF -> greater positivity; Step 3, LDF in deep task: b = -0.25), which was larger for shallow processing than deep processing too (Step 3 model, interaction term: b = -0.38).

Summary. The results confirmed that EoS had an effect on EEG amplitude across the centre of the scalp (especially the parietal area) around 400ms when people eventually accepted the metaphors. Although EoS alone did not affect EEG amplitude when people rejected the metaphors, it did interact with task at the end of the 1000ms epoch, suggesting differential activations of the embodied component per task. However, contrary to the hypothesis that EoS would be more involved in the deep interpretation generation task, this specific effect at the end of the epoch was in fact larger during the shallow sensibility judgement task.

The effect of LDF was only detected for “no” response, which was found to be moderated by task. This effect occurred around 200ms, prior to the effect of EoS, supporting the linguistic shortcut hypothesis. We found that the effect of LDF was in opposite directions for shallow and deep processing tasks, but the magnitude of the effect was not consistently larger during shallow processing. In other words, the evidence showed differential activations of the linguistic component per task, but did not consistently support the hypothesis that the linguistic component was more heavily utiliesd during shallow processing. Linguistic component had a larger effect during shallow than deep processing at a later stage of processing after 600ms from the adjective onset, which could be argued as support for the linguistic shortcut hypothesis.

General Discussion

The aim of the present study was to examine the nature and timecourse of conceptual representations during metaphor processing. From the behavioural and EEG study, we managed to see how the embodied and linguistic components interplayed to represent the meanings of a metaphor and inform the outcome of processing. Both components were activated within the initial 1000ms of processing, and lasted even until the response was made. Information from the linguistic distributional pattern was initially activated around 200ms after the adjective of a metaphor was presented, much earlier than when embodied simulation reached its peak. The effect on EEG approximated the typical P2 effect, usually associated with unexpected visual features of words in context (Federmeier, 2002). This early activation only appeared for metaphors that were eventually rejected as uninterpretable or nonsensical, highlighting the role of the linguistic component in deciding the eventual success/failure of processing. A metaphor that resulted in failed conceptual representation had already been marked out by the linguistic component from as early as 200ms, when what was generally regarded as semantic processing had not been initiated. It suggested that if the constituent words did not often co-occur, it violated the predictions created by the linguistic distributional patterns, which was informative to the eventual rejection of the metaphor.

After the initial activation of the linguistic components, embodied simulation commenced first for those metaphors which were eventually accepted. The temporal difference between embodied and linguistic components was consistent with LASS (Barsalou et al., 2008) and SIH (Louwerse, 2011), which suggested that the linguistic component should reach the peak of its activation before the embodied component. It also satisfied a necessary hypothesis of the linguistic shortcut hypothesis, that is, the linguistic component can be a quick and dirty guide for embodied component down the line (Connell & Lynott, 2014). The embodied component was prominent around and after 400ms, which was consistent with the typical N400 effect, commonly regarded as an indicator of the effort expended on integrating new stimulus into the semantic representation (Kutas & Federmeier, 2011). This finding offered further support to theories of embodied language comprehension, that is: the so-called semantic processing, a major part of language comprehension, depends on the reactivation of sensory, motor and other bodily experiences (Zwaan, 2004), which was also supported by the previous study on literal language processing (Liu et al., 2018a).

Judging from the neuroelectrical activations of EoS, if the embodied component was activated around 400ms, the metaphors would eventually be accepted. However, for those metaphors that ended up in rejection, the embodied component was not activated until the end of the 1000ms epoch, which suggested that the embodied component might have an effect on the processing of rejected metaphor beyond the initial 1000ms epoch, consistent with the behavioural finding the embodied component affected the rejection speed and rate too. The fact that this effect between 950ms and 1000ms was moderated by task was also consistent with the behavioural finding that people relied on simulation more during deep processing than shallow processing. It meant that simulation was not or all-or-nothing process, but was subject to the desired depth of processing.

On the other hand, the linguistic component affected the rejection of metaphors again around 750ms in the posterior region, as predicted by the tentative hypotheses based on results from literal langauge processing (Liu et al., 2018a). Furthermore, this effect was moderated by task too, lending support to the linguistic shortcut hypothesis. These activations, coinciding with the late positivity complex, were often associated with reconciling conflicting representations (Kuperberg, 2007; Paczynski & Kuperberg, 2012) as well as decision making (Collins et al., 2011;

Polich, 2007), which will be further discussed in the next section.

Linguistic Shortcut Hypothesis

The role of the linguistic component was of particular interest in this study, which aimed to test whether the linguistic component can act as a shortcut to guide future processing in the embodied component. This hypothesis was supported, in part, by both the ERP results and the behavioural results. The ERP study showed that the linguistic component was activated prior to the embodied component, satisfying the necessary condition of the linguistic shortcut hypothesis; it also showed that the linguistic component was activated, within the first 1000ms, for those metaphors that were eventually rejected, which supported the suggestion that one of the functions of the linguistic shortcut was to conserve limited energy if the processing is unlikely to succeed (Connell & Lynott, 2013). The behavioural study also showed that the linguistic component could affect the speed of accepting a metaphor directly. In other words, people could judge whether a metaphor was sensible/interpretable, based on the distributional patterns alone, even when the effect of embodied simulation was already accounted for.

However, our studies also showed evidence contrary to the linguistic shortcut hypothesis, specifically concerning the interaction between the linguistic component and task. According to the linguistic shortcut hypothesis, the linguistic component should have a greater effect during shallow processing; but the behavioural results showed that the lingusitic component had similar, or even larger, effect during deep processing compared to shallow processing. In other words, people did not prefer to rely on lingusitic distributional patterns more even when the task allowed for a superficial judgement. Similarly, the ERP results did not support the prediction either. Although different patterns were shown for shallow and deep processing (i.e., the regression coefficients of LDF at 200ms-250ms was larger for deep processing than shallow processing, but was larger for shallow than deep processing at 700ms-750ms), it did not necessarily mean that the linguistic component was more, or less, activated during deep processing because of the low spatial resolution of EEG. Since the waveform displayed the sum of all neural activities in the brain, which for certain included elements other than the activation of linguistic information (Luck, 2005), it was better concluded that the linguistic component was activated differently due to different depths of processing.

In summary, our studies did not fully support the linguistic shortcut hypothesis. Although the linguistic information was activated earlier than embodied simulation, it did not appear to be used heuristically (i.e., quickly in place of embodied simulation). In fact, the linguistic and embodied component appear to engage in a complex interplay to construct conceptual representations.

Interplay between Embodied and Linguistic Components

In this study, we found that the activations of embodied and linguistic components are flexible and dynamic, more complex than what was originally hypothesised. The two components were utilised differently based on the demand of the tasks and the outcome of processing as discussed before. Furthermore, unlike the schematic representation of embodied and linguistic components, as illustrated by Barsalou et al. (2008) for instance, embodied and linguistic information were activated at different stages of metaphor processing, possibly playing different roles.

The embodied information, on the one hand, was used for constructing semantic representation for successful processing, but it was also involved at a later stage and eventually affected the outcome of unsuccessful processing. It could be because embodied simulation took longer to initiate if the metaphors were difficult to simulate. Otherwise, it could also be because the embodied information was later involved in the decision making processing for those metaphors that were difficult to simulate. That is to say, people may not be able to construct a coherent representation, so they need to use the representation that is available to decide whether the sentence should be actually rejected as nonsensical/uninterpretable.

The linguistic information, on the other hand, was involved in decision making also in at least two stages. At the early stage, it influenced sub-lexical and lexical processing to signal which metaphors were likely to fail. At later stages, it interacted with the embodied component to decide which metaphors were to be rejected. Our findings in Experiment 1 replicated the findings by Liu et al. (2018a), which suggested that high distributional frequency (e.g., *Supply can be bright*) encouraged people to process the metaphor further even when the simulation was already difficult. Thus, it safeguarded against premature rejection of a metaphor. Although an opposite pattern was found in Experiment 2 (replicating Liu et al., 2017), that is low distributional frequency (e.g., *Solutions can be bright*) discouraged people to process a metaphor further with embodied simulation, it could be best explained by the fact that people had longer time to respond to the metaphors in Experiment 2, which had a 1s delay, than in Experiment 1.

In all, we painted a dynamic picture concerning how embodied and linguistic information is used to construct conceptual representations. Thanks to the high temporal resolution of EEG, we can see that both components were involved at different stages of processing, and their involvements were highly task dependent, being subject to the depth of processing required, the interaction between two components, and the design of the study. Indeed, we have suggested a process of conceptual representation that was more complex than theory such as LASS and SIH based largely on evidence from literal language processing. Future studies should consider if these complexities is could apply to the processing of other language types (e.g., abstract language).

References

Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia*, *49*(5), 970–983. doi:10.1016/j.neuropsychologia.2011.01.003

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. doi:10.1037/a0016261

Arzouan, Y., Goldstein, A., & Faust, M. (2007). Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Research*, *1160*, 69–81. doi:10.1016/j.brainres.2007.05.034

Auguie, B. (2017). *GridExtra: Miscellaneous functions for "grid" graphics.* Retrieved from https://CRAN.R-project.org/package=gridExtra

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/J.JML.2007.12.005

Bache, S. M., & Wickham, H. (2014). *Magrittr: A forward-pipe operator for r.* Retrieved from https://CRAN.R-project.org/package=magrittr

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). doi:10.1016/j.jml.2012.11.001

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In A. M. Glenberg & A. C. Graesser (Eds.), *Symbols, embodiment, and mean-*

ing (pp. 245–283). Oxford, England: Oxford University Press. Retrieved from http://psychology.emory.edu/cognition/barsalou/papers/Barsalou{_}et{_}al{_}chap{

Bates, D., & Maechler, M. (2017). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Brants, T., & Franz, A. (2006). Web 1T 5-gram Version 1 LDC2006T13. DVD. Philadelphia: Linguistic Data Consortium.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–43. doi:10.1016/j.brainres.2012.01.055

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. doi:10.1016/S0022-5371(73)80014-3

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, NJ: Erlbaum. Retrieved from https://books.google.co.uk/books?id=fAnSOgbdFXIC

Collins, J., Pecher, D., Zeelenberg, R., & Coulson, S. (2011). Modality Switching in a Property Verification Task: An ERP Study of What Happens When Candles Flicker after High Heels Click. *Frontiers in Psychology*, *2*(February), 10. doi:10.3389/fpsyg.2011.00010

Connell, L. (2018). What have labels ever done for us? The lingusitic shortcut in conceptual processing. *Language, Cognition and Neuroscience*.

doi:10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2013). Flexible and fast: linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, *20*(3), 542–50. doi:10.3758/s13423-012-0368-x

Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*(3), 390–406. doi:10.1111/tops.12097

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*(6), 958–968. doi:10.3758/BF03195780

Dantzig, S. van, Cowell, R. A., Zeelenberg, R., & Pecher, D. (2011). A sharp image or a sharp knife: norms for the modality-exclusivity of 774 concept-property items. *Behavior Research Methods*, *43*(1), 145–54. doi:10.3758/s13428-010-0038-8

Davenport, T., & Coulson, S. (2011). Predictability and novelty in literal language comprehension: an ERP study. *Brain Research*, *1418*, 70–82. doi:10.1016/j.brainres.2011.07.039

De Grauwe, S., Swain, A., & Holcomb, P. (2010). Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia*, *48*(7), 1965–1984. doi:10.1016/j.neuropsychologia.2010.03.017.Electrophysiological

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., … Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597–600. doi:10.1038/26967

Dowle, M., & Srinivasan, A. (2017). *Data.table: Extension of 'data.frame'.*

Retrieved from https://CRAN.R-project.org/package=data.table

Federmeier, K. D. (2002). Picture the difference: electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*, *40*(7), 730–747. doi:10.1016/S0028-3932(01)00193-2

Gelman, A., & Su, Y.-S. (2018). *Arm: Data analysis using regression and multilevel/hierarchical models*. Retrieved from https://CRAN.R-project.org/package=arm

Genz, A., & Bretz, F. (2009). *Computation of multivariate normal and t probabilities*. Heidelberg: Springer-Verlag.

