Brain & Language 174 (2017) 61–71

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Finding features, figuratively

Sarah H. Solomon [*], Sharon L. Thompson-Schill

University of Pennsylvania, Department of Psychology, 425 S. University Ave., Philadelphia, PA 19104-6018, United States

ARTICLE INFO

Article history:
Received 22 June 2016
Revised 10 April 2017
Accepted 2 July 2017

Keywords:
Figurative language
Metaphors
Conceptual flexibility
Object properties
Cognitive control
fMRI
Prefrontal cortex

ABSTRACT

Object concepts refer to unique clusters of properties that can be selectively activated or inhibited depending on what information is currently relevant. This conceptual "stretching" enables limitless new meanings to be generated, and figurative language provides a useful framework in which to study this conceptual flexibility. Here we probe the cognitive and neural mechanisms underlying the comprehension of novel metaphors as a means of understanding the conceptual flexibility inherent to language processing more generally. We show that novel metaphor comprehension involves the activation or inhibition of conceptual properties that are either relevant or irrelevant to the metaphor, and that left inferior frontal gyrus is recruited in this process, supporting a role for this region in the fine-tuning of conceptual meaning. Our results are consistent with a flexible, compositional account of conceptual structure in which semantic control mechanisms operate over conceptual properties during figurative language comprehension in order to create context-dependent meaning.

1. Introduction

Concepts are stretchy. *Blue* can be true of skies, forget-me-nots, political districts, and moods. Colorful insects, fluttery feelings in your stomach, and ballerinas can all be *butterflies*. This conceptual flexibility contributes to conceptual generativity, such that limitless new meanings can be created. Metaphors (e.g., "She is feeling blue", "The ballerina is a butterfly") are extreme examples of conceptual flexibility, in which a single concept appears to take on multiple, distinct meanings. However, a single word can take on a theoretically infinite number of meanings even in the absence of figurative language. Because meaning is always underdetermined by linguistic content (Carston, 2010), fine-tuning of meaning may be required for *every* utterance (c.f. Relevance Theory, Carston, 2010; Wilson & Carston, 2007), such that metaphorical meaning comprehension differs in degree (but not in kind) from literal meaning comprehension. Therefore, metaphorical language is a useful tool with which we can explore how concepts are stretched to encompass new meanings in language more generally.

Object concepts refer to unique clusters of information (e.g. color, taste, texture), but the context in which a concept is represented will determine which information is relevant. The concept RAISIN may include the properties *dried*, *wrinkled*, *sweet*, *chewy*, and *purple*, but not all of these properties are relevant to each raisin situation. While cooking, the properties *sweet* and *chewy* are relevant, whereas the properties *purple* and *wrinkled* are relevant while painting. This implies a process by which information contained in the RAISIN concept is adjusted to highlight the relevant information in each case, and empirical studies support this claim (e.g., Yee, Ahmed, & Thompson-Schill, 2012; Pecher, Zeelenberg, & Barsalou, 2004). This notion is central to theories of lexical pragmatics, such as Relevance Theory, which claim that the interpretation of single words often involves a meaning adjustment, in which concepts are either "narrowed" or "broadened" (Carston, 2010; Wilson & Carston, 2007). This constant fine-tuning of meaning occurs for action concepts (e.g. DANCE) as well as property concepts (e.g. FRESH). For example, the meaning of "dance" in the utterance "Let's dance" takes different forms in the contexts of a ballroom, a nightclub, or a sidewalk stroll (see Carston, 2010). Similarly, "fresh" conveys different information when paired with "vegetable", "shirt", and "idea" (Murphy & Andrew, 1993). However, pragmatic theories of meaning adjustment propose that the conceptual narrowing and broadening implicated in the comprehension process is captured by the formation of a new, ad hoc concept; this ad hoc concept is a separate entity from the original concept (e.g., DANCE vs. DANCE*). Our approach does not propose the formation of a new concept, but rather the flexible stretching of the concept as it is accessed in memory. We will describe our approach in more detail in order to motivate our empirical predictions.

A theory of conceptual adjustment necessarily includes a theory of conceptual structure. The most basic division in this set of

---

* Corresponding author.
E-mail address: sarahsol@sas.upenn.edu (S.H. Solomon).

http://dx.doi.org/10.1016/j.bandl.2017.07.002
0093-934X/© 2017 Elsevier Inc. All rights reserved.

theories is between those that posit concepts to be "atomic" or unstructured (e.g., Carston, 2010; Fodor & Lepore, 1998), and those that posit concepts to be "decompositional", that is, composed of simpler elements such as properties, features, or relations (e.g., Jackendoff, 2002; McRae, de Sa, & Seidenberg, 1997; Tyler & Moss, 2001). Here we adopt a decompositional account of conceptual structure, and see whether such an approach can provide insights into the cognitive and neural mechanisms underlying metaphor comprehension. We also adopt a radical view on conceptual representation that departs starkly from most theories relevant to the current discussion. The traditional way to frame figurative language comprehension is to ask how figurative meaning emerges, or is constructed, from literal meaning; what is meant by "literal" is usually a stable, context-free representation (e.g., Weiland, Bambini, & Schumacher, 2014) stored in memory. However, Searle (1978) argued that there is no context-free meaning of a sentence, and "literal" meaning only exists relative to a set of background assumptions. Additionally, recent empirical evidence suggests that the brain may represent concepts not as static structures, but as context-dependent, flexible representations (for a review, see Yee & Thompson-Schill, 2016). If one rejects the existence of stable conceptual representations, the notion of "literal" meaning becomes incongruous, since concepts have no default, context-free state. We therefore reframe the question of metaphor comprehension, and ask not how figurative meaning is derived from a literal concept, but rather how the structure of a concept is flexibly adjusted to meet the current referential and contextual demands. We therefore also reject the traditional distinction between "literal" and "figurative" language comprehension, and subscribe to what Wilson and Carston (2007) refer to as a radical, unified approach to pragmatics: The same meaning adjustment process is involved in all types of language comprehension, ranging from literal to figurative, with no clear division between them. Our approach thus assumes that (1) concepts are decompositional, (2) concepts are neither stable nor context-free, but rather flexible and context-dependent, and (3) the mechanisms of conceptual adjustment involved in metaphor comprehension are the same as those required in language comprehension more generally.

Here our questions concern the mechanism by which concepts are "stretched" in the process of metaphor comprehension. Prominent theories of metaphor processing can be roughly classified as comparison models (Ortony, 1979), in which properties are matched between the tenor and vehicle; domain-interaction models (Tourangeau & Rips, 1991; Tourangeau & Sternberg, 1982), in which properties of the vehicle are transformed into properties appropriate for the tenor's domain; class-inclusion models (Gernsbacher, Keysar, Robertson, & Werner, 2001; Glucksberg, 2003; Glucksberg & Keysar, 1990; Glucksberg, McGlone, & Manfredi, 1997), in which properties of the vehicle are used to create a superordinate category of which the tenor is asserted to be a member; and theories in the field of lexical pragmatics, such as Relevance Theory (Carston, 2010; Wilson & Carston, 2007), in which inferential processes work to narrow or broaden a concept and create a new, ad hoc category. Our current approach is that metaphorical meaning emerges as a result of conceptual "stretching", which encompasses the "narrowing" and "broadening" of the relevance theory account, but which operates over the set of properties that compose the concept of interest. Though we are agnostic regarding which theory most accurately predicts *why* a specific set of properties plays a role in the meaning of a particular metaphor, it seems clear that, for any particular metaphor, some properties are more relevant to metaphorical meaning than others.

