What's Big and Fluffy But Can't Be Seen? Selective Unimodal Processing of Bimodal Property Words

Louise Connell (louise.connell@manchester.ac.uk)
Dermot Lynott (dermot.lynott@manchester.ac.uk)
School of Psychological Sciences, University of Manchester
Oxford Road, Manchester M13 9PL, UK

Abstract

Recent work has shown that perceptual and conceptual processing share a common, modality-specific neural substrate and appear to share the same attentional mechanisms. However, this work has been largely limited to the conceptual processing of unimodal properties (i.e., that involve information from only one sensory modality) even though most perceptual properties are actually multimodal (i.e., involve information from more than one sensory modality). In two experiments, we investigate whether the conceptual processing of bimodal properties (e.g., *fluffy*, *jagged*) requires representation on both modalities or if it is instead possible for individual modalities to carry the representational burden. Results show that bimodal properties must be processed on both component modalities when attentional control is governed by incoming stimuli (i.e., exogenously), but a "quick and dirty" unimodal representation can suffice when selective conscious (i.e. endogenous) attention has time to suppress the non-target modality. We discuss these findings with reference to embodied views of cognition.

Introduction

How do we think about objects that are not in front of us at the time? Do we see with the mind's eye and touch with the mind's fingers? Embodied cognition research represents a recent trend to cease viewing conceptualisation and mental representation in terms of abstract information processing and rather in terms of grounded, sensorimotor simulation. As a common thread, embodied theories (Barsalou, 1999; Gibbs, 2003; Glenberg, 1997; Johnson-Laird, 1983; Pecher & Zwaan, 2005) hold that conceptual thought is grounded in the same neural systems that govern sensation, perception and action. For example, according to Barsalou's (1999, 2008) Perceptual Symbol Systems, concepts are essentially partial recordings of the neural activation that arises during perceptual and motor experiences and these recordings can later be re-enacted as a perceptual simulation of that concept. Such grounded accounts of cognition are in contrast to other accounts that assume concepts to be discrete representations stored in semantic memory, separated from systems governing perception and action (e.g., Collins & Quillian, 1969, Katz & Fodor, 1963; Kintsch & van Dijk, 1978; Fodor, 1975; Newell & Simon, 1972; Pylyshyn, 1984; Tulving, 1972).

A growing body of empirical work has emerged in support of embodied representations of conceptual information, and most research on modality-specific conceptual representations has focussed on unimodal object properties (i.e., properties that can be perceived using one sense alone). For example, using fMRI, Gonzáles and colleagues (2006) found that passively reading scent-related words (e.g., *cinnamon*) increased activation in the primary olfactory areas of the piriform cortex. Regarding visual processing, Simmons et al. (2007) showed that verifying colour properties in text (e.g., that a banana is *yellow*) led to activation in the same region of the left fusiform gyrus in the visual cortex as a perceptual task that involved judging colour sequences. Further comparisons by Goldberg, Perfetti and Schneider (2006) found that verification of colour, sound, touch and taste properties activated cortical regions respectively associated with encoding visual, auditory, haptic and gustatory experiences, illustrating that perceptual experience and conceptual knowledge share a common neural substrate.

Other recent work has focused on the emergence of perceptual phenomena, such as modality switching costs, in conceptual processing (e.g., Marques, 2006; Pecher, Zeelenberg & Barsalou, 2003; van Dantzig, Pecher, Zeelenberg & Barsalou, 2008). For example, Spence, Nicholls and Driver (2001; see also Turatto, Galfano, Bridgeman & Umiltà, 2004) asked people to indicate the left/right location of a series of perceptual stimuli, and found that switching modalities from one trial to the next (e.g., from a visual light flash to an auditory tone) incurred a processing cost. Pecher et al. (2003) replicated this paradigm in a conceptual task by asking people to verify a series of unimodal object properties presented as text onscreen, and found that people were slower to verify a property in a given modality (e.g., auditory *leaves:rustling*) after verifying a property in a different modality (e.g., visual *apple:shiny*) than after verifying a property in the same modality (e.g., auditory *blender:loud*), and that this effect was not due to associative priming. Van Dantzig et al. (2008) took the paradigm one step further by showing that property verification was also slowed when it followed a perceptual stimulus from a different modality (e.g., auditory *bee:buzzes* following a visual light flash). Like the modality switching costs found by Spence et al. during perceptual tasks, such costs during conceptual tasks result from the re-allocation of attention from one modality-specific system to another.