Harrell Jr, F. E. (2018). *Rms: Regression modeling strategies*. Retrieved from https://CRAN.R-project.org/package=rms

Harrell Jr, F. E., Charles Dupont, & others. (2018). *Hmisc: Harrell miscellaneous*. Retrieved from https://CRAN.R-project.org/package=Hmisc

Hauk, O., Pulvermüller, F., Ford, M., Marslen-Wilson, W. D., & Davis, M. H. (2009). Can I have a quick word? Early electrophysiological manifestations of psycholinguistic processes revealed by event-related regression analysis of the EEG. *Biological Psychology*, *80*(1), 64–74. doi:10.1016/j.biopsycho.2008.04.015

Hendrix, P., Bolger, P., & Baayen, R. H. (2017). Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 128–149. doi:10.1037/a0040332

Henry, L., & Wickham, H. (2018). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr

Hothorn, T. (2017). *TH.data: TH's data archive*. Retrieved from https://CRAN.R-project.org/package=TH.data

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Højsgaard, S., & Halekoh, U. (2018). *DoBy: Groupwise statistics, lsmeans, linear contrasts, utilities*. Retrieved from https://CRAN.R-project.org/package=doBy

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes Factors. *The Journal of Problem Solving*, *7*, 2–9. doi:10.7771/1932-6246.1167

Jeffreys, H. (1998). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press. Retrieved from https://global.oup.com/academic/product/the-theory-of-probability-9780198503682?cc=gb{\&}lang=en{\&}

Johns, B. T., & Jones, M. N. (2012). Perceptual Inference Through Global Lexical Similarity. *Topics in Cognitive Science*, *4*(1), 103–120. doi:10.1111/j.1756-8765.2011.01176.x

Koenker, R., & Ng, P. (2017). *SparseM: Sparse linear algebra*. Retrieved from https://CRAN.R-project.org/package=SparseM

Kuhn, M., Steve Weston, Wing, J., Forester, J., & Thaler, T. (2016). *Contrast: A collection of contrast methods*. Retrieved from https://CRAN.R-project.org/package=contrast

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*(1), 23–49. doi:10.1016/j.brainres.2006.12.063

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual*

Review of Psychology, *62*, 621–47. doi:10.1146/annurev.psych.093008.131123

Kutas, M., Federmeier, K. D., & Sereno, M. I. (1999). Current approaches to mapping language in electromagnetic space. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 359–392). New York: Oxford University Press. doi:10.1093/acprof:oso/9780198507932.003.0012

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13

Lebois, L. A. M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are Automatic Conceptual Cores the Gold Standard of Semantic Processing? The Context-Dependence of Spatial Meaning in Grounded Congruency Effects. *Cognitive Science*, *39*(8), 1764–1801. doi:10.1111/cogs.12174

Liu, P. Q., Connell, L., & Lynott, D. (2017). Can illness be bright? Metaphor comprehension depends on linguistic and embodied factors. In G. Gunzelmann, A. Hows, T. Tenbrink, & E. J. Davelaar (Eds.), *39th annual conference of cognitive science society* (pp. 2604–2609). Austin, TX: Cognitive Science Society. Retrieved from https://mindmodeling.org/cogsci2017/papers/0494/paper0494.pdf

Liu, P. Q., Connell, L., & Lynott, D. (2018a). Continuous neural activities of modality switching costs. *Manuscript in Preparation*.

Liu, P. Q., Connell, L., & Lynott, D. (2018b). Ease-of-Simulation Norms for 452 Adjective Metaphors. *Manuscript Under Review*, 1–27.

Liu, P. Q., Connell, L., & Lynott, D. (2018c). Effect of time constraints on the grounded representation of metaphor processing. *Manuscript in Preparation*, 1–30.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*(2), 273–302. doi:10.1111/j.1756-8765.2010.01106.x

Louwerse, M. M., & Connell, L. (2011). A taste of words: linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, *35*(2), 381–98. doi:10.1111/j.1551-6709.2010.01157.x

Louwerse, M. M., & Hutchinson, S. (2012). Neurological evidence linguistic processes precede perceptual simulation in conceptual processing. *Frontiers in Psychology*, *3*(October), 385. doi:10.3389/fpsyg.2012.00385

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*(1), 96–104. doi:10.1016/j.cognition.2009.09.002

Luck, S. J. (2005). Ten Simple Rules for Deisgning ERP Experiments. In *Event-related potentials: A methods handbook* (pp. 17–32). doi:10.1371/journal.pcbi.0020012

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41*(2), 558–64. doi:10.3758/BRM.41.2.558

Lynott, D., & Connell, L. (2010). Embodied conceptual combination. *Frontiers in Psychology*, *1*(November), 212. doi:10.3389/fpsyg.2010.00212

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of bayes factors for common designs*. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Nowosad, J. (2017). *Rcartocolor: 'CARTOColors' palettes*. Retrieved from https://CRAN.R-project.org/package=rcartocolor

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, *67*(4), 426–448. doi:10.1016/j.jml.2012.07.003

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2017). *nlme: Linear and nonlinear mixed effects models.* Retrieved from https://CRAN.R-project.org/package=nlme

Plate, T., & Heiberger, R. (2016). *Abind: Combine multidimensional arrays.* Retrieved from https://CRAN.R-project.org/package=abind

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for mcmc. *R News*, *6*(1), 7–11. Retrieved from https://journal.r-project.org/archive/

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–48. doi:10.1016/j.clinph.2007.04.019

Pynte, J., Besson, M., Robichon, F. H., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain and Language*, *316*(55), 293–316. Retrieved from http://www.sciencedirect.com/science/article/pii/S0093934X96901078

R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality research.* Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Robinson, D. (2018). *Broom: Convert statistical analysis objects into tidy data* *frames*. Retrieved from https://CRAN.R-project.org/package=broom

Sarkar, D. (2008). *Lattice: Multivariate data visualization with r*. New York: Springer. Retrieved from http://lmdvr.r-forge.r-project.org

Solt, F., & Hu, Y. (2015). interplot: Plot the effects of variables in interaction terms. Available at The Comprehensive R Archive Network (CRAN). Retrieved from https://CRAN.R-project.org/package=interplot

Terry M. Therneau, & Patricia M. Grambsch. (2000). *Modeling survival data: Extending the Cox model*. New York: Springer.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth.). New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247. doi:10.1515/LANGCOG.2009.011

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. doi:10.3758/BF03194105

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*(1), 1–29. Retrieved from http://www.jstatsoft.org/v40/i01/

Wickham, H., & Henry, L. (2017). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *Dplyr: A grammar of*

CHAPTER 8

General Discussions

8.1 Aims of thesis

This thesis set out to view metaphor processing as an aspect of language processing in general. Since language processing relies on conceptual representations that involve both embodied simulation and distributional patterns of linguistic symbols, metaphor processing could be examined using the same combinatorial approach. These two components should affect the process and performance of metaphor comprehension in similar ways. Therefore, I used both behavioural and neurophysiological measures to paint a detailed picture of how the embodied and lingusitic components acted and interplayed during real-time metaphor processing. Research presented in this thesis was to achieve three overarching aims: 1. to find evidence for the independent contributions of embodied and linguistic components; 2. to study how conceptual representations are constructed in response to different task demands; and 3. to examine whether the linguistic component could act as a shortcut for conceptual representations.

To achieve the first aim, I generated a body of metaphoric sentences which varied on two independent dimensions (Chapter 3), which were the ease-of-simulation measure (EoS) and the linguistic distributional frequencies (LDF). These two measures could indeed account for a unique portion of metaphor processing, including the speed and the outcome of the processing, independent from one another. In Chapters 4, 5 and 7, I found that EoS prominently affected metaphor processing, such that a metaphor which was typically regarded as easy to simulate was more likely to be accepted and accepted more quickly, whereas a metaphor which was typically regarded as difficult to simulate was more likely to be rejected and rejected more quickly. Meanwhile, LDF contributed to a unique portion of metaphor processing independent from EoS, though the results were more mixed. In Chapters 4 and 7, LDF was found to affect the outcome of processing, such that the more often two constituent words co-occurred, the more likely the metaphor was to be accepted after deep processing. LDF was also found to affect the speed of accepting a metaphor (i.e., the more often two words co-occurred, the more quickly the metaphor was accepted as sensible or interpretable).

To study how task demands affect metaphor processing (i.e., the second aim), throughout the thesis I used two tasks which required different depth of processing, that is the sensibility judgement task which required relatively shallower processing, and the interpretation generation task which required relatively deeper processing. I found that the required depth of processing affected the reliance on the embodied and linguistic components. Consistently throughout the thesis (Chapters 4, 5 and 7), EoS was found to have a greater effect during the interpretation generation task than the sensibility judgement task. LDF was also found to have a greater effect on the acceptance rate in the interpretation generation task in Chapter 4, and a greater effect on the speed to accept a metaphor in Chapter 7. In other words, the required depth of processing affected the engagement of the embodied component consistently, but it affected the linguistic component only occasionally. In later sections, we will further discuss the fact that the effect of task demands on the linguistic component was the opposite of what we predicted based on the linguistic shortcut hypothesis.

To examine the linguistic shortcut hypothesis (the third aim), I studied metaphor processing under time constraints (Chapter 5) and using EEG (Chapter 7). The EEG study found that the linguistic component reached the peak of its activation before the embodied component. Starting from the stimulus onset, LDF accounted for the neuroelectrical activations as early as 200ms, and EoS accounted for the activations from around 400ms. As such, the linguistic component was capable of providing some information about whether the processing would be successful. However, the linguistic shortcut was not always taken during metaphor processing. People would not hastily judge a metaphor based on the co-occurrence frequency of the constituent words alone. Instead, people would allow the embodied component to process the metaphor further, and the linguistic component seemed to only affect the outcome of processing at a later stage. For instance, in the study with time constraint (Chapter 5), LDF only had an effect on the speed and outcome of processing when people had plenty of time to process a metaphor. In the EEG study (Chapter 7), it was also found that embodied and linguistic components both affected the neural activations after 600ms from the stimulus onset. In a word, while the linguistic component was activated prior to the embodied component, it was only operative at a later stage when semantic representations had already been constructed.

8.2 Evaluation of findings

In this thesis, I conducted three behavioural studies using similar paradigms. In Chapters 4, 5 and the behavioral part of Chapter 7, participants all read metaphoric sentences and performed either a sensibility judgement task or an interpretation generation task (or both in Chapter 7), and their response decisions and RT were recorded. Because some of the findings were mixed, especially findings with LDF, it was necessary to consolidate them. More importantly, each study only provided a piece of evidence while no one of them should be considered more definitive than others, so it is also necessary to evaluate the findings holistically. Therefore, I performed a mini meta-analysis using the *z*- or *t*-values from the mixed-effect models using the meta.ttestBF function in the {BayesFactor} package in R (Morey & Rouder, 2015). The function can combine the *z*- or *t*-values from multiple regression models for a single meta-analytic test and yield a Bayes factor for a selected interval (e.g., $t > 0$) against the null (i.e., true effect size was equal to 0). When BF_{10} is larger than 3, it suggests that we have evidence to support the alternative hypothesis.

8.2.1 Effects of interest

The findings that need consolidating are the effects of EoS, LDF, and the moderation effect of tasks. EoS and LDF had direct effects on response decisions and RT, that is the effect of one variable (e.g., EoS) when other variables were at the reference level (e.g., when LDF was at its mean, and when task was 0). Meanwhile, they also had indirect effects, which were their interaction with one another, and the interaction between task and either of these variables. Here I examined both the direct effects and the indirect effects, and the alternative hypotheses were as follows:

1. Concerning the direct effects: EoS should have a positive effect on response decision (i.e., higher EoS $\rightarrow$ more acceptance; $z > 0$), a negative effect on "yes" RT (i.e., higher EoS $\rightarrow$ faster acceptance $t > 0$), and a positive effect on "no" RT (i.e., higher EoS $\rightarrow$ slower rejection; $t < 0$); LDF should also have a positive effect on response decision, a negative effect on "yes" RT, and a positive effect on "no" RT.

2. Concerning the indirect effects: task should interact with EoS with a positive term on response decision and "no" RT, and a negative term on "yes" RT (deep processing $\rightarrow$ larger EoS effect); and task should interact with LDF in the opposite way according to the linguistic shortcut hypothesis (shallow processing $\rightarrow$ larger LDF effect). EoS and LDF should have an interaction. From the pattern of results in the individual studies, I expected the interaction to affect the speed of rejection in particular, such that as LDF increased, the effect of EoS decreased (i.e., $t < 0$).

8.2.2 Models evaluated

To evaluate effects on response decision, I extracted results from the logistic mixed-effect models from Chapter 4 (i.e., between participant study with no time constraint) and the behavioural part of Chapter 7 (i.e., within participant study with long time constraint). The data from Chapter 5 were re-analysed in order to collapse the data cross time constraints. In other words, the fixed predictor time constraints was removed so that the effects evaluated here applied to the whole dataset. For the direct effects (i.e., effects of EoS or LDF alone) and the interaction between EoS and LDF, I used the following models:

```
Response decisions ~ EoS + LDF + EoS x LDF
```

To evaluate the effects on RT, I extracted results from the linear mixed-effect models from Chapters 4 and 7 as well. The data from Chapter 5 were also re-analysed to remove the effect of time constraint. The models used to extract the effects of EoS and LDF and their interaction were those that took on the same parameters as the models for response decisions above.