Most, if not all, of these theories would propose that the metaphor comprehension process involves activating relevant information and inhibiting irrelevant information. Though the kind of information under discussion differs according to each theory, this general finding has been supported in many empirical studies. Property-verification paradigms have supported the class-inclusion model by revealing that irrelevant basic-level properties (e.g., *fins*) are suppressed whereas relevant superordinate properties (e.g., *tenacious*) are activated subsequent to metaphor processing (e.g., "The lawyer is a shark"; Gernsbacher et al., 2001). Cross-modal priming paradigms have shown that irrelevant superordinate information (e.g., *buildings*) is inhibited and relevant distinct properties (e.g., *tall*) are active subsequent to metaphor processing (e.g., "The pine trees were skyscrapers"; Fernandez, 2007). Given that our account of metaphor comprehension involves the adjustment of structured concepts, rather than the creation of superordinate or ad hoc categories, we aim to replicate these findings using only basic-level properties. For example, in the metaphor "The ballerina is a butterfly", properties of BUTTERFLY that are relevant might include *delicate* and *colorful*, whereas other properties of BUTTERFLY such as *winged* are most likely irrelevant. The observed patterns of property inhibition should reflect the reshaping of the concept to match the metaphorical context.

In Study 1, we use a property-verification paradigm to test our prediction that semantic control mechanisms operate over basic-level properties during metaphor comprehension. We will restrict our analysis to nominative metaphorical assertions of the form *The X is a Y*, in which X and Y are considered the *tenor* and the *vehicle* of the metaphor, respectively. We specifically predict that, after reading a metaphor (e.g., "The ballerina is a butterfly"), it should be easier to verify that metaphor-relevant properties are true of the vehicle concept (*delicate* + BUTTERFLY), whereas it should be harder to verify that metaphor-irrelevant properties are true of the concept (*winged* + BUTTERFLY). Given that each metaphor involves a unique pairing of concepts, each with its own cluster of properties that might interact with each other in different ways, it is reasonable to assume that conceptual flexibility demands will differ across metaphors. We thus extracted a measure we name the P-index that reflects the particular inhibition demands for each of our metaphors.

Our next question concerns the neural mechanisms supporting this process of conceptual adjustment. In a recent meta-analysis, Bohrn, Altmann, and Jacobs (2012) report that figurative language recruits a bilateral frontotemporal network across studies, with a bias towards the left hemisphere. Consistently implicated regions include left and right inferior frontal gyrus, left temporal lobe, bilateral medial frontal gyri, and left amygdala; however, the neural structure with the largest effect distinguishing between figurative and literal language processing is the left inferior frontal gyrus (LIFG). Many neuroimaging studies have shown that LIFG is recruited for metaphor comprehension (e.g., Bambini, Gentili, Ricciardi, Bertinetto, & Pietrini, 2011; Cardillo, Watson, Schmidt, Kranjec, & Chatterjee, 2012; Eviatar & Just, 2006; Lee & Dapretto, 2006; Rapp, Leube, Erb, Grodd, & Kircher, 2004, 2007; Stringaris, Medford, Giampietro, Brammer, & David, 2007), but the attempts to link this neural region with a cognitive mechanism remain unsatisfying. Increased LIFG activation is often interpreted to reflect the suppression of the literal meaning, or selection or retrieval of the metaphorical meaning of a word (Cardillo et al., 2012; Lee & Dapretto, 2006). Because LIFG is thought to contribute to the selection of task-appropriate representations by biasing competitive interactions between incompatible representations (Thompson-Schill, D'Esposito, & Kan, 1999; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Thompson-Schill et al., 1998), and further, because LIFG has been implicated in homonym processing (Bedny, McGill, & Thompson-Schill, 2008), this interpretation would seem quite plausible. However, there is an important difference: In the case of homonyms, there are two (or more) pre-existing meanings that one must select between, whereas for novel metaphors, one *creates* the new meaning during the comprehension process. Since the figurative meaning is a function of the original concept, it seems implausible that suppressing the concept in its entirety would benefit metaphor comprehension. Fernandez (2007) explains this difference in terms of meaning disambiguation versus meaning construction, and uses a cross-modal lexical priming paradigm to highlight differences in the time courses of property suppression between the two cases. Rather than propose two separate mechanisms, we suggest that both involve the same inhibitory mechanism that operates over different kinds of information. In the case of homonyms, LIFG may operate over whole concepts, whereas in the case of metaphors, LIFG may operate over properties. Cognitive neuroscience approaches to conceptual processing suggest that neural representations of object concepts refer to unique combinations of features (Coutanche & Thompson-Schill, 2014; Lambon Ralph, Sage, Jones, & Mayberry, 2010), and LIFG appears to be involved in the modification of representations to reflect feature-specific changes (Hindy, Solomon, Altmann, & Thompson-Schill, 2015). Thus, the claim that LIFG supports the inhibition of conceptual properties during metaphor comprehension appears to be a neurally plausible one.

Bambini et al. (2011) attempted to link neural regions implicated in figurative language processing with specific sub-mechanisms proposed by pragmatic and cognitive theories of metaphor, and proposed that recruitment of bilateral IFG during figurative language processing reflects the context-sensitive activation of the conceptual system, that is, the fine-tuning of lexical meaning. However, at this point this claim is purely speculative, as the dependence of LIFG activation on levels of conceptual adjustment has not been directly tested. In Study 2, our goal is to explicitly link the conceptual flexibility demands engendered by each metaphor with levels of LIFG activation. We use fMRI to measure levels of LIFG activity during comprehension of each of our nominal metaphors relative to literal controls, and determine whether our index of conceptual adjustment (i.e. P-index extracted from Study 1) predicts the increase in LIFG response.

2. Study 1 (Behavioral)

2.1. Methods: Study 1

2.1.1. Participants

Forty-seven subjects from the University of Pennsylvania (31 female; mean age = 23.2 years, SD = 4.1) contributed data to this study, and were compensated $10/h for their time. Written consent was obtained for all participants in accordance with the University of Pennsylvania IRB.

2.1.2. Stimuli

All sentences were of the form *"The X is a Y"* (see Appendix for full list of experimental materials). We constructed 48 experimental pairs of metaphorical (e.g., "The train is a worm") and literal (e.g., "The creature is a worm") sentences, such that the vehicle in the metaphor (e.g., "worm") was a concept included in the McRae norms (McRae, Cree, Seidenberg, & McNorgan, 2005; a database that includes object concepts along with their most commonly reported properties); this word appeared in the metaphor (MET) and literal (LIT) version of each pair. The only difference between MET and LIT sentences in each pair was that the tenor (e.g., "train") of the metaphor was swapped for a different word (e.g., "creature") in the LIT version. An additional 60 filler sentences were constructed, half metaphorical and half literal; these were not constrained by the McRae norms.

For each experimental MET sentence, two properties of the vehicle (*worm*) were selected from the McRae database: one property was relevant to the metaphor (*slithers*), and one was irrelevant (*slimy*). We collected online survey data from a separate group of subjects ($N = 86$) to confirm that people's interpretations of the metaphors were consistent with our chosen relevant (REL) and irrelevant (IRR) properties. Subjects were asked to read each metaphor and rate the extent to which either the REL or IRR property was relevant to the meaning of the metaphor on a seven-point scale ranging from "Not at all relevant" to "Extremely relevant." REL properties ($M = 5.30$) were rated as more relevant to metaphorical meaning than IRR properties ($M = 2.13$; $t(47) = 13.73$, $p < 0.0001$). Across the 48 items, these properties were matched for production frequency (REL mean = 14.92, IRR mean = 14.58, $t(47) = 0.23$, $p > 0.8$). For each filler sentence, a property was selected that was not true of the vehicle (or last term in literal sentences).

Each participant saw either the MET or LIT version of each of the 48 experimental pairs, and saw either the relevant (REL) or irrelevant (IRR) property, resulting in a 2 (sentence-type) × 2 (property-type) design. All participants saw the same 60 filler trials.

2.1.3. Procedure and analysis

The experiment consisted of two tasks: (1) a property-verification task, and (2) a semantic 1-back task. The property-verification task provided the main data of interest; the 1-back task was an orthogonal task used to ensure attention during the experiment. The experiment was divided into four blocks. Participants completed a short practice run before starting the main experiment.