However, it is overly simplistic to assume that object properties are conceptually processed only in single, modality-specific regions of the brain. From philosopher

John Locke (1975/1690) to modern empirical studies (Amedi, von Kriegstein, van Atteveldt, Beauchamp & Naumer, 2005; Connell, 2007; Ernst & Bülthoff, 2004), many have pointed out that object properties exist in both multimodal (that can be perceived by multiple senses) and unimodal (that can be perceived by only one of our senses) forms. For example, while the colour *yellow* is normally perceived only through the single modality of vision (i.e., colour is a unimodal property), the property *round* would be considered multimodal as it can be perceived both haptically and visually. Indeed, the bimodal overlap of touch and vision in the conceptual processing of object properties has emerged in imaging work: Newman, Klatzky, Lederman and Just (2005) noted that the intraparietal sulcus, a region usually involved in visual imagery, was found to be highly activated when participants were processing the roughness of objects (e.g., which is rougher? *pear* or *egg*). The unimodal / multimodal distinction has recently been highlighted in a norming study conducted by Lynott and Connell (2009), who collected ratings of experiential strength on five perceptual modalities (visual, haptic, auditory, olfactory, gustatory) for hundreds of object properties. They found that most properties are multimodal rather than unimodal, with particular bimodal clustering of visual-haptic and olfactory-gustatory modalities.

The Current Study

Neuroimaging research suggests that automatic and extensive interaction between modalities may be the norm rather than the exception in perception (Shimojo & Shams, 2001). Furthermore, a strong interpretation of embodied theories of representation (e.g., Barsalou, 1999) would suggest that the conceptual processing of a visuohaptic word would involve simulating both the visual and haptic modalities (e.g., representing *round* would involve simulating both sight and touch). Since most object property words combine two or more perceptual modalities (Lynott & Connell, 2009), our aim in the following experiments is to examine whether conceptual processing involves a similar automatic interaction between modalities or if it is instead possible for individual modalities to carry the representative burden. If bimodal and unimodal properties entail different perceptual representations, then differential modality switching effects (Experiment 1) and modality detection rates (Experiment 2) should emerge during their conceptual processing as attention is directed to the perceptual modalities needed for simulation.

Experiment 1

In the modality switching effect (Marques, 2006; Pecher et al., 2003; van Dantzig et al., 2008), there is a processing cost involved in representing a modality-specific object property when attention has to be shifted from a different, already active perceptual modality. The present experiment uses a property verification task with the same basic methodology as Pecher et al. but will instead present a combination of unimodal and multimodal target properties. For example, a visual property (e.g., *window:misty*) may be followed by a unimodal visual target from the same modality (e.g., *pond:murky*) or a bimodal visuohaptic target that shares a component modality (e.g., *blade:jagged*).

Our main interest here is in how unimodal properties differentially facilitate targets from the same versus shared modality. If bimodal properties are always automatically represented using both component modalities, then it should take people longer to process bimodal targets (e.g., visual→visuohaptic) compared to unimodal targets (e.g., visual→visual), because there will be costs involved in allocating attention to the haptic modality.

Method

Participants Twenty-seven native speakers of English, with no reported reading or sensory deficits, participated in the experiment for course credit. Modal specificity (unimodal or bimodal) was manipulated within-participants, and each participant received one of nine counterbalanced lists of concept:property pairs.