For the moderation effect of task, only the models from Chapter 4 and 5 could be consolidated using BF, because they were both between-participant studies, while the Chapter 7 used a within-participant design. I used the following models:

```
Response decisions / RT ~ Task + EoS + LDF + Task x EoS + Task x LDF
+ Task x EoS x LDF
```

I extracted only the effect sizes of Task x EoS and Task x LDF for the meta-analysis.

8.2.3 Direct effects

EoS played a prominent and consistent role as predicted, that is the easier it was to simulate a concept, the more likely (Shallow: $BF_{10} = 4.02$ x 10^52; Deep: $BF_{10} = 2.38$ x 10^59) and speedily (Shallow: $BF_{10} = 1.40$ x 10^20; Deep: $BF_{10} =$ 2.02 x 10^26) it was to accept a metaphor; whereas the harder it was to simulate a concept, the more likely and speedily (Shallow: $BF_{10} = 1.31$ x 10^3; Deep: 2.15 x 10^7) it was to reject a metaphor. In short, the embodied component affected the speed and outcome of processing whether the processing was successful or not.

LDF on its own also had an effect on metaphor processing in some cases. The more often two constituent words co-occur in context, the more likely it was to accept a metaphor in both tasks (Shallow: $BF_{10} = 3.65$; Deep: $BF_{10} = 21.01$). Furthermore, LDF affected the speed to accept a metaphor during shallow processing ($BF_{10} = 26.56$), as well as during deep processing (3.33); but it did not affect the speed to reject a metaphor directly (Shallow: $BF_{10} = 0.35$; Deep: BF_{10} $= 1.54$).

8.2.4 Indirect effects

8.2.4.1 Interaction between EoS and LDF

The meta-analysis showed evidence for the interaction between EoS and LDF on the speed to reject a metaphor during deep interpretation generation

(BF_{10} = 4.48), with a negative interaction term. That is to say, although EoS had a positive effect on the speed of rejection when LDF was at its mean (i.e., high EoS $\rightarrow$ slow to reject), the effect decreased and became less positive as LDF increased. In other words, when the constituent words rarely co-occurred, people were very fast at deciding that they could not interpret a metaphor that was typically regarded as hard to simulate (e.g., *Illness can be bright*) and slow at deciding that they could not interpret a metaphor that was typically regarded as easy to simulate (e.g., *Solutions can be bright*). However, when the constituent words often co-occurred, people hesitated (i.e., were slower) when they needed to decide that they could not interpret a metaphor that was hard to simulate (e.g., *Supply can be bright*). We found a similar pattern of rejection speed in the shallow sensibility judgement task of Chapters 5 and 7, but that effect was not supported by strong evidence (BF_{10} = 1.65). Apart from this interaction on rejection speed, the meta-analyses showed that there were little or no evidence in favour of the interaction between EoS and LDF on response decisions and acceptance speed, compared with the null hypothesis that the true effect sizes of the interaction terms were 0 (Shallow response decision: BF_{10} = 2.44; Deep response decision: BF_{10} = 0.32; Shallow "yes" RT: BF_{10} = 7.43 x 10^-2; Deep "yes" RT: BF_{10} = 6.67 x 10^-2).

8.2.5 Moderation effects of tasks

The moderation effects of tasks were extracted from the two between-participant studies only (i.e., Chapter 4 and 5). There was strong evidence that tasks moderated the effect of EoS on the speed of processing in the direction predicted. That is to say, EoS had a larger, more negative effect on "yes" RT (BF_{10} = 2.85 x 10ˆ15), and a larger, more positive effect on "no" RT (BF_{10} = 2.89 x 10ˆ3). These effects were consistent with the results from the behavioral study of Chapter 7, which confirmed that people relied on the embodied component more during deeper processing than shallower processing, though the effect of EoS on response decisions was not moderated by task ($BF_{10} = 0.82$).

In contrast, there was strong evidence that task did not moderate the effect of LDF in the way we predicted. LDF did not have a larger effect during shallow processing on either response decision (BF_{10} = 5.21 x 10ˆ-2), or RT ("Yes": BF_{10} = 5.65 x 10ˆ-2; "No": $BF_{10} = 0.41$). In fact, the evidence suggested that the true effect of the interaction was more likely to be 0.

8.3 Contributions

This thesis took a novel approach to study metaphor processing as involving simulation-linguistic based conceptual representation. The results showed a dynamic, complex processing model that previous theories have not envisioned (e.g., Barsalou et al., 2008). First of all, the involvement of simulation and linguistic information was highly task-dependent. Depth of processing, length of response interval and task demands (e.g., EEG or behavioural experiment) all seemed to influence the conceptual representation constructed. I have mainly studied the effects of depth of processing and length of response interval and found that deeper processing and longer response intervals encouraged embodied simulation, whereas contrary to the lingusitic shortcut hypothesis shallower processing did not encourage the use of linguistic information and shorter response intervals eliminated its use altogether.

Furthermore, by combining EEG with behavioural methods, I achieved a detailed picture of the timecourse of activations of embodied and linguistic components. These two components were activated during various stages of processing and thus appeared to perform different roles. Lingusitic information was activated prior to simulation and was involved in lexical or sublexical processing. Afterwards, embodied simulation was used for semantic representation, which was its primary function. Later, both components seemed to be activated again before the respose was made. Thus, it was likely that both were used for decision making. The linguistic component in particular was suggested to be responsible for the integration of different representations.

Finally, simulation and linguistic information interacted. When people engaged in deep processing, they used high distributional frequency as an indicator that the metaphor was more likely to be interpretable. Thus, they spent longer processing the metaphor even when the simulation was difficult. In other words, the linguistic information was used to safeguard against premature rejection of metaphor when the processing time allowed. This finding was different from the hypothesis of the cognitive triage mechanism (Connell & Lynott, 2013) which suggested that the linguistic information was used to quickly signal which concepts were more likely to be non-sensical and should **not** be processed further. The safeguarding mechanism suggested by this thesis is more conservative than the triage mechanism in that people bias towards expending more effort to process a metaphor using the safeguarding method, whereas they would bias towards conserving effort and reject a metaphor using the triage mechanism.

8.3.1 Comparison with traditional approaches of metaphor comprehension

This combinatorial approach stood distinctly from the traditional disciplines of metaphor comprehension, such as the psycholinguistic tradition and the embodied metaphor theory, because it yielded results that could not be accounted for by any of these approaches.

1. The findings in this thesis showed that the combinatorial approach was better than the traditional psycholinguistic approach which studies metaphor comprehension based on individual factors such as salience, conventionality and aptness. In Chapter 1, I have argued that those factors suffer from problems with confounded definitions and inter-correlations. Salience and conventionality have both been construed in two ways, either as the familiarity to an expression (e.g., Blasko & Connine, 1993; Bowdle & Gentner, 2005; Glucksberg, 2003), or as the familiarity to a concept (e.g., Campbell & Raney, 2015; Cardillo et al., 2010; Gentner & Wolff, 1997; Giora, 1997). LDF, a minimalist measure of the linguistic distributional pattern, was better than them because it focused on the exposure to linguistic expressions alone. As a solely linguistic variable, LDF avoided the confound of conceptual familiarity; but it was still found to exert independent effects on metaphor processing, which could not be explained by the psycholinguistic factors if they referred to conceptual familiarity.

EoS was better than aptness as well for two reasons. Firstly, the definition of EoS was clear and focused, that is: it only measured the effort it took to construct a coherent conceptual representation (Chapter 3), unlike aptness which could measure either the appropriateness of an expression (i.e., how well *bright student* can convey the speaker's intention; Blasko & Connine, 1993; Campbell & Raney, 2015; Katz et al., 1988) or how well the source-domain property (e.g., "clever") fitted the target word (e.g., *student*; Chiappe & Kennedy, 1999; L. L. Jones & Estes, 2005; Roncero & Almeida, 2014b). Secondly, EoS varied independently from LDF, unlike aptness which strongly correlated with conventionality, familiarity and frequency (Roncero & Almeida, 2014a; Thibodeau & Durgin, 2011).

More importantly, the combinatorial approach has yielded results that cannot be explained by the psycholinguistic theories (e.g., graded salience hypothesis, career of metaphor, etc.). The debate between career of metaphor and the theory of class inclusion has centred on the question, that is: which factor, either conventionality or aptness, can account for metaphor comprehension. Instead, my research could be seen a synthesis of these two theories, and showed that both the lower-level linguistic variable (i.e., LDF) and the higher-level conceptual variable (i.e., EoS) played unique and independent roles. Furthermore, the involvement of these two variables depended on the context and task requirements, which has never been shown by the psycholinguistic studies.

2. My research went beyond recent research on the embodied metaphor theory. In this thesis, I proposed a way to operationalise the effort of embodied

simulation which was agnostic to the content of simulation during metaphor processing (Chapter 3). That is to say, I did not prescribe what sensory, motor or affect experiences in specific were simulated when people read a metaphor such as *Students can be bright*. It was very different from the embodied metaphor theory which insisted that the embodied simulation of the source term (e.g., visual experiences of something bright) was crucial for metaphor comprehension. In Chapters 1 and 3, I argued that such a view of embodied simulation was not consistent with current imaging evidence. Several fMRI studies found that, across a large temporal span (several hundred milliseconds), the sensory and motor cortices responsible to processing the source-domain property (e.g., visual cortex activated when processing the word *bright*) were not selectively activated during metaphor processing (e.g., *Students can be bright*) compared to its literal counterpart (e.g., *Students can be clever*; Citron & Goldberg, 2014; Desai et al., 2013; Romero Lauro et al., 2013; Yang & Shu, 2016). In fact, the activation of source-domain properties became less and less prominent as a metaphor became conventionalised (Cardillo et al., 2012). Even though Boulenger et al. (2012) found that source-domain properties activated the sensorimotor cortices as early as 150ms, such activations would be quickly suppressed (Desai et al., 2013; Mashal, 2013; Raposo et al., 2009). As a result, I argued that dwelling on the source-domain activations were not helpful for the understanding of conceptual representations during metaphor processing.

Although I did not make a claim about what is simulated during metaphor processing, the findings of this thesis still supported the proposal that embodied simulation was heavily used. The way EoS was operationalised was validated in my thesis in two ways. First, EoS was a combined measure of three linguistic tasks (Chapter 2), and it was found to outperform any of the individual ratings in accounting for the speed of sensibility judgement. Therefore, I have argued that EoS measured the underlying mechanism of language comprehension, which is simulation (Zwaan, 2004). Second, from the EEG studies in Chapters 6 and 7, it could be seen that EoS displayed a similar neuroelectrical activations as the perceptual modality switching costs, a staple evidence for embodied simulation. Metaphors that were typically regarded as difficult to simulate (i.e., low EoS; e.g., *Supply can be bright*) elicited a negativity in the centro-parietal region around 400ms in the same way a switch between perceptual modalities did (e.g., *BLEMDER - loud* $\rightarrow$ *SUN - bright*). It further suggested that EoS measured the difficulty for simulating bodily experiences during metaphor processing in the same way that a perceptual switch causes difficulties in simulating sensory experiences. Hence, the conceptual representation during metaphor processing indeed depends on embodied simulation.

In a word, the content of simulation is more than the bodily experiences of source-properties. For example, *bright student* does not necessarily activates the sensory experiences related with something bright (e.g., the sun, a light bulb), but the simulation of its meaning could still activate some sensory information (e.g., visual appearance of young adult), motor information (e.g., asking questions, debating), and even affective information (e.g., a teacher may be positively valenced towards a clever student; Citron &

Goldberg, 2014), as well as simulation of broader context (e.g., classroom or seminar discussion). Research on simulation like this are still elementary, so I will discuss future directions of embodied metaphor processing in the next section.

8.4 Limitations and future directions

From the work presented in this thesis, I have proposed a processing model of metaphor processing based on the combinatorial simulation-linguistic perspective of conceptual representations. While my model confirmed the general principles of the approach, such as the existence of embodied and linguistic components, and the linguistic shortcut hypothesis, it differed from previous suggestions regarding the roles of and the interplay between the two components. In order to support my model of conceptual representations, there remain some crucial questions to be answered.