Each trial included a presentation of a sentence followed by a property-verification task (Fig. 1). Each trial began with 1000 ms of fixation, a 3000 ms presentation of a cue (described below), and a 3000 ms presentation of the sentence, displayed at the center of the screen. The screen was cleared for 250 ms, followed by 1750 ms of fixation; at this point the last term in the preceding sentence (i.e., the "vehicle" of the MET), appeared above the central fixation cross, and a property (REL or IRR) appeared below the fixation cross. The object term and the property remained on screen for 3000 ms. While the words were on screen, participants were instructed to make a keyboard response as to whether the property was true ("j" key) or false ("f" key) of the object. If a response was not made within the 3000 ms in which the words were on screen, no response was recorded and that trial was considered incorrect. The screen was cleared for 2000 ms before the next trial.

For all experimental trials, the property in the property-verification task was always true of the object. We thus added the 60 filler trials, which served two purposes: (1) the property never matched the object in the property-verification task following filler trials, resulting in an approximately equal distribution for "yes" and "no" correct responses on this task, and (2) a subset of these filler trials (24) were used in the orthogonal 1-back task.

Since the property-verification task could theoretically be completed without paying attention to the sentences, an orthogonal task was also included. This was a 1-back semantic task in which participants were to respond whether the sentence was similar in meaning to the sentence that appeared on the previous trial (e.g., "Their relationship is a see-saw"/"Her emotions are a rollercoaster"). Each sentence was preceded by either a pound sign (#) or question mark (?); the question mark cue signified that, in addition to the property-verification task, an additional response was required for the 1-back semantic task. There were 24 cued trials (6 per block); cued sentences could either be literal or metaphorical. Participants were instructed to make a response while the sentence was displayed as to whether the meaning of the current sentence was similar ("j" key) or not similar ("f" key) to the preceding sentence. The essence of this task is similar to that used by Bambini et al. (2011), in which subjects judged which of two adjectives better matched a previous metaphorical or literal passage.

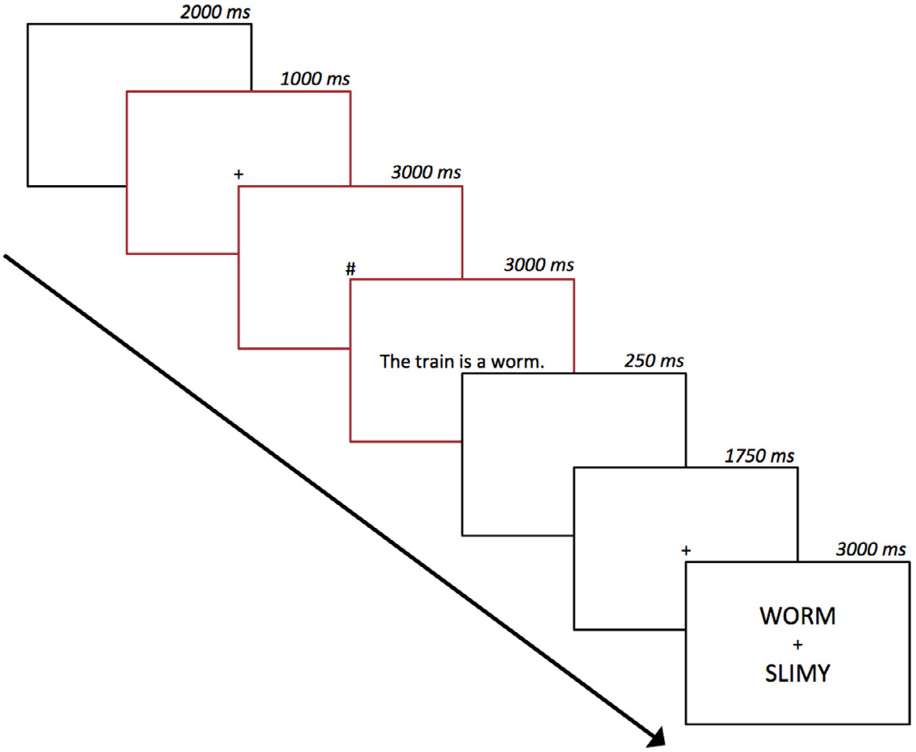

Fig. 1. Experimental procedure. Trials in Study 1 included all components; trials in Study 2 (fMRI) consisted only of the components outlined in red (in Study 2 fixation times were jittered between 3 and 12 s).

The orthogonal task here and the one used by Bambini et al. (2011) both require extraction of sentence meaning, though our task was more challenging in that we asked subjects to judge the similarity of an entire *sentence* with the preceding stimulus, rather than a single word. We analyzed the experimental data from Study 1 with the full sample, and also with a restricted sample based on performance on this orthogonal task (accuracy > 70%; $N = 32$), but the pattern was the same in both.

2.2. Results: Study 1

In the full sample ($N = 47$), accuracy on the property-verification task was high for both experimental (94.9%, SD = 6.5%) and filler (95.5% SD = 6.6%) trials; accuracy on the orthogonal 1-back semantic task was lower (76.6%, SD = 21.7%), but significantly above chance ($t(46) = 8.41$, $p < 0.0001$). Filler trials were excluded for all subsequent analyses. For each subject, we compared accuracy on the property-verification task between the four conditions (LIT-REL, LIT-IRR, MET-REL, MET-IRR). A 2×2 within-subject ANOVA revealed an interaction between sentence-type and property-type ($F(1,46) = 6.43$, $p = 0.015$) a main effect of property-type ($F(1,46) = 7.06$, $p = 0.011$), and no main effect of sentence-type ($F(1,46) = 1.81$, $p = 0.19$). These results are driven by lower accuracy in the IRR-MET condition: paired t-tests reveal that accuracy for IRR-MET trials was significantly lower than REL-MET ($t(46) = 3.36$, $p = 0.002$) and IRR-LIT ($t(46) = 2.65$, $p = 0.01$) trials. These data reflect the suppression of irrelevant properties subsequent to metaphor processing.

To explore whether reaction time (RT) for the property-verification task was influenced by the type of sentence that preceded it (LIT vs. MET), mean RTs for the four conditions were calculated for each subject, after removing incorrect trials and z-scoring RTs within subject. In a subject wise analysis, a 2×2 within-subject ANOVA revealed a marginal interaction between sentence-type and property-type ($F(1,47) = 2.93$, $p = 0.09$), in addition to a main effect of property-type ($F(1,47) = 22.84$, $p < 0.0001$). For the subjects that performed well on the orthogonal task ($N = 32$), this interaction was significant ($F(1,31) = 7.81$, $p = 0.009$), suggesting that performance on our orthogonal task did reflect some aspects of metaphor processing. This is the only statistical difference between our samples, so we included all 47 subjects in the remaining analyses.

An item analysis (Bedny, Aguirre, & Thompson-Schill, 2007) revealed a significant interaction between sentence-type and property-type ($F(1,47) = 7.84$, $p = 0.007$), a main effect of property-type ($F(1,47) = 8.45$, $p = 0.005$), and a marginal effect of sentence-type ($p = 0.08$). Paired t-tests revealed an effect of property-type after reading metaphors ($t(47) = 4.32$, $p < 0.0001$) but no effect of property-type after reading literal sentences ($t(47) = 0.92$, $p = 0.36$). This result seems to be driven by the RTs to the metaphor-irrelevant properties in the MET versus LIT conditions ($t(47) = 3.33$, $p = 0.002$); no significant difference between RTs to the metaphor-relevant properties was observed between MET and LIT conditions ($t(47) = 0.90$, $p = 0.37$). The mean RTs (unstandardized) for each condition are shown in Fig. 2. Relative to reading a literal sentence, reading a metaphorical sentence resulted in faster RTs to the REL property, and slower RTs to the IRR property.

Findings thus confirmed that, after reading a metaphorical sentence, subjects were faster to verify properties relevant to the metaphor, and slower to verify properties irrelevant to the metaphor. This suggests a dynamic activation or suppression of conceptual properties during metaphor comprehension.