Materials Fifty-four concept-property matches were created to act as test items, divided between unimodal (visual or haptic) and bimodal (visuohaptic) items. Property words were taken from Lynott & Connell's (2009) modality exclusivity norms. These norms comprise 423 adjectives, each describing an object property, with mean ratings (0-5) of how strongly that property is experienced through each of five perceptual modalities (auditory, gustatory, haptic, olfactory, visual) plus a number of other useful statistics. For this experiment, all words chosen had a familiarity score of at least 90% in the norms. Unimodal words had the highest strength rating in the visual or haptic modality (minimum strength of 3 on a 0-5 scale) with all other modalities at least one full point lower on the ratings scale. Likewise, bimodal words had joint highest strength ratings in visual and haptic modalities (both ratings over 3 and within one ratings point of each other) and all other modalities were at least one full point lower. There were no differences between unimodal and bimodal properties in target strength [$t(52) = 1.63$ $p > .1$]. Each property was then matched with an appropriate concept, such as *water:rippling* (unimodal visual), *draft:cold* (unimodal haptic), or *cactus:spiky* (bimodal visuohaptic). Two independent judges verified the appropriateness of all 54 attributions. There were no differences between unimodal and bimodal items in summed concept-property British National Corpus (BNC, 2001) word frequencies [$t(52) = 1.34$, $p > .15$], or orthographic length [$t(52) = 0.285$ $p > .8$]. We then formed pairs of concept-property items for sequential presentation by selecting a unimodal item (to be presented first) and pairing it with another unimodal or bimodal item (the target). The pairing of each target item with its preceding modality was fully rotated over nine lists: for example, a visual item would appear as the first item in a pair in one list and the second item in another, or a visuohaptic item would be presented following a visual item in one list and a haptic item in another. Each participant saw every item, but in only one of these nine possible critical pairs.

Table 1: Sample concept:property pairs per modal specificity condition in Experiment 1 with mean verification times and standard deviations (in milliseconds).

Modal specificity	Sample pairs	Transition type	*M*	*SD*
Unimodal	window:misty → pond:murky	(visual → visual)	1039	157
	sunburn:stinging → wool:itchy	(haptic → haptic)		
Bimodal	magazine:glossy → cactus:spiky	(visual → visuohaptic)	1064	151
	marble:cool → fabric:silky	(haptic → visuohaptic)		
Engagement cost			*25*	

A list of 96 concept-property fillers was also created, 72 false and 24 true, to provide an overall balance of 50:50 true:false responses per participant. As in Pecher et al.'s Experiment 1, most of the false fillers were associated in Nelson, McEvoy, and Schreiber's (2004) word association norms (e.g.. *oven:baked*, *coffin:dead*) in order to ensure participants could not perform the task using simple word association strategies (Solomon & Barsalou, 2004).

Procedure Participants read instructions that asked them to press the button labelled "true" (the comma key) if the property was usually true of the concept but to press the button labelled "false" (the full-stop key) if not. We used Pecher et al.'s (2003) example "carnation can be black" to highlight that, although carnations could theoretically be black, it would be highly unusual and should be judged as false. Each trial began with a fixation cross for 200 ms followed by the item in the form "concept can be property" which stayed onscreen until the participant responded. Participants received immediate feedback if they responded incorrectly or too slowly (more than 2000 ms), and each trial ended with a 200 ms blank screen. A practice session of 24 items, half true and half false, preceded the main experiment. Critical pairs and fillers appeared in a random order with a self-paced break every 48 trials.

Results & Discussion

Two participants' data were removed prior to analysis due to achieving less than 70% accuracy. Any targets that received error responses, and any targets where the preceding item in the pair was in error, were excluded from analysis (7.3% of data in total). There was no difference in the error rate between unimodal (M = 5.7%, SD = 7.1%) and bimodal (M = 4.0%, SD = 6.9%) targets, $t(24) = 0.96$, $p = .346$. Response time means (in milliseconds) were calculated as the mean of the medians per participant per condition to minimise the effect of outliers.