8.4.1 Are metaphors unique?

One question could be raised, regarding whether the current model can be generalised to the processing of all language types. Specifically, my thesis suggests that the linguistic component affected metaphor processing at a later stage (instead of before the peak of the embodied component), and it functioned not only as a shortcut, but also for decision making during shallow processing to reconcile different conceptual representations. An assumption for the present study was that metaphor processing should involve the components as literal language processing. It is of critical importance to assess whether such assumption holds true regarding the role of the linguistic component. It is conceivable that metaphors are processed in a different way from literal language. Thibodeau et al. (2017) showed that ratings on metaphors varied along a dimension called *figurativeness*, which accounted for ratings of how metaphoric, surprising and, to a small degree, familiar a metaphor was. It is possible that people can identify a sentence as metaphor, and then apply a different processing strategy. For instance, when they see a metaphor with low linguistic distributional frequency, they will not reject the metaphor straight away because it is common for metaphors be composed of words that do not often appear together (e.g., *lukewarm supporters*). Meanwhile, since it is also common for metaphors to have ambiguous meanings, people may be willing to accept a metaphor when the simulation is still at a relatively incomplete state, based on the joined results of linguistic and embodied components, especially during shallow processing.

If future research would be interested this question of whether the conceptual representation of metaphor processing is different from that of literal language processing, the top priority should be to come up with a better definition of *figurativeness*. At the moment, its definition is post hoc and circular. *Figurativeness* is a label put on the compound measure of metaphoricity, surprisingness and familiarity after the ratings have been collected. A sentence is regarded as figurative or metaphoric if it is less literal. Thus, we do not know how people are able to identify a sentence as a metaphor. The standard pragmatic approach used to suggest that a sentence could be identified as a metaphor if its literal meaning does not make sense (Grice, 1975), but later findings suggest that people responded to metaphors, at least familiar ones, as quickly as literal sentences (Gibbs, 1980; Glucksberg et al., 1982); and people could interpret a sentence as a metaphor even if the literal meaning is true (e.g., *Life is not a cruise*). Later with the pragmatic approach, the relevance theory suggests that a sentence is interpreted as a metaphor when the interpretation maximises its coherence with the linguistic and social context (Tendahl & Gibbs, 2008; D. Wilson & Sperber, 2004). On the other hand, taking a semantic approach, Winter (2016) recently suggested that a sentence is more figurative when the embodied simulation of source and target concepts have a greater mismatch. For example, the word *tie* has strong visual and haptic experiences but weak auditory, gustatory and olfactory experiences whereas the word *river* has strong visual and auditory experiences but weak experiences in the other senses, so *loud tie* may be rated as more metaphorical than *loud river*. Future research may follow either of these approaches for a better understanding of figurativeness.

8.4.2 What constitutes the embodied component?

As discussed before, this present thesis did not define simulation as the activation of source-domain properties and was agnostic concerning what was simulated during metaphor processing. As a result, EoS is an indirect measure (i.e., the end-product of successful language comprehension). Although we have demonstrated that EoS was a valid measurement of the effort to construct conceptual representations, these arguments may not satisfy some hardcore (or naive) embodiists, who would like to see metaphor processing fully grounded in bodily experiences. Meanwhile, it allows the embodiment skepticists to argue that this thesis does not support embodiment during metaphor processing at all, but that the effects of EoS could be fully reduced to amodal, symbolic language processing. Indeed, EoS was operationalised using three language processing tasks; sensibility and usability especially focuses on how the metaphors could be used in language. Without showing the specific sensorimotor experiences that are activated, the thesis did not provide direct evidence of embodied simulation of metaphoric meanings (see arguments in Casasanto & Gijssels, 2015; Leshinskaya & Caramazza, 2016; Mahon, 2015). To these criticisms, I do not have a ready fix in this thesis. Research on grounded conceptual representation needs to make substantial advance both theoretically and practically before we understand the content of simulation during metaphor processing.

On the theoretical side, we need to understand the status of embodied simulation in the grand scheme of what is called language processing. Although it has been demonstrated that language processing often recruits the distributed neural network responsible for sensory perception, motion or other bodily experiences, we cannot equate embodied simulation to language comprehension just yet. It is necessary to expand our scope of interest from literal, concrete words to other types of language, such as abstract concepts. Although abstract concepts are often defined as concepts represented by the linguistic system (Paivio, 1986), Barsalou et al. (2008) suggested that abstract concepts also activated experiences with mental states, events and situations under deep processing, in a similar way as concrete concepts (see also McRae, Nedjadrasul, Pau, Lo, & King, 2018; Vigliocco et al., 2014). Many have shown that the involvement of embodied simulation is task- and context-dependent (Lebois et al., 2015; Zwaan, 2014). Therefore, when studying the role of embodied simulation in language comprehension, it is crucial to study the conditions where embodied simulation is used. In addition, to understand what experiences in specific are simulated during metaphor processing, it again becomes necessary to study how metaphors are identified and how metaphoric meanings are generated. Both the relevance theory suggested above and the theory of structure mapping (see discussion about the career of metaphor in Chapter 1; Wolff & Gentner, 2011) could be useful to future research.

On the practical side, based on better understanding of the mechanism of simulation, better operationalisation of the embodied component should be devised. The relevance theory and structure alignment propose specific aspects of experiences that are simulated (i.e., contextual information and conceptual structures), which could be tested with property norms and priming studies. As for EoS proposed in the present thesis, it is worthy of further validation to see whether the same method could capture the ease of simulation in other types of language processing. If embodied simulation during literal language processing can be operationalised in the same way, we will have a stronger case to argue that the ease of simulation measure is valid for metaphors as well. Furthermore, imaging studies with better spatial resolution (e.g., TMS and MEG) could be used to validate the proposals above that sensorimotor and affective experiences related to the metaphoric meanings are activated.

8.4.3 Is LDF powerful enough to capture the linguistic component?

In my thesis, the role of LDF was found to be limited and to a much smaller scale than that of EoS. Apart from the possibility that metaphors are unique or that EoS could be explained by symbolic language processing as well, an alternative interpretation is that LDF is not a powerful enough definition to capture all there is to the linguistic component. Indeed, from the onset of this research, LDF was designed to be a minimalist rendition of the linguistic component. The rationale behind this decision was that if we were able to find evidence for the linguistic component with a simplistic measure (which we did), we could argue strongly that the linguistic component plays an important role, which was in fact the case for my thesis. The downside of this decision is that LDF may fail to detect instances where the linguistic component informs conceptual processing, and hence we are not able to draw a conclusion from the null effects.

The linguistic distributional patterns could be operationalised in many other ways. Some take into account a much larger window than the present LDF measure. For instance, latent semantic analysis (LSA) defines a word by distributional patterns within the context of a set of documents, and as a vector-based model it measures how often words occurred in similar context (Landauer & Dumais, 1997). In LSA, the meaning of *bright* is defined not only by the context in which it has a literal meaning, such as *sun*, *shining*, *lightbulb*; but also by the context in which it has a metaphoric meaning such as *student*, *clever*, *question*, *answer*, etc. As such, although *solution* and *bright* do not co-occur frequently in the same 5-word window, thus having a low n-gram frequency, they often co-occur with a similar set of words such as *students*, *answers*, etc. Therefore, the vector-based distributional models with a wider window might be able to capture the distributional pattern of *bright* and *solutions*, which could not be captured by the n-gram frequency. Apart from natural language processing techniques like LSA, there are also other distributional semantic models which take in small windows similar to the co-occurrence LDF in this thesis, such as probability models and predict models. Those models were also found to have outperformed n-gram frequencies, and the best model often depends on the task (Wingfield & Connell, 2018). Therefore, future research should take on more powerful operational definition, and also to be mindful about the fact that the use of the linguistic component is also flexible and task-dependent.

8.4.4 What is the best way to analyse ERP?

One of the missions of the thesis is to improve the method to analyse EEG data. The outcome method adopted by the present thesis was linear mixed-effect models to analyse the EEG data one electrode, one time slice at a time, and to use Bayes factor to solve the problems with null-hypothesis significance testing. This method is an improvement over traditional ANOVA; and treating EEG data in uniformed time slices is an especially important move towards component-free analysis, because the insistence on finding "components" has led to a proliferation of labels (e.g., early N400 vs N400, P600 vs LPC, etc.) and overstatements about the theoretical significance of raw waveform. However, even with linear mixed-effect models, a vast amount of data was still lost through averaging, and the continuity of data (i.e., similar data between electrodes and time slices) was still unaccounted for (Chapter 2) due to the lack of time and computing capacity practically. As such, the full potential of EEG's high temporal resolution is still not realised. To account for this systematicity across time and space, future research may use generalised additive models (GAMs) to create non-linear functions

(Hendrix et al., 2017).

Furthermore, to understand the neural substrates of metaphor processing, future research is in need of the combination multiple imaging approaches. While EEG can offer a high temporal resolution, thus fitting for the need of understanding the timecourse of conceptual representations, it does not inform us anything about the neural generator. To study how the conceptual representations are substrated in the brain, we are in need of both temporal and spatial precision such as source localisation with EEG or MEG. For the embodied component, high temporal and spatial resolutions could elucidate the relationship between source domain properties (i.e., the sensorimotor properties of the literal meaning) and the metaphoric properties. For the linguistic component, the combination of imaging approaches may shed light on its multiple roles (early semantic processing, decision making, etc.). A detailed picture of the interaction between neural clusters will improve our understanding of metaphor processing tremendously, and can help explain how the brain construct a concept grounded in a network of multi-modal experiences.

8.5 Concluding remarks

This thesis explored metaphor processing from the perspective of simulation-linguistic conceptual representations. It revealed a dynamic and flexible processing model, in which simulation and linguistic information interplayed subject to the demands of the tasks. It enriched current understandings of conceptual representations in general, pointing out that both embodied and linguistic components were activated at several stages of processing. Therefore, the depth of processing, the length of response intervals and design of tasks could all alter the reliance on the two components. Future studies on metaphor processing and language processing in general will greatly benefit if they could consider conceptual representation as such a dynamic process.

From the conception of the experiments to the producing of this thesis, I made it the top priority to enhance objectivity and replicability of my research. It is particularly shown by the EEG studies, for which I used a novel method of analysis that minimised researcher degrees of freedom. Meanwhile, the behavioural studies were pre-registered wherever possible, and I used the "21-word solution" for all the empirical papers (J. P. Simmons, Nelson, & Simonsohn, 2012). Furthermore, the entirety of thesis, including data analyses, manuscripts and all other sections (e.g., table of contents, literature reviews, etc.), was produced using R (see a list of R packages at the end of the thesis; scripts are available at https://goo.gl/SPvQgs).

List of Figures

List of Tables

List if R Packages

R packages used in this thesis include: R (Version 3.5.0; R Core Team, 2017) and the R-packages *abind* (Version 1.4.5; Plate & Heiberger, 2016), *arm* (Version 1.10.1; Gelman & Su, 2016), *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2015), *bindrcpp* (Version 0.2.2; Müller, 2017), *bookdown* (Version 0.7; Xie, 2016), *broom* (Version 0.4.4; Robinson, 2017), *citr* (Version 0.2.0; Aust, 2016), *coda* (Version 0.19.1; Plummer et al., 2006), *contrast* (Version 0.21; Kuhn et al., 2016), *cowplot* (Version 0.9.2; Wilke, 2017), *data.table* (Version 1.11.4; Dowle & Srinivasan, 2017), *doBy* (Version 4.6.1; Højsgaard & Halekoh, 2016), *dplyr* (Version 0.7.5; Wickham et al., 2017), *Formula* (Version 1.2.3; Zeileis & Croissant, 2010), *ggplot2* (Version 2.2.1; Wickham, 2009), *gridExtra* (Version 2.3; Auguie, 2017), *Hmisc* (Version 4.1.1; Harrell Jr et al., 2018), *interplot* (Version 0.1.5; Solt & Hu, 2015), *knitr* (Version 1.20; Xie, 2015), *lattice* (Version 0.20.35; Sarkar, 2008), *lme4* (Version 1.1.17; D. Bates et al., 2015), *lmerTest* (Version 3.0.1; Kuznetsova et al., 2017), *magrittr* (Version 1.5; Bache & Wickham, 2014), *MASS* (Version 7.3.50; Venables & Ripley, 2002), *Matrix* (Version 1.2.14; D. Bates & Maechler, 2017), *mgcv* (S. N. Wood, 2003, 2004, Version 1.8.23; 2011; S. Wood et al., 2016), *multcomp* (Version 1.4.8; Hothorn et al., 2008), *MuMIn* (Version 1.40.4; Bartoń, 2017), *mvtnorm* (Version 1.0.8; Genz & Bretz, 2009), *nlme* (Version 3.1.137; Pinheiro et al., 2017), *papaja* (Version 0.1.0.9735; Aust & Barth, 2017), *plyr* (Wickham, 2011; Version 1.8.4; Wickham et al., 2017), *psych* (Version 1.8.4; Revelle, 2017), *purrr* (Version 0.2.5; Henry & Wickham, 2017), *rcartocolor* (Version 0.0.22; Nowosad, 2017), *rms* (Version 5.1.2; Harrell Jr, 2018), *SparseM* (Version 1.77; Koenker & Ng, 2017), *survival* (Version 2.42.3; Terry M. Therneau & Patricia M. Grambsch, 2000), *TH.data* (Version 1.0.8; Hothorn, 2017), and *tidyr* (Version 0.8.1; Wickham & Henry, 2017)

Appendices

Appendix A

.1 Instructions of the metaphor norming tasks

.1.1 Sensibility:

You will read a list of sentences about everyday items (e.g. Socks are elastic). Your task is to rate **how much sense the sentence makes as if you had read it in a story or heard it in conversation**, on a scale of 1 – 7 where **1 = no sense at all** and **7 = perfect sense**. Take your time, and please do as many as you can even if you cannot finish them all.