2.2.1. P-Index

For each item, we extracted a measure of the inhibition demands during comprehension ((MET-IRR RT – MET-REL RT) – (LIT-IRR RT – LIT-REL RT)), which we refer to as the P-index. By

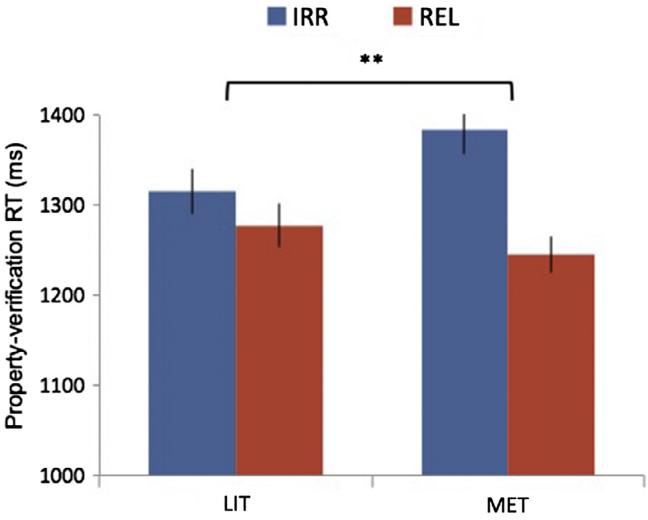

Fig. 2. RT results from Study 1: Mean reaction time to verify metaphor-relevant (REL) and metaphor-irrelevant (IRR) properties in both the literal (LIT) and metaphorical (MET) conditions Only correct trials were analyzed. Error bars represent standard errors of the mean.

subtracting out the LIT RT-effect, we removed any potential confounds between the object and the property (e.g., strength of association, reading times). That is, the P-index reflects the effect of metaphor on property activation or suppression. Though we calculate the P-index based on two particular properties (REL and IRR) of the vehicle concept, we consider this measure to be specific to the *metaphor* (i.e., the interaction of the tenor and vehicle concepts), and not specific to the vehicle concept alone or the properties we happened to use. The goal of Study 2 is to determine whether the P-index predicts activation in LIFG during metaphor comprehension.

3. Study 2 (fMRI)

3.1. Methods: Study 2

3.1.1. Participants

16 subjects from the University of Pennsylvania (8 female; mean age = 22.1 years, SD = 3.6) participated in this study, and were compensated $20/h for their time. All subjects were included in the analyses. Sample size was determined based on typical sample sizes in neuroimaging literature, and was decided upon prior to data collection. Written consent was obtained for all participants, in accordance with the University of Pennsylvania IRB. No subjects in Study 1 participated in Study 2.

3.1.2. Stimuli and procedure

The same experimental and filler sentences used in Study 1 were used in Study 2. However, in Study 2, each trial only consisted of a sentence preceded by fixation and a cue (Fig. 1); no properties were shown. Trials were separated by 3–12 s of fixation, and were pseudo-randomized for each subject within each of four scanning runs. Participants were instructed to perform the same 1-back sentence similarity task used in Study 1. Participants performed well on the semantic 1-back task with a mean accuracy of 85.5% (SD = 11.4). We will refer to the cued trials in the 1-back task as catch trials.

Following the sentence comprehension task, subjects completed a 10-min Stroop color-word interference task that was used to assess the sensitivity of our LIFG ROI to semantic conflict (Hindy, Altmann, Kalenik, & Thompson-Schill, 2012; Hindy et al., 2015; Milham et al., 2001). Subjects were instructed to press the button on the response pad (blue, yellow, green) that corresponded to the typeface color of the word displayed on the screen. Stimuli included four trial types: response-eligible conflict, response-ineligible conflict, and two groups of neutral trials. In response-eligible conflict trials, the word presented on the screen was a color term that matched a possible response (*"blue," "yellow," "green"*), but mismatched its typeface color (blue, yellow, or green). In response-ineligible conflict trials, the color term was not a possible response (*"orange," "brown," "red"*), and mismatched the typeface color. In neutral trials, non-color terms were used (e.g., *"farmer," "stage," "tax"*); these neutral trials were intermixed with the aforementioned conflict trials across a total of four blocks. Both response-eligible and response-ineligible conflict trial types have been demonstrated to induce conflict at nonresponse levels, and response-eligible conflict trial types also induce conflict at the level of motor response (Milham et al., 2001). To optimize power, we collapsed across these two types of conflict trials.

From the McRae norms, we extracted the production frequency and distinctiveness of the REL and IRR properties. In addition, we used Latent Semantic Analysis (http://lsa.colorado.edu/) to obtain pairwise comparison scores for each object concept with its REL and IRR property. We collected salience measures for the REL and IRR properties in a separate online survey using Amazon Mechanical Turk (AMT) (*N* = 77). Participants were shown the object concept and either the REL or IRR property and were asked, "When you think of [object], how likely are you to think of [property]" on a 7-point scale ranging from "Not likely at all" to "Extremely likely". We also collected familiarity ratings for our metaphors in a separate AMT survey (*N* = 51). Subjects were asked to read each metaphor and rate how familiar it is on a seven-point scale ranging from "Not at all familiar" to "Extremely familiar." Mean familiarity was 3.41 (*SD* = 0.88), and ranged from 1.90 ("The clam is a zebra") to 5.16 ("The rooster is an alarm clock"). We included the production frequency, distinctiveness, LSA scores, salience, and familiarity measures as covariates in our model of the fMRI data.

3.1.3. Data acquisition

Structural and functional data were collected on a 3-T Siemens Trio system and 32-channel array head coil. Structural data included axial T1-weighted localizer images with 160 slices and 1 mm isotropic voxels (TR = 1620 ms, TE = 3.87 ms, TI = 950 ms, FOV = 187 × 250 mm, Flip Angle = 15°). Functional data included four acquisitions of echo-planar fMRI performed in 42 axial slices and 3 mm isotropic voxels (TR = 3000 ms, TE = 30 ms, FOV = 192 × 192 mm, Flip Angle = 90°).

3.1.4. Data analysis

Image preprocessing and statistical analyses were performed using FMRIB Software Library (FSL). Before preprocessing, the four functional runs were concatenated into one time series for each subject. Functional data were processed using FEAT (FMRI Expert Analysis Tool) Version 6.00. Preprocessing included motion correction using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002), interleaved slice timing correction, spatial smoothing using a Gaussian kernel of FWHM 5 mm, grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor, and highpass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 50.0 s). Time-series statistical analysis was carried out using FILM with local autocorrelation correction (Woolrich, Ripley, Brady, & Smith, 2001). Each experimental sentence was modeled in a separate covariate as a 3 s boxcar function convolved with a double-gamma hemodynamic response function. Filler trials, catch trials (including all time points in which a response was made), and run covariates were added to the model,

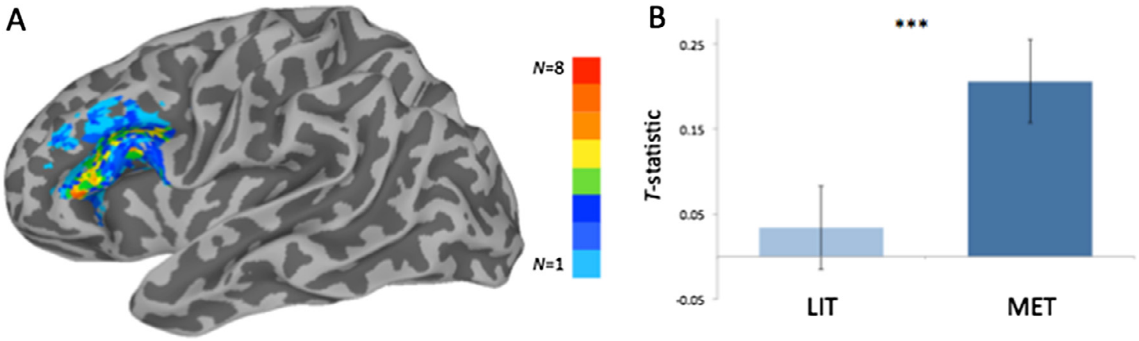

Fig. 3. Increased LIFG response to metaphors. (A) An overlay of the subject-specific ROIs in LIFG, defined as the 100 voxels within our anatomical mask most responsive to all sentences versus baseline (highest overlap, N = 8). (B) Within these subject-specific ROIs, there was greater activation for the MET sentence than the LIT sentence within each item pair ($t(47) = 3.52$ $p = 0.001$). Error bars represent standard error of the difference.

with motion outliers as covariates of no interest. Functional data were then normalized to a standard template in Talairach space.