As predicted, people were slower to verify bimodal properties than unimodal properties (see Table 1: directional $t(24) = 1.96$, $p = .031$). When a single modality was already active, people encountered a processing cost for bimodal properties that shared a component modality (i.e., a partial overlap) compared to processing the same single modality again (i.e, a complete overlap). This effect is not a switching cost, as described by Pecher et al. (2003) or Spence et al. (2001), because attention is not being decoupled from one modality and coupled to another; rather, the effect is an *engagement cost* because attention must be split and coupled to an additional modality while still remaining focused on the original. In other words, results suggest that verifying a bimodal property such as *fluffy* requires representation on both visual and haptic component modalities because both sight and touch are involved in its perceptual simulation.

Experiment 2

In property verification tasks, each new stimulus that comes along directs attention to its particular modality (or modalities) for processing. This type of attentional mechanism is exogenous control, where the modality involved in processing a word (or perceptual stimulus) automatically and obligatorily grabs attention. There is also endogenous attentional control, where participants consciously and voluntarily focus their attention on the target modality. Perceptual studies have shown that endogenous attention on a particular modality creates anticipatory activation in the relevant area of the cortex (Foxe, Simpson, Ahlfors & Saron, 2005) and allows information from the target modality to be processed faster than information from other modalities (Spence et al., 2001; Turatto et al., 2004). Furthermore, endogenous attention on a particular modality during presentation of bimodal stimuli can suppress activation in the cortex corresponding to the unattended modality (e.g., attending to vision for an audiovisual stimulus results in suppression in the auditory cortex: Johnson & Zattore, 2005). Our aim in this experiment is, therefore, to see if selective endogenous attention can overcome the obligatory exogenous grab of attention that we observed for bimodal stimuli in Experiment 1. We use a modality detection task to examine conceptual processing of unimodal and bimodal property words, in a variant of the paradigm used to examine the positive/negative detection of emotionally affective words at near-subliminal thresholds (Dijksterhuis & Aarts, 2003). Endogenous attention will be directed to a target modality (visual or haptic) for a particular block of stimuli and participants will be asked to judge whether each presented property corresponds to the target modality for that block. For example, in a visual block, participants should detect both unimodal (e.g., *misty*, *green*) and bimodal (e.g., *big*, *fluffy*) stimuli as properties with visual information. We expect accuracy to improve from near-chance performance over successive blocks, both because of practice effects and because longer display durations increase the probability of successful detection. However, by measuring accuracy rates for a range of increasing display times, we can test how endogenous and exogenous attention interact.

If bimodal properties must always be represented using both component modalities, then even focusing endogenous attention on a single modality will not be enough to prevent a bimodal word exogenously directing attention towards its other modality (e.g., processing visuohaptic *fluffy* in a visual block will still need haptic attention). In this case, we would expect accuracy rates for bimodal properties to be lower than those for unimodal properties in the same block. On the other hand, if the effects of selective attention found for perceptual processing (Foxe et al., 2005; Johnson & Zattore, 2005) extend to conceptual processing, then it should be possible to represent bimodal properties unimodally if a single component modality is the focus of endogenous attention, and so we would expect the bimodal and unimodal properties in a block to be detected equally accurately.

Method

Participants Sixty native speakers of English, with no reported reading or sensory deficits, participated in the experiment for course credit. Modal specificity (unimodal or bimodal) and display duration (17ms, 33ms, 50ms, 67ms, 100ms) were both manipulated within-participants. and each participant received one of two counterbalanced lists.