.1.2 Usability:

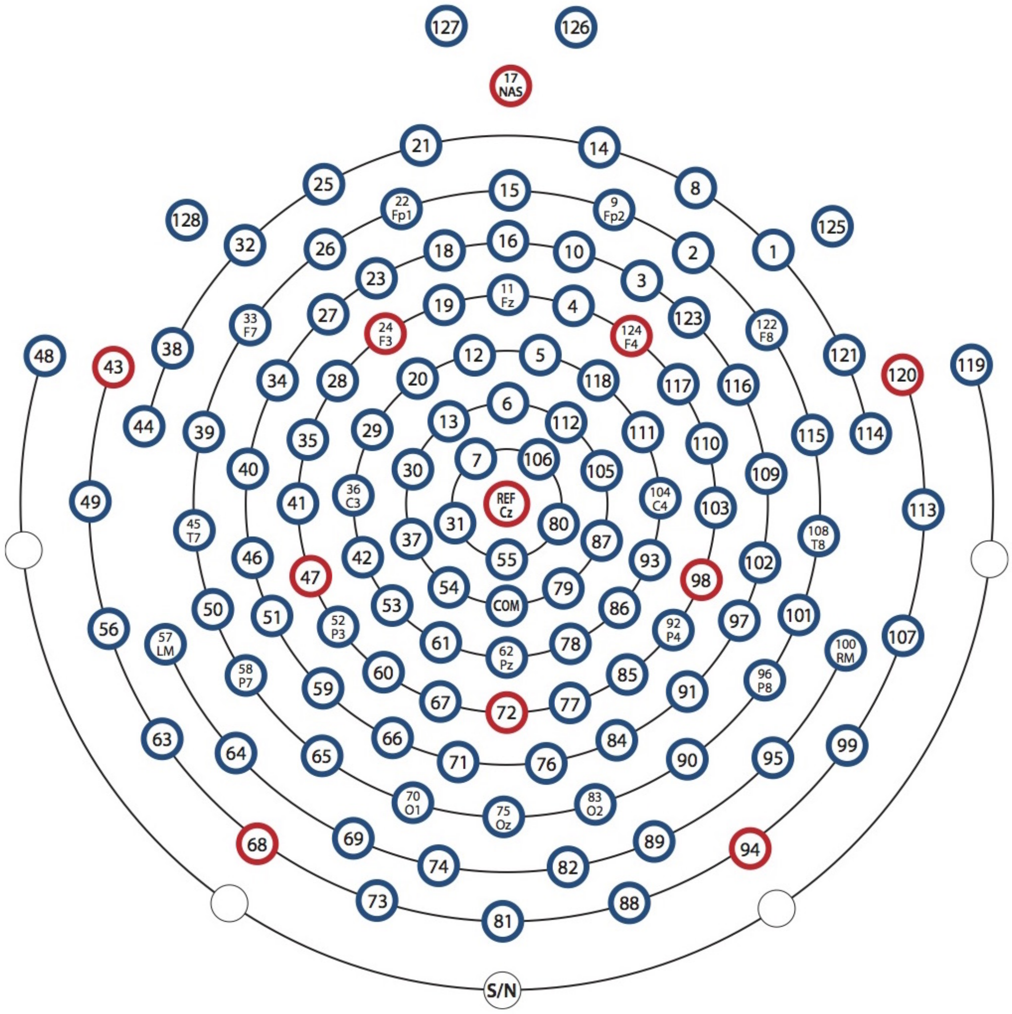
You will read a list of sentences about everyday items (e.g. Socks are elastic). Your task is to rate **how easy it would be for you to use the sentence in writing or in conversation**, on a scale of 1 – 7 where **1 = impossible to use** and **7 = very easy to use**. Take your time, and please do as many as you can even if you cannot finish them all.

.1.3 Imaginability:

You will read a list of sentences about everyday items (e.g. Socks are elastic). Your task is to rate **how easy it is for you to imagine the concept described in the sentence** (e.g., how easily you can imagine that socks might be elastic), on a scale of 1 – 7 where **1 = impossible to imagine** and **7 = very easy to imagine**. Take your time, and please do as many as you can even if you cannot finish all.

Appendix B

.2 128-channel EEG net layout



Consolidated References

Allport, D. A. (1985). Distributed memory, modular subsystems and dysphasia. In S. Newman & R. Epstein (Eds.), *Current perspectives in dysphasia* (pp. 207–244). Edinburgh: Churchill Livingstone. Retrieved from https://www.scribd.com/document/17676263/Allport-1985-Distributed-Memory-Modular-Subsystems-and-Dysphasia http://scholar.google.com/scholar?hl=en{\

Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia*, *49*(5), 970–983. doi:10.1016/j.neuropsychologia.2011.01.003

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. doi:10.1037/a0016261

Aravena, P., Hurtado, E., Riveros, R., Cardona, J. F., Manes, F., & Ibáñez, A. (2010). Applauding with closed hands: neural signature of action-sentence compatibility effects. *PloS One*, *5*(7), e11751. doi:10.1371/journal.pone.0011751

Arzouan, Y., Goldstein, A., & Faust, M. (2007). Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Research*, *1160*, 69–81. doi:10.1016/j.brainres.2007.05.034

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics By 8-Month-Old Infants. *Psychological Science*, *9*(4), 321–324.

Auguie, B. (2017). *GridExtra: Miscellaneous functions for "grid" graphics*. Retrieved from https://CRAN.R-project.org/package=gridExtra

Aust, F. (2016). *Citr: 'RStudio' add-in to insert markdown citations*. Retrieved from https://CRAN.R-project.org/package=citr

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent Embodied Representations for Visually Presented Actions and Linguistic Phrases Describing Actions. *Current Biology*, *16*(18), 1818–1823. doi:10.1016/j.cub.2006.07.060

Baayen, R. H. (2010). A real experiment is a factorial experiment? *The Mental Lexicon*, *5*(1), 149–157. doi:10.1075/ml.5.1.06baa

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/J.JML.2007.12.005

Bache, S. M., & Wickham, H. (2014). *Magrittr: A forward-pipe operator for r.* Retrieved from https://CRAN.R-project.org/package=magrittr

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, *26*(9), 1338–1367. doi:10.1080/01690965.2010.542671

Bagiella, E., & Sloan, R. P. (2000). Mixed-effects models in psychophysiology. *Psychophysiology*, *37*(January), 13–20. doi:10.1111/1469-8986.3710013

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–59. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3193910{\&}tool=pmcentr

Barber, H. A., Otten, L. J., Kousta, S.-T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, *125*(1), 47–53. doi:10.1016/j.bandl.2013.01.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). doi:10.1016/j.jml.2012.11.001

Barrós-Loscertales, A., González, J., Pulvermüller, F., Ventura-Campos, N., Bustamante, J. C., Costumero, V., … Ávila, C. (2012). Reading salt activates gustatory brain regions: fMRI evidence for semantic grounding in a novel sensory modality. *Cerebral Cortex (New York, N.Y. : 1991)*, *22*(11), 2554–63. doi:10.1093/cercor/bhr324

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*(4), 577–609; discussion 610–60. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1693222{\&}tool=pmcentr

Barsalou, L. W. (2016). On staying grounded and avoiding Quixotic dead ends. *Psychonomic Bulletin & Review*, *23*, 1122–1142. doi:10.3758/s13423-016-1028-3

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In A. M. Glenberg & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 245–283). Oxford, England: Oxford University Press. Retrieved from http://psychology.emory.edu/cognition/barsalou/papers/Barsalou{_}et{_}al{_}cha

Bartoń, K. (2016). *MuMIn: Multi-Model Inference*. Retrieved from https://cran.r-project.org/package=MuMIn

Bartoń, K. (2017). *MuMIn: Multi-model inference.* Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., & Maechler, M. (2017). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

doi:10.18637/jss.v067.i01

Bernabeu, P., Willems, R. M., & Louwerse, M. M. (2017). Modality switch effects emerge early and increase throughout conceptual processing: Evidence from ERPs. In *39th annual conference of cognitive science society*. PsyArXiv. Retrieved from https://osf.io/preprints/psyarxiv/a5pcz/

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 295–308. doi:10.1037//0278-7393.19.2.295

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022. Retrieved from http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Bonner, M. F., & Grossman, M. (2012). Gray matter density of auditory association cortex relates to knowledge of sound concepts in primary progressive aphasia. *Journal of Neuroscience*, *32*(23), 7986–7991. doi:10.1523/JNEUROSCI.6241-11.2012

Borghi, A. M., & Binkofski, F. (2014). *Words as Social Tools: An Embodied View on Abstract Concepts*. New York, NY: Springer New York. doi:10.1007/978-1-4614-9539-0

Borghi, A. M., Caramelli, N., & Setti, A. (2005). Conceptual information on objects' locations. *Brain and Language*, *93*(2), 140–151. doi:10.1016/j.bandl.2004.09.004

Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition*, *75*(1), 1–28. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10815775

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*, 1–22. doi:10.1006/cogp.2001.0748

Borreggine, K. L., & Kaschak, M. P. (2006). The Action-Sentence Compatibility Effect: It's All in the Timing. *Cognitive Science*, *30*(6), 1097–1112. doi:10.1207/s15516709cog0000_91

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., & Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, *46*(2), 743–756. doi:10.1016/j.neuropsychologia.2007.10.007

Boulenger, V., Shtyrov, Y., & Pulvermüller, F. (2012). When do you grasp the idea? MEG evidence for instantaneous idiom understanding. *NeuroImage*, *59*(4), 3502–3513. doi:10.1016/j.neuroimage.2011.11.011

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*(1), 193–216. doi:10.1037/0033-295X.112.1.193

Brants, T., & Franz, A. (2006). Web 1T 5-gram Version 1 LDC2006T13. DVD. Philadelphia: Linguistic Data Consortium.

Brouillet, T., Heurley, L., Martin, S., & Brouillet, D. (2010). The embodied cognition theory and the motor component of "yes" and "no" verbal responses.

Acta Psychologica, *134*(3), 310–317. doi:10.1016/j.actpsy.2010.03.003

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–43. doi:10.1016/j.brainres.2012.01.055

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, *27*(1), 45–50. doi:10.1177/0963721417727521

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cognitive Brain Research*, *24*(3), 355–63. doi:10.1016/j.cogbrainres.2005.02.020

Buckner, R. L., & Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. *Trends in Cognitive Sciences*, *17*(12), 648–665. doi:10.1016/j.tics.2013.09.017

Campbell, S. J., & Raney, G. E. (2015). A 25-year replication of Katz et al.'s (1988) metaphor norms. *Behavior Research Methods*, *48*(1), 330–340. doi:10.3758/s13428-015-0575-2

Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods*, *42*(3), 651–64. doi:10.3758/BRM.42.3.651

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: tuning the brain for metaphors. *NeuroImage*, *59*(4), 3212–21. doi:10.1016/j.neuroimage.2011.11.079

Casasanto, D., & Gijssels, T. (2015). What makes a metaphor an embodied metaphor? *Linguistic Vanguard*, *1*(1), 327–337. doi:10.1515/lingvan-2014-1015

Chan, K. Q., Tong, E. M. W., Tan, D. H., & Koh, A. H. Q. (2013). What do love and jealousy taste like? *Emotion*, *13*(6), 1142–1149. doi:10.1037/a0033758

Chemero, A. (2009). *Radical embodied cognitive science* (p. 252). MIT Press.