Analyses focused on an ROI in left inferior frontal gyrus (LIFG). The ROI was anatomically constrained based on probabilistic anatomical atlases (Eickhoff et al., 2005) transformed into Talairach space, and was defined as the combination of pars triangularis (Brodmann area 45), pars opercularis (Brodmann area 44), and the anterior half of the inferior frontal sulcus. On average, these anatomical ROIs comprised 892 voxels. Within these anatomical boundaries, subject-specific ROIs were created by restricting the ROI to the 100 voxels that had the highest *t*-statistics in a subject-specific contrast of all sentences > fixation in the sentence comprehension task. In our item analyses, each sentence was modeled individually and contrasted to a baseline of filler sentences, resulting in a measure of neural response (*t*-statistic) for each item. We then calculated the difference in response to the metaphor and literal version of each item, to obtain a measure comparable to the standardized P-index in Study 1. All statistical tests were assessed at the two-tailed $p < 0.05$ level of significance.

3.2. Results: Study 2

3.2.1. Increased LIFG activation for metaphors

We compared activation to LIT and MET sentences to see if we replicated the finding of increased response to metaphors in LIFG. For each subject, we extracted the mean *t*-statistic for each item compared to the filler baseline within the 100-voxel LIFG ROI; we collapsed across subjects to get the mean LIFG response to the literal and metaphorical version of each of the 48 items (Fig. 3). Metaphorical sentences resulted in significantly greater LIFG activation than literal sentences ($t(47) = 3.52$, $p = 0.001$), such that the metaphor sentence of each pair tended to result in greater LIFG response than the literal sentence. Though here we report results from 100-voxel ROIs, this result was robust across a wide range of ROI sizes (10–800 voxels). It is this increase in LIFG activation to metaphors that we aim to explain with the data obtained in Study 1. That is, can we predict the extent to which, for each object concept, metaphors increased activation in LIFG?

3.2.2. Sensitivity to conflict in LIFG

We extracted the *t*-statistics from the contrast of incongruent > neutral trials in the Stroop color-word interference task within each subject's ROI. Across subjects, the LIFG ROI was sensitive to Stroop-conflict ($t(15) = 2.36$, $p = 0.032$). This result was robust across a wide range of voxel sizes (90–800 voxels).

3.2.3. P-index predicts LIFG response to metaphors

Our goal here is to predict the extent to which LIFG activity increases during metaphor comprehension, relative to literal sentence comprehension, on an item level. We hypothesized that if metaphors are comprehended via a process that activates or suppresses conceptual properties (as suggested by Study 1), then LIFG could house the responsible neural mechanism. For each of the 48 items, we averaged the mean activation in the 100 voxel subject-specific ROIs across subjects to obtain a measure of LIFG response for the LIT and MET version of each pair; we then calculated the difference between these activations (MET-LIT). We thus had, for each item, a difference-score that represented the extent to which LIFG response was greater for the metaphor than for the literal sentence.

We predicted that the P-index for each item would predict the magnitude of the increased LIFG response during metaphor comprehension. A rank-based correlation between these measures reveals a marginal effect ($r(47) = 0.28$, $p = 0.053$; Fig. 4). If the P-index was constructed from the high-accuracy sample, this relationship is significant ($r(47) = 0.32$, $p = 0.029$); the P-index includes the full sample in all following analyses. We also fit a multiple regression model to test the strength of relationship after controlling for salience, production frequency, distinctiveness of metaphor-relevant property, semantic distance of the properties to the vehicle concept (LSA), and metaphor familiarity. The P-index reliably predicts the LIFG effect in this model ($B = 0.22$, $p = 0.02$). P-index was the only measure that reliably predicted the MET-LIT difference in LIFG response. Though not reaching significance, semantic distance ($p = 0.06$) and frequency ($p = 0.08$) also appear to be somewhat predictive of increased LIFG response to metaphors.

Since the P-index is constructed out of a comparison between relevant and irrelevant property activation, the variables we included in the regression also controlled for the *difference* between the metaphor-relevant and -irrelevant properties (as in Study 1). We nevertheless confirmed that these results hold when controlling for only the metaphor-relevant properties, since the metaphor-irrelevant properties were completely unrelated to the fMRI task in Study 2. Results from these two regression models are shown in Table 1.

In summary, we found that our measure of property-selection and property-suppression obtained in Study 1 (P-index) reliably predicted the increased LIFG response to metaphors in Study 2, supporting our hypothesis that LIFG is involved in activating or suppressing conceptual properties during metaphor comprehension.

3.2.4. Specificity to LIFG

In order to test the specificity of this effect, we ran the same analyses in other regions putatively involved in metaphor comprehension: right inferior frontal gyrus (RIFG), left fusiform gyrus (FG), and left middle temporal gyrus (MTG). ROIs in these regions were

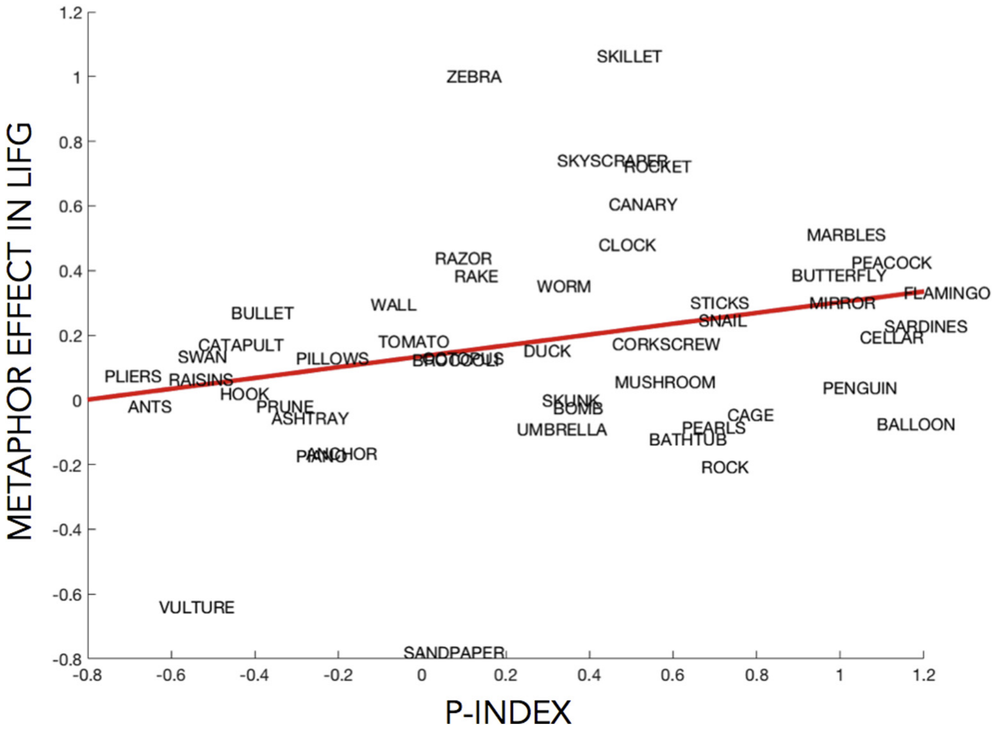

Fig. 4. P-index predicts increased LIFG response to metaphors. The P-index measure obtained in Study 1, which reflects the extent to which conceptual properties are selected during metaphor comprehension, predicts increase in LIFG response to metaphors relative to literal sentences (B = 0.16, $p = 0.025$) after controlling for property salience, production frequency, distinctiveness, pairwise distance from object concept (LSA), and metaphor familiarity.

