Can Illness be Bright?
Metaphor Comprehension Depends on Linguistic and Embodied Factors

Pei Liu (p.liu1@lancaster.ac.uk)
Louise Connell (l.connell@lancaster.ac.uk)
Dermot Lynott (d.lynott@lancaster.ac.uk)
Department of Psychology, Fylde College, Lancaster University, Bailrigg, Lancaster, LA1 4YF, UK

Abstract
Conceptual representations in language processing employ both linguistic distributional and embodied information. Here, we aim to demonstrate the roles of these two components in metaphor processing. The linguistic component is captured by linguistic distributional frequency (LDF), that is, how often the constituent words appear together in context. The embodied component, on the other hand, refers to how easy it is to generate an embodied simulation, operationalised by a previous norming study. In the current study, we looked at the interplay of these components in metaphor processing, and investigated their roles at different depths of processing in two experiments. Thus, we required participants to engage in shallow processing (Experiment 1: Sensibility Judgement), or deep processing (Experiment 2: Interpretation Generation). Results showed that the increase of both variables made it more likely to accept a metaphor. However, whereas ease of simulation (EoS) contributed to the speed of processing at both levels of depth, LDF only affected the speed in shallow processing. Specifically, LDF acted as a heuristic, 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 views of language processing that emphasise the importance of both linguistic and embodied components according to task goals.

Keywords: metaphor processing; embodied cognition; linguistic distribution; simulation; depth of processing

Traditional views of metaphor processing
In a metaphoric expression, a word or a phrase (the source) is applied to an object or an action (the target) to which it cannot be literally applied. In the expression “a bright student”, a student is not an object to which the visual property of bright is usually applied. Nevertheless, we can comprehend it effortlessly meaning “clever or intelligent students”. How is this comprehension achieved? Many factors have been implicated, namely the familiarity, conventionality and aptness of a metaphor. These factors can not only affect the speed of processing (Giora, 2007; Pierce & Chiappe, 2008), but conventionality and aptness are also suggested to determine the mechanism of processing, whether by comparison or by categorisation (Bowdle & Gentner, 2005; Jones & Estes, 2006).

However, when it comes to understanding exactly how these three factors affect metaphor comprehension, they have problems with their theoretical specificity, and subsequently their operationalisation. Familiarity and conventionality are often used interchangeably, and they face the same problem concerning their operational definitions. They are sometimes assumed to refer to how often people have encountered the metaphor expression itself (e.g., how often is “bright” used to describe “students”? e.g., Cardillo, Watson, Schmidt, Kranjec, & Chatterjee, 2012; Roncero & de Almeida, 2014a) and sometimes to how accustomed people are to relating the expression to its metaphorical meaning (e.g., “bright” meaning intelligent and quick-witted: Campbell & Raney, 2015; Mashal, Faust, Hendler, & Jung-Beeman, 2009), 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., purple prose), or a metaphorical meaning might be encountered reasonably often via a different expression to the one supplied (e.g., “Solution can be bright”).

Aptness also has received different definitions. It is sometimes assumed to reflect a very general, high-level quality or goodness of a metaphor and is often operationalised as such (Haught, 2014), whereas at other times represents a much more low-level specification of how well the metaphorical meaning (e.g., intelligent and quick-witted) fits or overlaps with the target (e.g., “student”: Chiappe & Kennedy, 1999). In addition, aptness appears to be theoretically confounded with familiarity and conventionality. Only apt metaphors are likely to become conventionalised or familiar, as a metaphor that does not work well is unlikely to become widely used by speakers of a language. Because familiarity and conventionality depend on usage patterns of metaphors across a language, and usage patterns depend to some extent on aptness, it means that 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: Campbell & Raney, 2015; Roncero & de Almeida, 2014a), as are ratings of aptness and corpus frequency counts of the metaphor expression (r = .41-.57: Roncero & de Almeida, 2014b; Thibodeau & Durgin, 2011).