Chiappe, D. L., & Kennedy, J. M. (1999). Aptness predicts preference for metaphors or similes, as well as recall bias. *Psychonomic Bulletin & Review*, *6*(4), 668–76. doi:10.3758/BF03212977

Chiappe, D. L., Kennedy, J. M., & Smykowski, T. (2003). Reversibility, Aptness, and the Conventionality of Metaphors and Similes. *Metaphor and Symbol*, *18*(January 2015), 85–105. doi:10.1207/S15327868MS1802_2

Chow, H. M., Mar, R. A., Xu, Y., Liu, S., Wagage, S., & Braun, A. R. (2014). Embodied Comprehension of Stories: Interactions between Language Regions and Modality-specific Neural Systems. *Journal of Cognitive Neuroscience*, *26*(2), 279–295. doi:10.1162/jocn_a_00487

Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical Sentences Are More

Emotionally Engaging than Their Literal Counterparts. *Journal of Cognitive Neuroscience*, *26*(11), 2585–2595. doi:10.1162/jocn_a_00654

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. doi:10.1016/S0022-5371(73)80014-3

Clement, C. A., Mawby, R., & Giles, D. E. (1994). The Effects of Manifest Relational Similarity on Analog Retrieval. *Journal of Memory and Language*, *33*(3), 396–420. doi:10.1006/jmla.1994.1019

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, NJ: Erlbaum. Retrieved from https://books.google.co.uk/books?id=fAnSOgbdFXIC

Collins, J., Pecher, D., Zeelenberg, R., & Coulson, S. (2011). Modality Switching in a Property Verification Task: An ERP Study of What Happens When Candles Flicker after High Heels Click. *Frontiers in Psychology*, *2*(February), 10. doi:10.3389/fpsyg.2011.00010

Columbus, G., Sheikh, N. A., Côté-Lecaldare, M., Häuser, K., Baum, S. R., & Titone, D. (2015). Individual differences in executive control relate to metaphor processing: an eye movement study of sentence reading. *Frontiers in Human Neuroscience*, *8*(January), 1–12. doi:10.3389/fnhum.2014.01057

Colunga, E., & Smith, L. B. (2003). The emergence of abstract ideas: Evidence from networks and babies. doi:10.1098/rstb.2003.1306

Connell, L. (2007). Representing object colour in language comprehension. *Cognition*, *102*(3), 476–85. doi:10.1016/j.cognition.2006.02.009

Connell, L. (2018). What have labels ever done for us? The lingusitic shortcut in conceptual processing. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2018.1471512

Connell, L., & Lynott, D. (2011). Modality Switching Costs Emerge in Concept Creation as Well as Retrieval. *Cognitive Science*, *35*(4), 763–778. doi:10.1111/j.1551-6709.2010.01168.x

Connell, L., & Lynott, D. (2012). Principles of representation: Why you can't represent the same concept twice. *Under Review*, *526*. Retrieved from http://scholar.google.com/scholar?hl=en{\&}btnG=Search{\&}q=intitle:Principles+of

Connell, L., & Lynott, D. (2013). Flexible and fast: linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, *20*(3), 542–50. doi:10.3758/s13423-012-0368-x

Connell, L., & Lynott, D. (2014a). I See/Hear What You Mean: Semantic Activation in Visual Word Recognition Depends on Perceptual Attention. *Journal of Experimental Psychology. General*, *143*(2), 527–533. doi:10.1037/a0034626

Connell, L., & Lynott, D. (2014b). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*(3), 390–

406. doi:10.1111/tops.12097

Connell, L., & Lynott, D. (2016). Do we know what we're simulating? Information loss on transferring unconscious perceptual simulation to conscious imagery. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(8), 1218–32. doi:10.1037/xlm0000245

Connell, L., Lynott, D., & Dreyer, F. (2012). A functional role for modality-specific perceptual systems in conceptual representations. *PloS One*, *7*(3), e33321. doi:10.1371/journal.pone.0033321

Coulson, S. (2007). Electrifying results: ERP data and cognitive linguistics. In R. W. Gibbs (Ed.), *Methods in cognitive linguistics* (3rd ed., pp. 400–423). Amsterdam; Philadelphia: Cambridge University Press. Retrieved from http://books.google.com/books?hl=en{\&}lr={\&}id=EuTYvdYZM0gC{\&}oi=fnd{\

Coulson, S., & Severens, E. (2007). Hemispheric asymmetry and pun comprehension: When cowboys have sore calves. *Brain and Language*, *100*(2), 172–187. doi:10.1016/j.bandl.2005.08.009

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*(6), 958–968. doi:10.3758/BF03195780

Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we able to run any psycholinguistic experiments at all in 1990? *Cognition*, *10*(1-3), 65–70. doi:10.1016/0010-0277(81)90026-3

Dantzig, S. van, Cowell, R. A., Zeelenberg, R., & Pecher, D. (2011). A sharp image or a sharp knife: norms for the modality-exclusivity of 774 concept-property items. *Behavior Research Methods*, *43*(1), 145–54. doi:10.3758/s13428-010-0038-8

Davenport, T., & Coulson, S. (2011). Predictability and novelty in literal language comprehension: an ERP study. *Brain Research*, *1418*, 70–82. doi:10.1016/j.brainres.2011.07.039

Davenport, T., & Coulson, S. (2013). Hemispheric asymmetry in interpreting novel literal language: an event-related potential study. *Neuropsychologia*, *51*(5), 907–21. doi:10.1016/j.neuropsychologia.2013.01.018

De Grauwe, S., Swain, A., & Holcomb, P. (2010). Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia*, *48*(7), 1965–1984. doi:10.1016/j.neuropsychologia.2010.03.017.Electrophysiological

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., … Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597–600. doi:10.1038/26967

Desai, R. H., Binder, J. R., Conant, L. L., Mano, Q. R., & Seidenberg, M. S. (2011). The neural career of sensory-motor metaphors. *Journal of Cognitive Neuroscience*, *23*(9), 2376–86. doi:10.1162/jocn.2010.21596

Desai, R. H., Conant, L. L., Binder, J. R., Park, H., & Seidenberg, M. S. (2013). A piece of the action: Modulation of sensory-motor regions by action idioms and metaphors. *NeuroImage*, *83*, 862–869. doi:10.1016/j.neuroimage.2013.07.044

Dowle, M., & Srinivasan, A. (2017). *Data.table: Extension of 'data.frame'*. Retrieved from https://CRAN.R-project.org/package=data.table

Engelen, J., Bouwmeester, S., Bruin, A. B. de, & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, *110*(4), 659–75. doi:10.1016/j.jecp.2011.06.009

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. doi:10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. doi:10.3758/BF03193146

Faust, M., & Mashal, N. (2007). The role of the right cerebral hemisphere in processing novel metaphoric expressions taken from poetry: a divided visual field study. *Neuropsychologia*, *45*(4), 860–70. doi:10.1016/j.neuropsychologia.2006.08.010

Federmeier, K. D. (2002). Picture the difference: electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*, *40*(7), 730–747. doi:10.1016/S0028-3932(01)00193-2

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., … Seidenberg, M. S. (2016). Concept Representation Reflects Multimodal Abstraction: A Framework for Embodied Semantics. *Cerebral Cortex*, *26*(5), 2018–2034. doi:10.1093/cercor/bhv020

Ferreira, F., Bailey, K., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. Retrieved from http://cdp.sagepub.com/content/11/1/11.short

Firth, J. R. (1957). A synopsis of linguistic theory 1930-55. In *Studies in linguistic analysis (special volume of the philological society)* (Vols. 1952-59, pp. 1–32). Oxford, England: Blackwell Publishers. Retrieved from https://books.google.co.uk/books/about/A{_}Synopsis{_}of{_}Linguistic{_}The

Forgacs, B., Bardolph, M. D., Amsel, B. D., DeLong, K. a., & Kutas, M. (2015). Metaphors are physical and abstract: ERPs to metaphorically modified nouns resemble ERPs to abstract language. *Frontiers in Human Neuroscience*, *9*(February), 1–11. doi:10.3389/fnhum.2015.00028

Gagné, C. L., & Shoben, E. J. (1997). Influence of Thematic Relations on the Comprehension of Modifier–Noun Combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 71–87. doi:10.1037/0278-7393.23.1.71

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The Search for "Common Sense": An Electrophysiological Study of the Comprehension of Words and Pictures in Reading. *Journal of Cognitive Neuroscience*, *8*(2), 89–106. doi:10.1162/jocn.1996.8.2.89

Gelman, A., & Su, Y.-S. (2016). *Arm: Data analysis using regression and multilevel/hierarchical models*. Retrieved from https://CRAN.R-project.org/package=arm

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, *7*(2), 155–170. doi:10.1016/S0364-0213(83)80009-3

Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language*, *37*(3), 331–355. doi:10.1006/jmla.1997.2527

Genz, A., & Bretz, F. (2009). *Computation of multivariate normal and t probabilities*. Heidelberg: Springer-Verlag.

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, *8*(2), 149–156. doi:10.3758/BF03213418

Gibbs, R. W. (2006). Metaphor interpretation as embodied simulation. doi:10.1111/j.1468-0017.2006.00285.x

Gibbs, R. W., Costa Lima, P. L., & Francozo, E. (2004). Metaphor is grounded in embodied experience. *Journal of Pragmatics*, *36*(7), 1189–1210. doi:10.1016/j.pragma.2003.10.009

Gineste, M.-D., Indurkhya, B., & Scart, V. (2000). Emergence of Features in Metaphor Comprehension. *Metaphor and Symbol*, *15*(3), 37–41. doi:10.1207/S15327868MS1503

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183–206. Retrieved from http://www.tau.ac.il/{~}giorar/files/Giora{_}97{_}GradedSalienceHypothesis.pdf

Giora, R. (2002). Literal vs. figurative language: Different or equal? *Journal of Pragmatics*, (34), 487–506. Retrieved from http://www.sciencedirect.com/science/article/pii/S0378216601000455

Giora, R. (2007). Is metaphor special? *Brain and Language*, *100*(2), 111–4. doi:10.1016/j.bandl.2006.08.001

Giora, R., & Fein, O. (1999). On understanding familiar and less-familiar figurative language. *Journal of Pragmatics*, *31*(12), 1601–1618. doi:10.1016/S0378-2166(99)00006-5

Giora, R., Zaidel, E., & Soroker, N. (2000). Differential Effects of Right- and Left-Hemisphere Damage on Understanding Sarcasm and Metaphor. *Metaphor and ...*, *15*(1986), 63–83. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/10926488.2000.9678865

Glantz, S. A., & Slinker, B. K. (1990). *Primer of Applied Regression and Analysis of Variance* (p. 777). McGraw-Hill Education. Retrieved from https://books.google.com/books?id=UcR6QgAACAAJ{\&}pgis=1

Glenberg, A. M., & Gallese, V. (2012). Action-based language: a theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905–22.

doi:10.1016/j.cortex.2011.04.010

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–65. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12412897

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., & Buccino, G. (2008). Processing abstract language modulates motor system activity. *The Quarterly Journal of Experimental Psychology*, *61*(6), 905–919. doi:10.1080/17470210701625550

Glucksberg, S. (1991). Beyond Literal Meanings : The Psychology of Allusion. *Psychological Science*, *2*(3), 146–152. Retrieved from http://pss.sagepub.com/content/2/3/146.short

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, *7*(2), 92–96. doi:10.1016/S1364-6613(02)00040-2

Glucksberg, S., & Haught, C. (2006a). Can Florida become like the next Florida? When metaphoric comparisons fail. *Psychological Science*, *17*(11), 935–938. doi:10.1111/j.1467-9280.2006.01807.x

Glucksberg, S., & Haught, C. (2006b). On the Relation Between Metaphor and Simile: When Comparison Fails. *Mind and Language*, *21*(3), 360–378. doi:10.1111/j.1468-0017.2006.00282.x

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, *97*(1), 3–18. Retrieved from http://psycnet.apa.org/journals/rev/97/1/3/

Glucksberg, S., Gildea, P., & Bookin, H. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, (21), 85–98. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022537182904674

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual Knowledge Retrieval Activates Sensory Brain Regions. *Journal of Neuroscience*, *26*(18), 4917–4921. doi:10.1523/JNEUROSCI.5389-05.2006

Grice, H. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Speech acts: Syntax and semantics, vol (3)* (pp. 41–58). New York: Academic Press. Retrieved from http://books.google.com/books?hl=en{\&}lr={\&}id=84RpGGJ

Günther, F., Dudschig, C., & Kaup, B. (2017). Symbol Grounding Without Direct Experience: Do Words Inherit Sensorimotor Activation From Purely Linguistic Context? *Cognitive Science*, 1–39. doi:10.1111/cogs.12549

Hald, L. A., Hocking, I., Vernon, D., Marshall, J.-A., & Garnham, A. (2013). Exploring modality switching effects in negated sentences: further evidence for grounded representations. *Frontiers in Psychology*, *4*(February), 93. doi:10.3389/fpsyg.2013.00093

Hald, L. A., Marshall, J.-A., Janssen, D. P., & Garnham, A. (2011). Switching Modalities in A Sentence Verification Task: ERP Evidence for Embodied Language Processing. *Frontiers in Psychology*, *2*(March), 45.

doi:10.3389/fpsyg.2011.00045

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1-3), 335–346. doi:10.1016/0167-2789(90)90087-6

Harrell Jr, F. E. (2018). *Rms: Regression modeling strategies.* Retrieved from https://CRAN.R-project.org/package=rms

Harrell Jr, F. E., Charles Dupont, & others. (2018). *Hmisc: Harrell miscellaneous.* Retrieved from https://CRAN.R-project.org/package=Hmisc

Haught, C. (2013). A Tale of Two Tropes: How Metaphor and Simile Differ. *Metaphor and Symbol*, *28*(4), 254–274. doi:10.1080/10926488.2013.826555

Haught, C. (2014). Spain is Not Greece: How Metaphors are Understood. *Journal of Psycholinguistic Research*, *43*(4), 351–356. doi:10.1007/s10936-013-9258-2

Hauk, O. (2016). Only time will tell – why temporal information is essential for our neuroscientific understanding of semantics. *Psychonomic Bulletin & Review*, *23*(4), 1072–1079. doi:10.3758/s13423-015-0873-9

Hauk, O., & Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, *21*(3), 191–201. doi:10.1002/hbm.10157

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, *41*(2), 301–307. doi:10.1016/S0896-6273(03)00838-9

Hauk, O., Pulvermüller, F., Ford, M., Marslen-Wilson, W. D., & Davis, M. H. (2009). Can I have a quick word? Early electrophysiological manifestations of psycholinguistic processes revealed by event-related regression analysis of the EEG. *Biological Psychology*, *80*(1), 64–74. doi:10.1016/j.biopsycho.2008.04.015

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory.* New York: Hardcover; Wiley.