Table 1

Predicting LIFG Response to Metaphors. Results of our regression models. Variables included in the model were the P-index (calculated based on results of Study 1), property salience, property frequency, property distinctiveness, and semantic distance of property and object concept. (A) Controlling for difference between metaphor-relevant and -irrelevant properties. (B) Controlling for metaphor-relevant property only.

(A) *IRR-REL*	*B*	*t*	*p*
P-index	0.22	2.44	0.02*
Salience	0.06	1.37	0.18
Frequency	−0.01	−1.79	0.08
Distinctiveness	−0.15	−1.41	0.16
Semantic Distance (LSA)	−0.39	−1.96	0.06
Familiarity	0.01	0.16	0.87
(B) *REL*	*B*	*t*	*p*
P-index	0.19	2.13	0.04*
Salience	0.11	1.42	0.16
Frequency	−0.01	−1.38	0.17
Distinctiveness	−0.18	−1.17	0.25
Semantic Distance (LSA)	−0.47	−1.30	0.20
Familiarity	−0.02	−0.29	0.77

* the asterisk was intended to mark the significant regression results.

defined as 123-voxel spheres centered around the peak voxel of a cluster that emerged in a whole brain contrast of MET > LIT (no clusters were significant, but threshold was reduced to reveal regions sensitive to this contrast). P-index did not correlate with metaphor-related activity in RIFG ($p > 0.3$) or left MTG ($p = 0.10$). We ran the same regression model described above in these ROIs, and no variables (including P-index) significantly predicted metaphor-related neural activity. However, we found that P-index was negatively correlated with metaphor-related activity in left fusiform gyrus (FG) ($r(47) = -0.31$, $p = 0.034$). When we controlled for other variables, this relationship was no longer significant ($p = 0.13$). These results suggest that, within the neural regions sensitive to metaphor processing, LIFG is uniquely involved in the dynamic selection of properties during metaphor comprehension.

4. Discussion

Metaphorical language "stretches" concepts, enabling a wide range of meanings to emerge. Our results suggest that the cognitive mechanism involved in this process operates on the level of conceptual properties: semantic control mechanisms work to activate relevant properties or suppress irrelevant properties during metaphor comprehension. Further, we provide evidence that the specific property inhibition demands engendered by each metaphor predict activity in LIFG during metaphor comprehension, thus providing an explanation for prior reports of this region's increased response to figurative versus literal language.

Conceptual properties of the tenor and vehicle play a prominent role in most cognitive theories of metaphor comprehension (Gernsbacher et al., 2001; Glucksberg, 2003; Glucksberg & Keysar, 1990; Glucksberg et al., 1997; Ortony, 1979; Tourangeau & Rips, 1991; Tourangeau & Sternberg, 1982), and our finding that metaphor comprehension entails the inhibition of irrelevant property information is in line with previous empirical findings (Gernsbacher et al., 2001; Glucksberg, Newsome, & Goldvarg, 2001; Fernandez, 2007; Taira & Kusumi, 2012). Though we are not the first to show that metaphor comprehension involves the selective inhibition of information, we go beyond the previous literature by using levels of this metaphor-specific property inhibition to predict the magnitude of cortical, language-related activity.

There have been many neuroimaging studies examining the neural correlates of the metaphor comprehension process (Bambini et al., 2011; Cardillo et al., 2012; Eviatar & Just, 2006; Lee & Dapretto, 2006; Rapp et al., 2004, 2007; Stringaris et al., 2007), and LIFG is the region most consistently implicated in these tasks (Bohrn et al., 2012). Our fMRI task elicited contributions from RIFG, left fusiform gyrus, and left middle temporal gyrus; though these regions did not reach significance on the group-level, they are consistent with prior findings (see Bohrn et al., 2012 for a meta-analysis). These cortical regions undoubtedly play important roles in figurative language processing, but we will not speculate on their functions here. However, we did find that the P-index for each item negatively predicted metaphor-related activity in left fusiform gyrus (FG), a region that has been reported to represent conceptual properties of objects (e.g. Kan, Barsalou, Olseth Solomon, Minor, & Thompson-Schill, 2003; Simmons et al., 2007). It is thus possible that a decreased response in left FG during comprehension of high P-index metaphors reflects the fact that more property information is being suppressed for these items. That is, metaphors with high conceptual adjustment demands require more property information to be suppressed, and if this information is represented in left FG, then less activity will be observed in these cases.

Our goal was to specifically target the cognitive mechanisms related to LIFG recruitment, rather than to uncover the broad functional network that supports metaphor processing. Despite the many studies reporting involvement of this cortical region, none of them have, to our knowledge, directly probed the cognitive mechanisms associated with LIFG during figurative language processing. We aimed to bridge the gap between cognitive and neural theories, and empirically test the relationship between conceptual properties, semantic control processes recruited during conceptual adjustment, and the increased response in left prefrontal cortex.

We extracted a measure of property-inhibition for each item (P-index) from our RT data from Study 1, and interpret this measure as reflecting the particular property inhibition demands engendered by each metaphor. Though this measure was constructed using one specific relevant property and one specific irrelevant property, we consider the P-index to be a relatively stable property of a metaphor, and robust to the choice of properties used in the property-verification task. (In order to support this interpretation, we matched the relevant and irrelevant properties on frequency, salience, distinctiveness, and semantic distance from the target concept across items.) Each metaphor is thus assigned a P-index value that represents the cost of suppressing irrelevant property information during comprehension, which also can be thought to represent the degree of conceptual adjustment required during the comprehension process. As discussed above, we do not believe that this process is unique to figurative language processing, but rather a ubiquitous process of language comprehension more generally. Just as conceptual fine-tuning is required in figurative language, this fine-tuning is necessary in more simple utterances in which the meaning is under-determined by the linguistic content (Carston, 2010; Wilson & Carston, 2007). One can also imagine a measure analogous to our P-index in the cases of "literal" contexts: the attribute concept FRESH relates to a different cluster of properties when referring to shirts (e.g., *scented, soft*) and vegetables (e.g., *edible, crunchy*), thus the comprehension of "fresh shirt" requires the fine-tuning of the FRESH concept to match the contextual demands (Murphy & Andrew, 1993). We might therefore expect a cost of information inhibition, or a fine-tuning cost, in even these simple cases. This inhibition occurs during comprehension itself, and is not triggered by a subsequent task (e.g. the property-verification task in Study 1), enabling us to incorporate the P-index into our simple comprehension task in Study 2.

The metaphor-specific inhibition demands captured in the P-index varied across items, and we predicted that this variability would correspond to degree of LIFG recruitment. We have not, however, provided a concrete explanation for the *source* of variability in the P-index measure itself. The P-index for any given metaphor might depend on sentence-level features such as aptness and familiarity (Blasko & Connine, 1993), or conventionality and interactivity (Taira & Kusumi, 2012). However, metaphor familiarity did not predict the P-index measure in our data, suggesting an alternate source of variation. We speculate that variation in P-index most likely relates to property-specific features such as distinctiveness (Randall, Moss, Rodd, Greer, & Tyler, 2004; Taylor, Devereux, Acres, Randall, & Tyler, 2012; Tyler, Moss, Durrant-Peatfield, & Levy, 2000), salience (e.g., Giora, 1999), or property domain (e.g., color, taste, texture) and the ways in which these property features interact with each other across tenor and vehicle concepts. That is, the P-index reflects not (just) the ways in which the vehicle's properties interact with each other, but the way in which these properties interact with the surrounding conceptual context provided by the metaphor.