Materials A set of 128 words were taken from Lynott & Connell's (2009) modality exclusivity norms: 64 test items and 64 fillers. Unimodal and bimodal test items (32 of each) were selected with the same criteria as in Experiment 1. Bimodal properties were split into two lists: one to appear in visual blocks and one in haptic blocks (counterbalanced). There was no difference between unimodal and bimodal words in target modality strength (i.e., properties were equally perceptible by sight in visual blocks [$t(46) = 1.65$ $p > .1$] and by touch in haptic blocks [$t(46) = 1.01$ $p > .3$]). In addition, lexical decision times [$t(62) = 0.45$ $p > .6$] and accuracy [$t(62) = 0.42$ $p > .6$] were equivalent for unimodal and bimodal words (English Lexicon Project database: Balota et al., 2007).

Thirty-two filler items were selected per block so that each filler word had a low strength rating (less than 2) on the target modality. This meant that all fillers had significantly lower strength on the target modality than the corresponding test words [$t(158) = 54.97$ $p < .0001$]. In order to minimise possible interference with bimodal stimuli, filler items also had lower strength on the unattended modality than bimodal test items (e.g., in visual blocks, the haptic strength of filler items was significantly less than the haptic strength of visuohaptic items), [$t(158) = 9.47$ $p < .0001$].

Procedure Participants were instructed that they would be asked to judge whether or not words appearing onscreen could be experienced through a particular sense; either felt through touch or seen. They were told that words would appear onscreen one at a time and be covered very quickly by a row of Xs, and that they should press "Yes" (the comma key) if the word could be perceived through that sense or "No" (the full stop key) if it could not. Stimuli were arranged into blocks of test and filler words for each modality; since all test items pertained to the given modality and all fillers did not, there was an equal ratio of yes:no responses within each block. At the start of each block, participants were told which sensory modality they would be making judgements about. When participants had completed both haptic and visual modality blocks with a display duration of 17s, the same blocks were repeated at 33ms, then 50ms, 67ms, and lastly 100ms (the presentation of visual or haptic blocks first was counterbalanced across participants). Items were presented randomly within each block, with each trial beginning with a central fixation (250ms), followed by a word (displayed for different durations depending on the block), followed by a mask (a row of Xs) until the participant responded. Response times were measured from mask onset to keypress[1].

Table 2: Mean percentage accuracy, with standard deviations, per modal specificity and display duration of properties in Experiment 2.

Display Duration (ms)	Modal specificity			
	Unimodal		Bimodal	
	M	*SD*	*M*	*SD*
17	47.4	20.3	42.9	21.3
33	63.4	23.1	65.7	20.1
50	71.6	20.5	73.5	20.5
67	76.8	13.6	79.0	13.6
100	79.2	13.9	79.5	15.1

Results & Discussion

Responses to test words less than 200 ms or more than three standard deviations away from a participant's mean per display duration were removed as outliers (2.1% of data). The percentage of correctly detected test words per modal specificity per display time is shown in Table 2.

As expected, there was an overall main effect of display duration [$F(4, 236) = 80.07$, $p < .0001$], with planned contrasts showing that people became more accurate with each increasing duration up to 67ms (all ps $< .003$) and performance levelling out between 67ms and 100ms ($p > .2$). Modal specificity had no main effect [$F(1, 59) = 0.09$, $p = .771$] but did interact significantly with duration [$F(4, 236) = 4.78$, $p = .001$]. In simple effects analysis, accuracy for bimodal properties was worse than that for unimodal properties at the shortest display duration [17ms: $t(59) = 2.43$, $p = .018$], even though the strength on the target modality and the lexical decision times for each word were equal for both unimodal and bimodal words. Longer exposure to the properties, however, made this difference disappear [33ms: $t(59) = 1.16$, $p > .2$; 50ms $t(59) = 1.04$, $p > .3$; 67ms $t(59) = 1.27$, $p > .2$; 100ms $t(59) = 0.17$, $p > .8$].