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 continuing to vary and refine how these factors are conceptualised, we propose a different approach to seek clearer predictors of metaphor processing that (a) are
Grounded views of language processing

Research in conceptual representation 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; Connell & Lynott, 2014). The first component relies on the statistical, distributional pattern of how words co-occur across contexts (Landauer & Dumais, 1997). The second type of representation is the embodied (also known as the grounded, sensorimotor or situated) component, which relies on the process of simulation; that is the partial reactivation of past perceptual, motor, affective, introspective and other experiences (Barsalou, 1999).

Together, the linguistic and embodied components can explain language processing better than either alone (e.g., Andrews, Vigliocco, & Vinson, 2009). In particular, research in the grounded linguistic-embodied approach has demonstrated that the linguistic distributional information provides a powerful tool for superficial language processing because activity of the linguistic component peaks earlier than that of the embodied simulation component (Louwiser & Jeuniaux, 2008). People are more likely to rely on the embodied component when deeper processing is specifically cued in the task; but people will be reliant upon the linguistic component to generate a good-enough approximation (Ferreira, Bailey, & Ferraro, 2002) when shallow processing can suffice.

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 costly embodied simulation. If the linguistic component indicates that future processing is likely to fail (e.g., the words rarely co-occur in the same context and so their combined meaning might not be simulated successfully), then the processing cold be abandoned 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 (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.

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, operate in theory across all types of language comprehension. In this study, we asked participants to process metaphors that systematically varied on these two dimensions. The linguistic component is quantified by linguistic distributional frequency (LDF), that is how often constituent words (“bright” and “student”) of a metaphor co-occur in a large corpus.

The embodied component, on the other hand, is operationalised by ease of simulation (EoS), a new normed metric that quantifies how easy it is to come to a mental representation of a metaphor (Liu, Connell, & Lynott, 2016). This metric was extracted using a principle components analysis from the ratings on the sensibility (how much sense the sentence make if read or heard), usability (how easy it is to use it in conversation or writing), and imaginability (how easy it is to describe the concept) in the norming study. When combined into a single measure, these ratings offer a proxy for how easy it is to simulate a concept. That is, if people find it easy to make sense of and use the metaphor, as well as imagine the concept, they would find it easy to generate embodied simulations.

Although LDF and EoS may correlate to a certain degree, we expected them to play distinctive roles in metaphor processing after the common variance between them is removed. Both variables would independently affect the acceptability rate and speed of metaphor processing. Specifically, increase in both variables would make it more likely and faster for people to accept a metaphor, and meanwhile slower to reject the very metaphor. More crucially, participants performed one of two tasks: a sensibility judgement task (Experiment 1), which required relatively shallower processing because participants made only yes/no response; or an interpretation-generation task (Experiment 2), which was deeper because they specified the meaning verbally. We predicted that EoS would play a greater role in deep processing or when people accepted metaphors because it indicated successful simulation, while LDF should play a larger role in shallow processing especially when people rejected a metaphor as not being sensible.

Experiment 1: Sensibility Judgement Task

Method

Participants Twenty-eight participants took part (five male and 23 female), all of whom were students at Lancaster University and native speakers of English with mean age of 19.1 years (SD = 1.1). The sample size was determined beforehand to achieve the same level of variability as the conceptual combination study in Connell and Lynott (2013).

Materials We used a total of 452 metaphor sentences taken from Liu et al., (2016). All sentences took the form “Noun can be 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 that were capable of eliciting metaphoric meanings that vary independently on the following two dimensions (see examples in Table 1):

Theoretically and operationally distinct, and (b) are able to independently account for speed and accuracy performance in metaphor processing.
Procedure Participants read one sentence in each trial and decided whether or not the sentence made sense. All trials had the same structure. Participants first saw a fixation cross for 1000 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 onscreen 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. 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 decisions (“yes” to accept the metaphor as sensible; or “no” to reject the metaphor as nonsensical), and the response time in milliseconds (RT) from onset of the adjective were recorded as dependent variables.

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 < 200ms). Furthermore, individual trials with RT more than 3SD from each participant’s mean per response decision were removed as outliers: 1.33% of “yes” responses and 2.20% of “no” responses.