Hendrix, P., Bolger, P., & Baayen, R. H. (2017). Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 128–149. doi:10.1037/a0040332

Henry, L., & Wickham, H. (2017). *Purrr: Functional programming tools.* Retrieved from https://CRAN.R-project.org/package=purrr

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical Signs of Selective Attention in the Human Brain. *Science*, *182*(4108), 177–180. doi:10.1126/science.182.4108.177

Hothorn, T. (2017). *TH.data: TH's data archive.* Retrieved from https://CRAN.R-project.org/package=TH.data

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Huang, H. W., Lee, C. L., & Federmeier, K. D. (2010). Imagine that! ERPs provide evidence for distinct hemispheric contributions to the processing of concrete and abstract concepts. *NeuroImage*, *49*(1), 1116–1123. doi:10.1016/j.neuroimage.2009.07.031

Hurtienne, J., & Meschke, O. (2016). Soft Pillows and the Near and Dear. In *Proceedings of the tei '16: Tenth international conference on tangible, embedded, and embodied interaction - tei '16* (pp. 324–331). New York, New York, USA: ACM Press. doi:10.1145/2839462.2839483

Højsgaard, S., & Halekoh, U. (2016). *DoBy: Groupwise statistics, lsmeans, linear contrasts, utilities*. Retrieved from https://CRAN.R-project.org/package=doBy

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes Factors. *The Journal of Problem Solving*, *7*, 2–9. doi:10.7771/1932-6246.1167

Jeffreys, H. (1998). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press. Retrieved from https://global.oup.com/academic/product/the-theory-of-probability-9780198503682?cc=gb{\&}lang=en{\&}

Johns, B. T., & Jones, M. N. (2012). Perceptual Inference Through Global Lexical Similarity. *Topics in Cognitive Science*, *4*(1), 103–120. doi:10.1111/j.1756-8765.2011.01176.x

Jones, L. L., & Estes, Z. (2005). Metaphor comprehension as attributive categorization. *Journal of Memory and Language*, *53*(1), 110–124. doi:10.1016/j.jml.2005.01.016

Jones, L. L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language*, *55*(1), 18–32. doi:10.1016/j.jml.2006.02.004

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, *9*(11), 512–8. doi:10.1016/j.tics.2005.09.009

Katsumata, S. (1980). *Kenkyusha's new dictionary of English collocations*. Kenkyusha. Retrieved from https://books.google.co.uk/books?id=7UBvswEACAAJ

Katz, A. N., Paivio, A., Marshark, M., & Clark, J. M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor and Symbolic Activity*, *3*(4), 191–214.

Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(47), 12224–30. doi:10.1523/JNEUROSCI.3579-08.2008

Kintsch, W., & Dijk, T. V. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. Retrieved from http://psycnet.apa.org/journals/rev/85/5/363/

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), 4–5. doi:10.1016/S0010-0277(02)00004-5

Koenker, R., & Ng, P. (2017). *SparseM: Sparse linear algebra.* Retrieved from https://CRAN.R-project.org/package=SparseM

Koles, Z. J. (1998). Trends in EEG source localization. *Electroencephalography and Clinical Neurophysiology*, *106*(2), 127–137. doi:10.1016/S0013-4694(97)00115-6

Kuhn, M., Steve Weston, Wing, J., Forester, J., & Thaler, T. (2016). *Contrast: A collection of contrast methods.* Retrieved from https://CRAN.R-project.org/package=contrast

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*(1), 23–49. doi:10.1016/j.brainres.2006.12.063

Kurby, C. A., & Zacks, J. M. (2013). The activation of modality-specific representations during discourse processing. *Brain and Language*, *126*(3), 338–349. doi:10.1016/J.BANDL.2013.07.003

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–47. doi:10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. doi:10.1038/307161a0

Kutas, M., Federmeier, K. D., & Sereno, M. I. (1999). Current approaches to mapping language in electromagnetic space. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 359–392). New York: Oxford University Press. doi:10.1093/acprof:oso/9780198507932.003.0012

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13

Lai, V. T., & Curran, T. (2013). ERP evidence for conceptual mappings and comparison processes during the comprehension of conventional and novel metaphors. *Brain and Language*, *127*(3), 484–496. doi:10.1016/j.bandl.2013.09.010

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, *1284*, 145–155. doi:10.1016/j.brainres.2009.05.088

Lakoff, G., & Johnsen, M. (2003). *Innholdsfortegnelse i Word.*

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought.* Basic Books. Retrieved from http://books.google.co.uk/books?id=KbqxnX3{_}uc0C

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–40. doi:10.1037/0033-295X.104.2.211

Laurent, J. P., Denhières, G., Passerieux, C., Iakimova, G., & Hardy-Baylé, M. C. (2006). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research*, *1068*(1), 151–160. doi:10.1016/j.brainres.2005.10.076

Lebois, L. A. M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are Automatic Conceptual Cores the Gold Standard of Semantic Processing? The Context-Dependence of Spatial Meaning in Grounded Congruency Effects. *Cognitive Science*, *39*(8), 1764–1801. doi:10.1111/cogs.12174

Leshinskaya, A., & Caramazza, A. (2016). For a cognitive neuroscience of concepts: Moving beyond the grounding issue. doi:10.3758/s13423-015-0870-z

Liu, P. Q., Connell, L., & Lynott, D. (2017). Can illness be bright? Metaphor comprehension depends on linguistic and embodied factors. In G. Gunzelmann, A. Hows, T. Tenbrink, & E. J. Davelaar (Eds.), *39th annual conference of cognitive science society* (pp. 2604–2609). Austin, TX: Cognitive Science Society. Retrieved from https://mindmodeling.org/cogsci2017/papers/0494/paper0494.pdf

Liu, P. Q., Connell, L., & Lynott, D. (2018a). Can illness be bright? Metaphor comprehension depends on linguistic and embodied factors. *Manuscript Under Review*, 1–54.

Liu, P. Q., Connell, L., & Lynott, D. (2018b). Continuous neural activities of modality switching costs. *Manuscript in Preparation.*

Liu, P. Q., Connell, L., & Lynott, D. (2018c). Ease-of-Simulation Norms for 452 Adjective Metaphors. *Manuscript Under Review*, 1–27.

Liu, P. Q., Connell, L., & Lynott, D. (2018d). Ease-of-Simulation norms for 452 Adjective Metaphors. *Manuscript Under Review.*

Liu, P. Q., Connell, L., & Lynott, D. (2018e). Effect of time constraints on the grounded representation of metaphor processing. *Manuscript in Preparation*, 1–30.

Liu, P. Q., Connell, L., & Lynott, D. (2018f). Switching costs between both perceptual and linguistic modalities. *Manuscript in Preparation.*

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*(2), 273–302. doi:10.1111/j.1756-8765.2010.01106.x

Louwerse, M. M., & Benesh, N. (2012). Representing Spatial Structure Through Maps and Language: Lord of the Rings Encodes the Spatial Structure of Middle Earth. *Cognitive Science*, *36*(8), 1556–1569. doi:10.1111/cogs.12000

Louwerse, M. M., & Connell, L. (2011). A taste of words: linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, *35*(2), 381–98. doi:10.1111/j.1551-6709.2010.01157.x

Louwerse, M. M., & Hutchinson, S. (2012). Neurological evidence linguistic processes precede perceptual simulation in conceptual processing. *Frontiers in Psychology*, *3*(October), 385. doi:10.3389/fpsyg.2012.00385

Louwerse, M. M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. In A. Glenberg & A. C. Graesser (Eds.), *Embodiment and meaning: A debate* (pp. 309–326). Oxford, England: Oxford University Press. Retrieved from http://www.madresearchlab.org/Selected{_}Publications{_}fil

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*(1), 96–104. doi:10.1016/j.cognition.2009.09.002

Louwerse, M. M., Hutchinson, S., & Cai, Z. (2012). The Chinese route argument: Predicting the longitude and latitude of cities in China and the Middle East using statistical linguistic frequencies. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 695–700. Retrieved from http://mindmodeling.org/cogsci2012/papers/0130/paper0130.pdf

Louwerse, M. M., Hutchinson, S., Tillman, R., & Recchia, G. (2015). Effect size matters: the role of language statistics and perceptual simulation in conceptual processing. *Language, Cognition and Neuroscience*, *30*(4), 430–447. doi:10.1080/23273798.2014.981552

Luck, S. J. (2005). Ten Simple Rules for Deisgning ERP Experiments. In *Event-related potentials: A methods handbook* (pp. 17–32). doi:10.1371/journal.pcbi.0020012

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. doi:10.3758/BF03204766

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41*(2), 558–64. doi:10.3758/BRM.41.2.558

Lynott, D., & Connell, L. (2010). Embodied conceptual combination. *Frontiers in Psychology*, *1*(November), 212. doi:10.3389/fpsyg.2010.00212

Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: the relationship between perceptual experience and surface word form. *Behavior Research Methods*, *45*(2), 516–26. doi:10.3758/s13428-012-0267-0

MacGregor, L. J., Bouwsema, J., & Klepousniotou, E. (2015). Sustained meaning activation for polysemous but not homonymous words: Evidence from EEG. *Neuropsychologia*, *68*, 126–138. doi:10.1016/j.neuropsychologia.2015.01.008

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, *30*(4), 420–429. doi:10.1080/23273798.2014.987791

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology Paris*, *102*(1-3), 59–70. doi:10.1016/j.jphysparis.2008.03.004

Mashal, N. (2013). The role of working memory in the comprehension of unfamiliar and familiar metaphors. *Language and Cognition*, *5*(4), 409–436.

doi:10.1515/langcog-2013-0024

Mashal, N., & Faust, M. (2008). Right hemisphere sensitivity to novel metaphoric relations: Application of the signal detection theory. *Brain and Language*, *104*(2), 103–112. doi:10.1016/j.bandl.2007.02.005

Mashal, N., Faust, M., & Hendler, T. (2005). The role of the right hemisphere in processing nonsalient metaphorical meanings: application of principal components analysis to fMRI data. *Neuropsychologia*, *43*(14), 2084–100. doi:10.1016/j.neuropsychologia.2005.03.019

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman, M. (2009). An fMRI study of processing novel metaphoric sentences. *Laterality*, *14*(1), 30–54. doi:10.1080/13576500802049433

McGlone, M. S., & Manfredi, D. A. (2001). Topic-vehicle interaction in metaphor comprehension. *Memory & Cognition*, *29*(8), 1209–1219. doi:10.3758/BF03206390

McRae, K., Nedjadrasul, D., Pau, R., Lo, B. P.-H., & King, L. (2018). Abstract Concepts and Pictures of Real-World Situations Activate One Another. *Topics in Cognitive Science*, 1–15. doi:10.1111/tops.12328

Miller, J., & Hackley, S. A. (1992). Electrophysiological Evidence for Temporal Overlap Among Contingent Mental Processes. *Journal of Experimental Psychology: General*, *121*(2), 195–209. Retrieved from http://web.missouri.edu/{~}hackleys/docs/MillerHackley1992.pdf

Miller, T. M. C., Schmidt, T. T., Blankenburg, F., & Pulvermüller, F. (2018). Verbal labels facilitate tactile perception. *Cognition*, *171*, 172–179. doi:10.1016/j.cognition.2017.10.010

Monaghan, P., Chang, Y.-N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, *93*, 1–21. doi:10.1016/J.JML.2016.08.003