These results show a mixture of complete and "quick and dirty" conceptual processing of bimodal properties,

[1]It could be argued that the button-pressing nature of the task could interfere with the simultaneous processing of haptic words (e.g., Kaschak et al., 2005). However, in a related study (Connell & Lynott, 2009) we compared modality detection performance using this methodology to that using a verbal task (where participants respond with a voice trigger rather than a keypress) and found no evidence of any such interference.

suggesting that endogenous attention can selectively modulate modality-specific representation at least some of the time. When a word is displayed for only 17ms, and people are not necessarily conscious of having read it, bimodal properties are more difficult to detect as visible or touchable than unimodal properties. This difference suggests that bimodal properties are being simulated on both component modalities and are exogenously directing attention towards whichever modality is not the current subject of endogenous focus. Dividing attention in this way means that 17ms display time is not enough to process whether a visuohaptic word like *fluffy* or *round* corresponds to the target sense of vision (or touch) as easily as a unimodal word like *glossy* (or *clammy*). With longer display durations, however, people are able to resolve the difficulties caused by dividing attention between modalities and so bimodal accuracy closely follows unimodal accuracy. The lack of difference between unimodal and bimodal performance suggests that 33ms exposure offers enough opportunity to suppress the exogenous attentional grab of bimodal stimuli (Johnson & Zattore, 2005) and allow the bimodal property to be processed only on the target modality that is the subject of endogenous attention. In short, this experiment's findings suggest that, although the conceptual processing of a bimodal property such as *big* or *fluffy* automatically attempts representation on both sight and touch, endogenous attention can effect a "quick and dirty" perceptual simulation on just one of those modalities.

General Discussion

This study investigated an issue largely neglected in conceptual processing research – the representational nature of multimodal properties – and specifically asked whether bimodal properties must always be represented bimodally (i.e., an obligatorily complete simulation) or whether a partial unimodal representation can sometimes suffice (i.e., a "quick and dirty" simulation). Results showed that both perspectives were partly right: processing bimodal properties such as *fluffy* or *round* automatically attempts representation on both visual and haptic component modalities, but conscious attention on one of these modalities can selectively produce a unimodal representation. These findings support the embodied view that the conceptual system utilises modality-specific perceptual resources (e.g., Barsalou, 1999) and adds novel insights into the role of attentional mechanisms in modality-specific conceptual processing.

In Experiment 1, we found people were faster to verify a unimodal property that used exactly the same modality as its predecessor (e.g., visual→visual) than a bimodal property that only shared one component modality with its predecessor (e.g., visual→visuohaptic). Simply processing stimuli as they arrive allows attention to be exogenously grabbed by whatever perceptual modality is needed, and so bimodal properties incur a processing cost when an additional modality must be engaged. For example, when the visual modality is already active, verifying a bimodal property such as *blade:jagged* requires additional attention to be allocated to the haptic modality because both sight and touch are involved in its perceptual simulation. Consciously focusing endogenous attention on a particular perceptual modality, on the other hand, limits the ease with which incoming stimuli can grab attention. Experiment 2 showed that the processing of modality-specific information is rapid and automatic, with performance differences between unimodal and bimodal words after just 17ms exposure. For example, people found it more difficult to detect bimodal words like *fluffy* as pertaining to the sense of vision than unimodal words like *colourful* because the haptic component of *fluffy* exogenously grabbed attentional control. Endogenous attention on the visual modality, however, was able to suppress this unwanted haptic simulation if the word was displayed for longer (33ms onwards). When bimodal perceptual stimuli are presented, endogenous attention on one modality can suppress activation in the cortex corresponding to the unattended modality (Johnson & Zattore, 2005). The current findings suggest a similar mechanism operates in the conceptual processing of bimodal properties.