Our hypothesis on the role of LIFG in metaphor comprehension stems from prior findings that this region is recruited for selecting between multiple, competing representations (e.g., Thompson-Schill et al., 1997, 1998, 1999; Bedny et al., 2008; Hindy et al., 2012, 2015; Solomon et al., 2015). Stroop-conflict is a useful way to assess sensitivity to this general form of competition. Here we asked whether metaphor comprehension involves the selection between multiple conceptual properties: The fact that LIFG response correlated with our behavioral measure of property inhibition (P-index) and is also sensitive to Stroop-conflict provides support for our hypothesis that LIFG is involved in selecting amongst conceptual properties during metaphor comprehension.

Determining which conceptual properties are relevant to a metaphor is crucial for novel metaphors, for which the figurative meaning is generated during the comprehension process. LIFG involvement could thus be reduced for familiar or conventionalized metaphors, where the meaning may not have to be generated online. Cardillo et al. (2012) found that LIFG response to metaphors did indeed decrease with increased metaphor familiarity, suggesting that conventionalization of metaphors tunes activity within this region. Though it has been suggested that the right hemisphere is brought online for the processing of novel vs. conventionalized metaphors (Graded Salience Hypothesis; Giora, 1997), this does not contradict the possibility that the left hemisphere is similarly influenced by the increased effort needed to generate novel meanings. Our current results are consistent with the idea that LIFG activation is associated with conceptual adjustment during the comprehension of novel metaphors, though the potential contribution of RIFG to this process should be further explored in future studies.

Regarding current theories of metaphor processing, our data support the claim made by both the class-inclusion model (Gernsbacher et al., 2001; Glucksberg et al., 2001) and Relevance Theory (Carston, 2010; Wilson & Carston, 2007; Fernandez, 2007) that comprehension of metaphors involves an inhibition of irrelevant information, and additionally show that this pattern holds when only basic-level properties are considered (without referring to superordinate categories or ad hoc concepts). Our question of interest focused on the levels of property activation that result from metaphor comprehension, but our analyses do not directly speak to the role of "literal" meaning in the construction of this interpretation (i.e. the direct vs. indirect views of metaphor processing), since our irrelevant/relevant properties were presented after, and not before, metaphor presentation (see Weiland et al., 2014 for a more direct comparison of these theories). But, given our view that metaphorical meaning emerges as a result of "stretching" a target concept, our views on this matter most closely align with Recanati (1995) in which literal meaning plays only a "local" role: the literal meaning of a constituent (i.e., a concept) is activated, and non-literal interpretations are derived from this initial content. However, our current studies did not aim to differentiate between theories of metaphor, but rather to make an explicit link between cognitive mechanisms and neural recruitment during the comprehension process

Here we had a narrow goal, namely to link metaphor-specific conceptual adjustment demands with recruitment of LIFG during metaphor processing. Our predictions were motivated by the combination of (1) a decompositional, flexible approach to conceptual structure, (2) theories of metaphor that propose that irrelevant information is inhibited during comprehension, and (3) the notion that LIFG is a neural region involved in inhibiting irrelevant information during conceptual and linguistic tasks. We found that the degree of property inhibition required by each metaphor predicted the increase in LIFG activity for metaphors versus their literal controls, providing a concrete mechanism for this region's involvement in figurative language comprehension.

Our data fit within a flexible theory of conceptual structure in which concepts refer to unique combinations of properties that are dynamically changed, activated, or strengthened depending on semantic context (Musz & Thompson-Schill, 2015), event context (Hindy et al., 2012, 2015; Solomon et al., 2015), primed sensory modality (Pecher et al., 2004), or motor experience (Yee, Chrysikou, Hoffman, & Thompson-Schill, 2013). We show that using a decompositional view of concepts in conjunction with a radical, unified view of figurative language comprehension holds up in empirical data, and is neurally plausible. Just as thinking of the concept RAISIN would activate different properties for a chef (e.g., *sweet*, *chewy*) than for a painter (e.g., *purple*, *wrinkled*), representing the concept RAISIN in a metaphorical context (e.g., "Her eyes are raisins") will similarly involve a conceptual stretching such that properties are reweighted to match referential demands. Figurative language provides a fertile ground on which to test the ways concepts are stretched and fine-tuned to support language comprehension more generally.

5. Material

Metaphor	Control	REL	IRR	P-index
The yoga student is a flamingo.	The bird is a flamingo.	*one leg*	*pink*	1.15
The prisoners are sardines.	The fish are sardines.	*canned*	*salty*	1.10
Her heart is a balloon.	Her gift is a balloon.	*expands*	*parties*	1.09
His mind is a cellar.	The room is a cellar.	*dark*	*basement*	1.05
The prom queen is a peacock.	His pet is a peacock.	*blue*	*eggs*	1.03
The groom is a penguin.	The cartoon is a penguin.	*black*	*beaked*	0.96
The pond is a mirror.	The wall is a mirror.	*shiny*	*breakable*	0.93
The hailstones are marbles.	The toys are marbles.	*smooth*	*games*	0.92
The ballerina is a butterfly.	The insect is a butterfly.	*delicate*	*winged*	0.88
His cubicle is a cage.	The container is a cage.	*trap*	*metal*	0.73
The bagel is a rock.	The specimen is a rock.	*hard*	*grey*	0.67
The tourist is a snail.	The critter is a snail.	*slow*	*shell*	0.66
Her legs are sticks.	The branches are sticks.	*thin*	*tree*	0.64
Her teeth are pearls.	The earrings are pearls.	*white*	*round*	0.62
The river is a bathtub.	The receptacle is a bathtub.	*washing*	*taps*	0.54
The sprinter is a rocket.	The vehicle is a rocket.	*fast*	*large*	0.48
The cloud is a mushroom.	The appetizer is a mushroom.	*stemmed*	*fungus*	0.46
His tail is a corkscrew.	The tool is a corkscrew.	*curly*	*wine*	0.45
The crayon is a canary.	The bird is a canary.	*yellow*	*sings*	0.45
The rooster is an alarm clock.	The object is an alarm clock.	*keeps time*	*ticking*	0.42
The sidewalk is a skillet.	The pan is a skillet.	*hot*	*kitchen*	0.42
Her parents are skyscrapers.	The buildings are skyscrapers.	*tall*	*elevators*	0.32
His temper is a bomb.	His weapon is a bomb.	*explosive*	*dropped*	0.31
The cigar is a skunk.	The shadow is a skunk.	*smelly*	*furry*	0.29
The train is a worm.	The creature is a worm.	*slither*	*slimy*	0.27
The scuba diver is a duck.	The animal is a duck.	*swims*	*quacks*	0.24
The tree is an umbrella.	His symbol is an umbrella.	*protective*	*handle*	0.23
Her fingers are rakes.	The implement is a rake.	*prongs*	*leaves*	0.08
The clam is a zebra.	The animal is a zebra.	*striped*	*hooved*	0.06
Her insult is a razor.	Her purchase is a razor.	*sharp*	*shaving*	0.03
The chair is an octopus.	The creature is an octopus.	*tentacles*	*suction cups*	0.00
His hair is broccoli.	The vegetable is broccoli.	*stalks*	*green*	−0.02
His skin is sandpaper.	The material is sandpaper.	*rough*	*brown*	−0.04
His face is a tomato.	His topping is a tomato.	*red*	*seeds*	−0.10
Her silence is a wall.	The barrier is a wall.	*separating*	*brick*	−0.12
Her father is an anchor.	The artifact is an anchor.	*stationary*	*boats*	−0.28
Her lips are pillows.	The decorations are pillows.	*soft*	*beds*	−0.30
His mouth is a piano.	His instrument is a piano.	*keys*	*pedals*	−0.30
The street is an ashtray.	The item is an ashtray.	*butts*	*glass*	−0.36
His grandmother is a prune.	His dessert is a prune.	*wrinkled*	*purple*	−0.40
Her answer is a bullet.	Her weapon is a bullet.	*injure*	*gun*	−0.46
His nose is a hook.	His accessory is a hook.	*curved*	*hanging*	−0.48

(*continued on next page*)

(*continued*)

Metaphor	Control	REL	IRR	P-index
His promotion is a catapult.	His artifact is a catapult.	*launching*	*medieval*	−0.54
The dancer is a swan.	The blur is a swan.	*graceful*	*long neck*	−0.58
Her eyes are raisins.	The snacks are raisins.	*dried*	*sweet*	−0.61
The opponent is a vulture.	The predator is a vulture.	*scavenger*	*talons*	−0.63
The campers are ants.	The intruders are ants.	*colonies*	*antennae*	−0.70
His arms are pliers.	The tools are pliers.	*pulling*	*steel*	−0.76

Funding sources

This work was supported by the National Institutes of Health grants R01DC015359 and R01EY021717 awarded to STS, an NSF IGERT Fellowship awarded to SHS, and National Institutes of Health grant P30NS45839.