Among 3105 valid trials, 1413 (45.51%) were accepted as sensible (“yes” responses) and 1692 (54.59%) were rejected as nonsensical (“no” responses). Logistic regression showed evidence for net suppression. This means that while metaphors with the “yes” response had a higher mean LDF than those with the “no” response (i.e., the higher LDF was, the more likely it should be to respond “yes”), the effect of LDF in the mixed effects logistic regression turned out negative (i.e., the higher LDF was, the less likely to respond “yes”). This suggested that LDF enhanced the effect of
Table 3: Results from mixed effects regressions of RT per response in Experiment 1.

<table>
<thead>
<tr>
<th>Response (&quot;yes&quot;)</th>
<th>Variable</th>
<th>$\beta$</th>
<th>95% CI</th>
<th>df</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept</td>
<td>Intercept</td>
<td>1172.17</td>
<td>[1051.79, 1292.53]</td>
<td>25.7</td>
<td>19.09</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>EoS</td>
<td>-134.88</td>
<td>[-161.98, -107.79]</td>
<td>357.3</td>
<td>-9.76</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>LDF</td>
<td>-27.52</td>
<td>[-53.10, -1.93]</td>
<td>248.1</td>
<td>-2.11</td>
<td>.036</td>
</tr>
<tr>
<td></td>
<td>EoS * LDF</td>
<td>-13.00</td>
<td>[-37.91, 11.91]</td>
<td>394.6</td>
<td>-1.02</td>
<td>.307</td>
</tr>
<tr>
<td>Reject (&quot;no&quot;)</td>
<td>Intercept</td>
<td>1155.41</td>
<td>[1010.94, 1299.88]</td>
<td>26.7</td>
<td>15.68</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>EoS</td>
<td>72.28</td>
<td>[47.27, 97.29]</td>
<td>323.8</td>
<td>5.66</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>LDF</td>
<td>17.73</td>
<td>[-6.40, 41.83]</td>
<td>189.1</td>
<td>1.44</td>
<td>.152</td>
</tr>
<tr>
<td></td>
<td>EoS * LDF</td>
<td>23.33</td>
<td>[0.51, 46.14]</td>
<td>295.3</td>
<td>2.00</td>
<td>.046</td>
</tr>
</tbody>
</table>

EoS by explaining the residuals of EoS rather than the variance of response decision. 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 correlating at $r = .27$) by orthogonalising the variables using a principal components analysis with varimax rotation and Kaiser normalization on a covariance matrix (Glantz & Slinker, 2000). What this did was to create two new orthogonal variables ($r = 0$), each correlating highly with one original variable ($r = .99$). These two new variables were thus named orthogonal EoS and orthogonal LDF. We re-ran the logistic regression with them and obtained results as follows (further analyses all used orthogonal variables).

The logistic regression with orthogonal variables showed that both variables had a positive effect on response decision. As the EoS increased by one unit, the odds to accept a metaphor as sensible increased 3.42 times ($z(3101) = 25.03, p < .001, \beta = 1.23$). As the LDF increased, the odds to accept a metaphor increased with a marginally significant effect ($1.084$ times, $z(3101) = 1.88, p = .06, \beta = 0.08$).

RT was also analysed using orthogonal variables. Table 2 shows full results of the omnibus analysis. Overall, EoS had a positive effect on RT ($M = 1139$ms, $SD = 587$ms), and it critically interacted negatively with response decision, suggesting that the direction of the EoS effect differed by the response type, and was greater for “yes” than “no” responses. 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 analysed their RTs separately. Results are given in Table 3. For “yes” responses (i.e. metaphors that were accepted as sensible; RT: $M = 1150$ms, $SD = 589$ms), the easier a metaphor was to simulate, the less time people took to accept it as sensible. Also, the more often the words in a metaphor co-occurred in language, the faster people were to accept it as sensible.