The modality engagement cost we report in the present paper is different to the modality switching cost found for unimodal processing of perceptual (Spence et al., 2001; Turatto et al., 2004) and conceptual (Marques, 2006; Pecher et al., 2003; van Dantzig et al., 2008) stimuli because it does not involve decoupling attention from the original modality. This raises the question of whether modality switching costs are actually composed of two summed costs: the time required to decouple attention from the first perceptual modality plus the time required to engage attention with a second modality. Our findings suggest that the modality engagement, at least, incurs a sizeable processing cost. Future research will investigate whether modality decoupling is similarly costly in processing terms.

Selective endogenous attention in perception is an efficient means of filtering the complex stream of incoming information according to task demands. But is there any such efficiency benefit for selective attention in conceptual processing? Or is the role of attentional mechanisms in conceptual processing merely an artifact of the conceptual system co-opting the perceptual system for representational purposes? We would suggest that there are indeed some advantages in selective processing of certain aspects of conceptual information. For example, if selective attention allows people to create a partial "quick and dirty" perceptual simulation when task demands do not require anything more complex, it frees up cognitive resources for other tasks. Whether or not this benefit emerged from the adaptation of the attentional system to offline processing, or whether it is a happy accident of shared neural substrate between perception and conception, remains an open question.

Acknowledgments

This work was funded by grant RES-000-22-2407 from the UK Economic and Social Research Council to the first author. Thanks to Felix Dreyer for help with data collection.

References

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal detection and object recognition. *Experimental Brain Research, 166*, 559-571.

Balota, D. A., Yap, M. J., Cortese, M.J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences 22*, 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617–645.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behaviour, 8*, 240-248.

Connell, L. (2007). Representing object colour in language comprehension. *Cognition, 102*, 476-485.

Connell, L., & Lynott, D. (2009). Hard to put your finger on it: Haptic modality disadvantage in conceptual processing. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society.*

Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science, 14*, 14–18.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*, 162–169.

Fodor, J. A. (1975). *The language of thought.* New York: Crowell.

Foxe, J. J., Simpson, G. V., Ahlfors, S. P., & Saron, C. D. (2005). Biasing the brain's attentional set: I. Cue driven deployments of intersensory selective attention. *Experimental Brain Research, 166*, 370–392.

Gibbs, R. W. (2003). Embodied experience and linguistic meaning. *Brain and Language, 84*, 1-15.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20,* 1–55.

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *Journal of Neuroscience, 26*, 4917-4921.

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage, 32*, 906-912.

Johnson, J. A., & Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events: Behavioral and neural correlates. *Cerebral Cortex, 15,* 1609–1620.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, *94*, B79–B89.

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language, 39*, 170-210.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Locke, J. (1975). *An essay concerning human understanding*. In P.H. Nidditch (Ed.), Oxford: Clarendon Press. (Original work published 1690).

Lynott, D. & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods, 41*, 558-664.

Marques, J. M. (2006). Specialization and semantic organization: Evidence for multiple semantics linked to sensory modalities. *Memory & Cognition, 34*, 60–67.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*, 402-407.

Newell, A. & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Newman, S. D., Klatzky, R. L., Lederman, S. J., Just, M. A. (2005). Imagining material versus geometric properties of objects: An fMRI study. *Cognitive Brain Research, 23*, 235-246.

Pecher, D., Zeelenberg, R., & Barsalou, L.W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychological Science, 14,* 119-124.

Pecher, D., & Zwaan, R. A. (2005). Introduction to grounding cognition. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: the role of perception and action in memory, language, and thinking*. Cambridge: CUP.

Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion In Neurobiology, 11*, 505-509.

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*, 2802-2810.

Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition, 32*, 244-259.

Spence, C., Nicholls, M. E. R., & Driver, J. (2000). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics, 63*, 330-336.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization and memory*. New York: Academic Press.

Turatto, M., Galfano, G., Bridgeman, B., & Umiltà, C. (2004). Space-independent modality-driven attentional capture in auditory, tactile and visual systems. *Experimental Brain Research, 155*, 301-310.

van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L.W. (2008). Perceptual processing affects conceptual processing. *Cognitive Science, 32,* 579-5.