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of bayes factors for common designs*. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Moscoso Del Prado Martín, F., Hauk, O., & Pulvermüller, F. (2006). Category specificity in the processing of color-related and form-related words: An ERP study. *NeuroImage*, *29*(1), 29–37. doi:10.1016/j.neuroimage.2005.07.055

Müller, K. (2017). *Bindrcpp: An 'rcpp' interface to active bindings*. Retrieved from https://CRAN.R-project.org/package=bindrcpp

Nelson, L. D., Simmons, J. P., & Simonsohn, U. (2018). Psychology's renaissance. *Annu. Rev. Psychol*, *69*(17), 1–24. doi:10.1146/annurev-psych-122216

Niedenthal, P. M. (2007). Embodying Emotion. *Science*, *316*(5827), 1002–1005. doi:10.1126/science.1136930

Nowosad, J. (2017). *Rcartocolor: 'CARTOColors' palettes*. Retrieved from https://CRAN.R-project.org/package=rcartocolor

Olivares, E. I., Iglesias, J., & Antonieta Bobes, M. (1999). Searching for face-specific long latency ERPs: a topographic study of effects associated with mismatching features. *Cognitive Brain Research*, *7*(3), 343–356. doi:10.1016/S0926-6410(98)00038-X

Olofsson, J. K., Hurley, R. S., Bowman, N. E., Bao, X., Mesulam, M.-M., & Gottfried, J. A. (2014). A designated odor-language integration system in the human brain. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(45), 14864–73. doi:10.1523/JNEUROSCI.2247-14.2014

Otten, L. J., & Rugg, M. D. (2005). Interpreting event-related brain potentials. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 3–16). Cambridge: MIT Press. Retrieved from http://discovery.ucl.ac.uk/185452/

Öttl, B., Dudschig, C., & Kaup, B. (2017). Forming associations between language and sensorimotor traces during novel word learning. *Language and Cognition*, *9*(01), 156–171. doi:10.1017/langcog.2016.5

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, *67*(4), 426–448. doi:10.1016/j.jml.2012.07.003

Paivio, A. (1986). *Mental Representations : a Dual Coding Approach.* (p. 335). New York: Oxford University Press, Incorporated. Retrieved from https://books.google.co.uk/books?hl=en{\&}lr={\&}id=hLGmKkh{_}4K8C{\&}oi= representations{\%}3A A dual coding approach.{\&}f=false

Paivio, A., Yuille, J., & Madigan, S. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1), 1–25. Retrieved from http://psycnet.apa.org/journals/xge/76/1p2/1/

Papesh, M. H. (2015). Just out of reach: On the reliability of the action-sentence compatibility effect. *Journal of Experimental Psychology: General*, *144*(6), e116–e141. doi:10.1037/xge0000125

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, *14*(2), 119–24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12661672

Peleg, O., Giora, R., & Fein, O. (2001). Salience and Context Effects: Two Are Better Than One. *Metaphor and Symbol*, *16*(3-4), 173–192. doi:10.1080/10926488.2001.9678894

Pierce, R. S., & Chiappe, D. L. (2008). The Roles of Aptness, Conventionality, and Working Memory in the Production of Metaphors and Similes. *Metaphor and Symbol*, *24*(1), 1–19. doi:10.1080/10926480802568422

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2017). *nlme: Linear and nonlinear mixed effects models.* Retrieved from https://CRAN.R-project.org/package=nlme

Plate, T., & Heiberger, R. (2016). *Abind: Combine multidimensional arrays.*

Retrieved from https://CRAN.R-project.org/package=abind

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for mcmc. *R News*, *6*(1), 7–11. Retrieved from https://journal.r-project.org/archive/

Pobric, G., Mashal, N., Faust, M., & Lavidor, M. (2008). The Role of the Right Cerebral Hemisphere in Processing Novel Metaphoric Expressions: A Transcranial Magnetic Stimulation Study. *Journal of Cognitive Neuroscience*, *20*(1), 170–181. doi:10.1162/jocn.2008.20005

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–48. doi:10.1016/j.clinph.2007.04.019

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain Signatures of Meaning Access in Action Word Recognition. *Journal of Cognitive Neuroscience*, *17*(6), 884–892. doi:10.1162/0898929054021111

Pynte, J., Besson, M., Robichon, F. H., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain and Language*, *316*(55), 293–316. Retrieved from http://www.sciencedirect.com/science/article/pii/S0093934X96901078

R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rabovsky, M., & McRae, K. (2014). The N400 component of the ERP as semantic network error: Insights from feature-based connectionist attractor models of word meaning. *Cognition*, *132*, 68–89. doi:10.3389/conf.fncom.2012.55.00242

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47*(2), 388–396. doi:10.1016/j.neuropsychologia.2008.09.017

Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality research.* Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*(2), 303–345. doi:10.1111/j.1756-8765.2010.01111.x

Robinson, D. (2017). *Broom: Convert statistical analysis objects into tidy data frames.* Retrieved from https://CRAN.R-project.org/package=broom

Romero Lauro, L. J., Mattavelli, G., Papagno, C., & Tettamanti, M. (2013). She runs, the road runs, my mind runs, bad blood runs between us: Literal and figurative motion verbs: An fMRI study. *NeuroImage*, *83*, 361–371. doi:10.1016/j.neuroimage.2013.06.050

Roncero, C., & Almeida, R. G. de. (2014a). Semantic properties, aptness, familiarity, conventionality, and interpretive diversity scores for 84 metaphors and similes. *Behavior Research Methods*. doi:10.3758/s13428-014-0502-y

Roncero, C., & Almeida, R. G. de. (2014b). The importance of being apt: metaphor comprehension in Alzheimer's disease. *Frontiers in Human Neuroscience*, *8*, 973. doi:10.3389/fnhum.2014.00973

Rundell, M. (2007). *Macmillan English Dictionary for Advanced Learners*. Macmillan. Retrieved from https://books.google.co.uk/books?id=OHLOLAAACAAJ

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. doi:10.1016/S0010-0277(98)00075-4

San Roque, L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M., ... Majid, A. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics*, *26*(1), 31–60. doi:10.1515/cog-2014-0089

Santana, E., & Vega, M. de. (2011). Metaphors are embodied, and so are their literal counterparts. *Frontiers in Psychology*, *2*(MAY), 90. doi:10.3389/fpsyg.2011.00090

Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Language and Cognition*, *3111*, 83–119. doi:10.1515/

Sarkar, D. (2008). *Lattice: Multivariate data visualization with r*. New York: Springer. Retrieved from http://lmdvr.r-forge.r-project.org

Schmidt-Snoek, G. L., Drew, A. R., Barile, E. C., & Agauas, S. J. (2015). Auditory and motion metaphors have different scalp distributions: an ERP study. *Frontiers in Human Neuroscience*, *9*(March), 1–9. doi:10.3389/fnhum.2015.00126

Searle, J. R. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences*, *3*(3), 417–457. doi:10.1017/S0140525X00005756

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. doi:10.1177/0956797611417632

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *SSRN Electronic Journal*. doi:10.2139/ssrn.2160588

Simmons, W. K., & Barsalou, L. W. (2003, May). The similarity-in-topography principle: Reconciling theories of conceptual deficits. doi:10.1080/02643290342000032

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology Paris*, *102*(1-3), 106–119. doi:10.1016/j.jphysparis.2008.03.014

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Marin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802–2810. doi:10.1016/j.neuropsychologia.2007.05.002.A

Slotnick, S. D. (2005). Source Localisazation of ERP generators. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 149–166). Cambridge: The MIT Press.

Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, *32*(2), 244–259. doi:10.3758/BF03196856

Solt, F., & Hu, Y. (2015). interplot: Plot the effects of variables in interaction terms. Available at The Comprehensive R Archive Network (CRAN). Retrieved from https://CRAN.R-project.org/package=interplot

Speed, L. J., & Majid, A. (2017). Dutch modality exclusivity norms: Simulating perceptual modality in space. *Behavior Research Methods*, *49*(6), 2204–2218. doi:10.3758/s13428-017-0852-3

Speed, L. J., Dam, W. O. van, Hirath, P., Vigliocco, G., & Desai, R. H. (2017). Impaired Comprehension of Speed Verbs in Parkinson's Disease. *Journal of the International Neuropsychological Society*, *23*(05), 412–420. doi:10.1017/S1355617717000248

Spence, C., Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*(2), 330–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11281107

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447–481. doi:10.1016/S0010-0285(02)00503-0

Stevenson, A. (2010). *Oxford Dictionary of English*. OUP Oxford. Retrieved from https://books.google.co.uk/books?id=TaZZSAAACAAJ

Storms, G., & Wisniewski, E. J. (2005). Does the order of head noun and modifier explain response times in conceptual combination? *Memory & Cognition*, *33*(5), 852–61. doi:10.3758/BF03193080

Tendahl, M., & Gibbs, R. W. (2008). Complementary perspectives on metaphor: Cognitive linguistics and relevance theory. *Journal of Pragmatics*, *40*(11), 1823–1864. doi:10.1016/j.pragma.2008.02.001

Terry M. Therneau, & Patricia M. Grambsch. (2000). *Modeling survival data: Extending the Cox model*. New York: Springer.

Thibodeau, P. H., & Durgin, F. H. (2011). Metaphor Aptness and Conventionality: A Processing Fluency Account. *Metaphor and Symbol*, *26*(November), 206–226. doi:10.1080/10926488.2011.583196

Thibodeau, P. H., Sikos, L., & Durgin, F. H. (2017). Are subjective ratings of metaphors a red herring? The big two dimensions of metaphoric sentences. *Behavior Research Methods*, (May). doi:10.3758/s13428-017-0903-9

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522. doi:10.1038/381520a0

Van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cognitive Science*, *32*(3), 579–90. doi:10.1080/03640210802035365

van 't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12. doi:10.1016/j.jesp.2016.03.004

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Vermeulen, N., Corneille, O., & Niedenthal, P. M. (2008). Sensory load incurs conceptual processing costs. *Cognition*, *109*(2), 287–94. doi:10.1016/j.cognition.2008.09.004

Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex*, *24*(7), 1767–1777. doi:10.1093/cercor/bht025

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247. doi:10.1515/LANGCOG.2009.011

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1656–1674. doi:10.1037/0096-1523.24.6.1656

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. doi:10.3758/BF03194105

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & Maas, H. L. J. van der. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. doi:10.1037/a0022790

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*(1), 1–29. Retrieved from http://www.jstatsoft.org/v40/i01/

Wickham, H., & Henry, L. (2017). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Francois, R., Henry, L., & Müller, K. (2017). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wilke, C. O. (2017). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. Retrieved from https://CRAN.R-project.org/package=cowplot

Wilson, D., & Carston, R. (2006). Metaphor, Relevance and the 'Emergent

Property' Issue. *Mind & Language*, *21*(3), 404–433. doi:10.1111/j.1468-0017.2006.00284.x

Wilson, D., & Sperber, D. (2004). Relevance theory. In *Handbook of pragmatics* (pp. 607–632). doi:10.1016/j.pragma.2009.09.021

Wingfield, C., & Connell, L. (2018). Different cognitive tasks require different distributional language models: a systematic comparison of n-gram, count and predict models. *Manuscript in Preparation*.

Winter, B. (2016). Re-analyzing the synesthetic metaphor hierarchy using modality norms. In *Perception metaphor workshop*.

Wolff, P., & Gentner, D. (2011). Structure-mapping in metaphor comprehension. *Cognitive Science*, *35*(8), 1456–88. doi:10.1111/j.1551-6709.2011.01194.x

Wood, S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*, *65*(1), 95–114.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99*(467), 673–686.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36.

Wood, S., N., Pya, & S"afken, B. (2016). Smoothing parameter and model selection for general smooth models (with discussion). *Journal of the American Statistical Association*, *111*, 1548–1575.

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.name/knitr/

Xie, Y. (2016). *Bookdown: Authoring books and technical documents with R markdown*. Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://github.com/rstudio/bookdown

Yang, J., & Shu, H. (2016). Involvement of the Motor System in Comprehension of Non-Literal Action Language: A Meta-Analysis Study. *Brain Topography*, *29*(1), 94–107. doi:10.1007/s10548-015-0427-5

Zeileis, A., & Croissant, Y. (2010). Extended model formulas in R: Multiple parts and multiple responses. *Journal of Statistical Software*, *34*(1), 1–13. doi:10.18637/jss.v034.i01

Zwaan, R. A. (2004). The immersed experiences: toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35–62). New York, NY, US: Elseview Science. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.588.3040{\&}

Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences*, *18*(5), 229–234. doi:10.1016/j.tics.2014.02.008

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, *23*(4), 1028–1034. doi:10.3758/s13423-015-0864-x

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: six replication attempts. *PloS One*, *7*(12), e51382. doi:10.1371/journal.pone.0051382

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review*, *10*(4), 954–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15000544