For “no” responses (RT: $M = 1114$ms, $SD = 603$ms), the effects ran in the opposite direction (Table 3). As predicted, people were faster to reject metaphors that were normally regarded as difficult to simulate. Furthermore, it interacted with LDF positively, such that the effect of EoS was reduced when LDF was low (the $\beta$ for the interaction term was positive).

**Experiment 2: Interpretation Generation Task**

**Method**

In this study, we asked 40 participants (native speakers of English, 11 males, age: $M = 19.65, SD = 2.08$) to judge whether they could think of a meaning for the metaphoric sentences instead of judging their sensibility. The procedure was the same as Experiment 1, except that participants needed to provide their interpretation of the sentences after they responded “yes”. To reduce the possibility of fatigue, each participant saw half (56-57) of the items of Experiment 1.

**Results and Discussion**

Data cleaning was performed as in Experiment 1. Furthermore, we also identified accept (“yes”) trials with invalid interpretations (e.g., blank, “I don’t know”). Two participants were excluded from analysis for providing more than 50% invalid interpretations. Amongst the remaining participants, 2.33% of interpretations were identified as invalid. For individual trials, 2.10% of “yes” responses and 2.00% of “no” responses were identified as outliers.

Among 2103 valid trials, 1302 (61.91%) were accepted as interpretable whereas 801 (38.09%) were rejected as uninterpretable. The logistic regression showed that both EoS and LDF had positive effects on response decision. For every unit of increase in EoS, the odds of accepting a metaphor as interpretable increased 2.826 times ($z(2099) = 17.49, p < .001, \beta = 1.04$); and for every unit of increase in LDF, the odds of accepting a metaphor increased 1.286
Based on metaphor varied by task: the odds to accept a metaphor increased 3.24 times in deep interpretation generation compared to shallow sensibility judgement; \( z(5200) = 3.39, p = .001, \beta = 1.18 \). As for response time, EoS interacted with task, showing that the effect is larger for deep interpretation generation than for shallow sensibility judgement, as predicted (“Yes”: \( t(2461.2) = -6.39, p < .001, \beta = -4.029 \), “No”: \( t(1936.6) = 3.37, p = .001, \beta = 145.00 \). The interaction between EoS and LDF also varied across tasks \( t(1933.3) = -2.10, p = .036, \beta = -81.67 \), which was larger for shallow than deep processing.

### General Discussion

Our goal in taking this grounded approach was to move the investigation of metaphor processing beyond the traditional factors of familiarity, conventionality, and aptness, which – while having a long history of use – have been increasingly criticised for theoretical and operational problems that limit their utility in explaining what makes one metaphor easier to understand than another. Indeed, our study generated complex results that could not be accounted for by traditional theories with single factors.

The current study shows for the first time that both linguistic component (based on linguistic distributional frequency) and embodied component (based on ease of simulation norms) affect metaphor comprehension independently. Their roles are statistically distinct from each other.
other after we managed to remove their common variance with a principle components analysis. Whereas ease of simulation often had a large effect overall and was more prominent for the “yes” response because that was when simulation was eventually successful; linguistic distributional frequency represents a relatively coarse-grained, but nonetheless highly useful, approximation of whether a particular source and target have previously formed a metaphor. It informs people’s responses, not only making acceptance more likely and faster when the words are likely to constitute a meaningful representation, but also flags up potentially unsuccessful simulation to be rejected right away without further processing when distributional frequency is low. Our findings are consistent with the grounded views which suggest that conceptual representation relies on both embodied simulation and linguistic distributional pattern ([Barsalou et al., 2008; Connell & Lynott, 2014; Louwerse & Jeuniaux, 2008).

However, against predictions, the main effect of linguistic distributional frequency did not differ between shallow and deep processing tasks according to the cross-experiment analysis (i.e., linguistic distributional frequency itself did not interact with task). This null effect could be because the tasks disincentivised using the linguistic shortcut by allowing people as much time as needed to make a response. That is, they had unlimited time resource to form a mental representation using the embodied component. In future research, we will impose time constraints on the task in order to further examine the utility of linguistic distributional information during metaphor processing, and provide an additional test of the linguistic shortcut hypothesis.

References