References

Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M., & Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin, 86*(3), 203–216.

Bedny, M., Aguirre, G. K., & Thompson-Schill, S. L. (2007). Item analysis in functional magnetic resonance imaging. *Neuroimage, 35*(3), 1093–1102.

Bedny, M., McGill, M., & Thompson-Schill, S. L. (2008). Semantic adaptation and competition during word comprehension. *Cerebral Cortex, 18*(11), 2574–2585.

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.

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language—A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia, 50*(11), 2669–2683.

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), 3212–3221.

Carston, R. (2010). Lexical pragmatics, ad hoc concepts and metaphor: From a relevance theory perspective. *Italian Journal of Linguistics, 22*(1), 153–180.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Creating concepts from converging features in human cortex. *Cerebral Cortex*, bhu057.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage, 25*(4), 1325–1335.

Eviatar, Z., & Just, M. A. (2006). Brain correlates of discourse processing: An fMRI investigation of irony and conventional metaphor comprehension. *Neuropsychologia, 44*(12), 2348–2359.

Fernandez, P. R. (2007). Suppression in metaphor interpretation: Differences between meaning selection and meaning construction. *Journal of Semantics, 24*(4), 345–371.

Fodor, J. A., & Lepore, E. (1998). The emptiness of the lexicon: Reflections on James Pustejovsky's The Generative Lexicon. *Linguistic Inquiry, 29*(2), 269–288.

Gernsbacher, M. A., Keysar, B., Robertson, R. R., & Werner, N. K. (2001). The role of suppression and enhancement in understanding metaphors. *Journal of Memory and Language, 45*(3), 433–450.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics (includes Cognitive Linguistic Bibliography), 8*(3), 183–206.

Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics, 31*(7), 919–929.

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences, 7*(2), 92–96.

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review, 97*(1), 3.

Glucksberg, S., McGlone, M. S., & Manfredi, D. (1997). Property attribution in metaphor comprehension. *Journal of Memory and Language, 36*(1), 50–67.

Glucksberg, S., Newsome, M. R., & Goldvarg, Y. (2001). Inhibition of the literal: Filtering metaphor-irrelevant information during metaphor comprehension. *Metaphor and Symbol, 16*(3–4), 277–298.

Hindy, N. C., Altmann, G. T., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object changes on event processing: Do objects compete with themselves? *The Journal of Neuroscience, 32*(17), 5795–5803.

Hindy, N. C., Solomon, S. H., Altmann, G. T., & Thompson-Schill, S. L. (2015). A cortical network for the encoding of object change. *Cerebral Cortex, 25*(4), 884–894.

Jackendoff, R. (2002). *Foundation of language-brain, meaning, grammar, evolution*. NY: Oxford University Press (247 p.).

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17*(2), 825–841.

Kan, I. P., Barsalou, L. W., Olseth Solomon, K., Minor, J. K., & Thompson-Schill, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology, 20*(3–6), 525–540.

Lee, S. S., & Dapretto, M. (2006). Metaphorical vs. literal word meanings: fMRI evidence against a selective role of the right hemisphere. *Neuroimage, 29*(2), 536–544.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*(4), 547–559.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126*(2), 99.

Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., & Kramer, A. F. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research, 12*, 467–473.

Murphy, G. L., & Andrew, J. M. (1993). The conceptual basis of antonymy and synonymy in adjectives. *Journal of Memory and Language, 32*(3), 301.

Musz, E., & Thompson-Schill, S. L. (2015). Semantic variability predicts neural variability of object concepts. *Neuropsychologia*.

Ortony, A. (1979). Beyond literal similarity. *Psychological Review, 86*(3), 161.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin & Review, 11*(1), 164–167.

Ralph, M. A. L., Sage, K., Jones, R. W., & Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences, 107*(6), 2717–2722.

Randall, B., Moss, H. E., Rodd, J. M., Greer, M., & Tyler, L. K. (2004). Distinctiveness and correlation in conceptual structure: behavioral and computational studies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 393.

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., & Kircher, T. T. (2004). Neural correlates of metaphor processing. *Cognitive Brain Research, 20*(3), 395–402.

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., & Kircher, T. T. (2007). Laterality in metaphor processing: Lack of evidence from functional magnetic resonance imaging for the right hemisphere theory. *Brain and Language, 100*(2), 142–149.

Recanati, F. (1995). The alleged priority of literal interpretation. *Cognitive Science, 19*(2), 207–232.

Searle, J. R. (1978). Literal meaning. *Erkenntnis, 13*(1), 207–224.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia, 45*(12), 2802–2810.

Solomon, S. H., Hindy, N. C., Altmann, G. T., & Thompson-Schill, S. L. (2015). Competition between mutually exclusive object states in event comprehension. *Journal of Cognitive Neuroscience*.

Stringaris, A. K., Medford, N. C., Giampietro, V., Brammer, M. J., & David, A. S. (2007). Deriving meaning: Distinct neural mechanisms for metaphoric, literal, and non-meaningful sentences. *Brain and Language, 100*(2), 150–162.

Taira, T., & Kusumi, T. (2012). Relevant/irrelevant meanings of topic and vehicle in metaphor comprehension. *Metaphor and Symbol, 27*(3), 243–257.

Taylor, K. I., Devereux, B. J., Acres, K., Randall, B., & Tyler, L. K. (2012). Contrasting effects of feature-based statistics on the categorisation and basic-level identification of visual objects. *Cognition, 122*(3), 363–374.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, 94*(26), 14792–14797.

Thompson-Schill, S. L., D'Esposito, M., & Kan, I. P. (1999). Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron, 23*(3), 513–522.

Thompson-Schill, S. L., Swick, D., Farah, M. J., D'Esposito, M., Kan, I. P., & Knight, R. T. (1998). Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. *Proceedings of the National Academy of Sciences, 95*(26), 15855–15860.

Tourangeau, R., & Rips, L. (1991). Interpreting and evaluating metaphors. *Journal of Memory and Language, 30*(4), 452–472.

Tourangeau, R., & Sternberg, R. J. (1982). Understanding and appreciating metaphors. *Cognition, 11*(3), 203–244.

Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences, 5*(6), 244–252.

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., & Levy, J. P. (2000). Conceptual structure and the structure of concepts: A distributed account of category-specific deficits. *Brain and Language, 75*(2), 195–231.

Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP. *Frontiers in Human Neuroscience, 8*, 583.

Wilson, D., & Carston, R. (2007). A unitary approach to lexical pragmatics: Relevance, inference and ad hoc concepts. In N. Burton-Roberts (Ed.), *Pragmatics* (pp. 230–2259). Basingstoke: Palgrave.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage, 14*(6), 1370–1386.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic bulletin & review, 23*(4), 1015–1027.

Yee, E., Ahmed, S. Z., & Thompson-Schill, S. L. (2012). Colorless green ideas (can) prime furiously. *Psychological science, 23*(4), 364–369.

Yee, E., Chrysikou, E. G., Hoffman, E., & Thompson-Schill, S. L. (2013). Manual experience shapes object representations. *Psychological Science* 0956797612464658.