Linguistically modulated perception and cognition: the label-feedback hypothesis

Gary Lupyan*

Department of Psychology, University of Wisconsin–Madison, Madison, WI, USA

How does language impact cognition and perception? A growing number of studies show that language, and specifically the practice of labeling, can exert extremely rapid and pervasive effects on putatively non-verbal processes such as categorization, visual discrimination, and even simply detecting the presence of a stimulus. Progress on the empirical front, however, has not been accompanied by progress in understanding the mechanisms by which language affects these processes. One puzzle is how effects of language can be both deep, in the sense of affecting even basic visual processes, and yet vulnerable to manipulations such as verbal interference, which can sometimes nullify effects of language. In this paper, I review some of the evidence for effects of language on cognition and perception, showing that performance on tasks that have been presumed to be non-verbal is rapidly modulated by language. I argue that a clearer understanding of the relationship between language and cognition can be achieved by rejecting the distinction between verbal and non-verbal representations and by adopting a framework in which language modulates ongoing cognitive and perceptual processing in a flexible and task-dependent manner.

Keywords: language and thought, perception, categorization, labels, top-down effects, linguistic relativity, Whorf

INTRODUCTION

Are the faculties of perception, categorization, and memory – capacities humans share with other animals – shaped by the human-specific faculty of language? Does language simply allow us to communicate about our experiences, albeit with much greater flexibility compared to other animal communication systems? Or, does language also transform cognition and perception, allowing humans to access and manipulate mental representations in novel ways? This question has been of longstanding interest to philosophers (see Lee, 1996 for a historical review), and goes to the core of understanding human cognition (Carruthers, 2002; Spelke, 2003). Many have speculated on the transformative power of language on cognition (James, 1890; Whorf, 1956; Cassirer, 1962; Vygotsky, 1962; Dennett, 1994; Clark, 1998). A growing number of studies show that language can exert rapid and pervasive effects on putatively non-verbal processes. For contemporary reviews of the "language and thought debate" (see Gumperz and Levinson, 1996; Gentner and Goldin-Meadow, 2003; Gleitman and Papafragou, 2005; Casasanto, 2008; Borodit-sky, 2010; Wolff and Holmes, 2011). Despite progress on the empirical front showing apparent effects of language in domains ranging from basic perceptual tasks such as color perception (see below), motion perception (Meteyard et al., 2007), visual search (Lupyan, 2008a), and simple visual detection (Lupyan and Spivey, 2010a), to categorization in infancy (e.g., Waxman and Markow, 1995) and adulthood (Lupyan et al., 2007), to recognition memory (e.g., Lupyan, 2008b; Fausey and Borodit-sky, 2011) and relational thinking (Loewenstein and Gentner, 2005), there has been a lack in progress on the theoretical front. In this work, I will argue that significant theoretical progress can be made by taking a interactive-processing perspective (e.g.,

McClelland and Rumelhart, 1981) on the question of the relationship between language and thought.

The paper is divided into four parts: First, I discuss an apparent paradox that has stymied both critics and proponents of the "language and thought" research program (Gleitman and Papafragou, 2005; Wolff and Holmes, 2011): how can effects of labels be both deep, apparently affecting basic even perceptual processing, and yet be easily disrupted by manipulations such as verbal interference? Second, I present a proposed solution to the paradox in the form of the *label-feedback hypothesis,* on which the classic distinction between verbal and non-verbal processes is replaced with an emphasis on the role of language as a *modulator* of a distributed and interactive system (see also Kemmerer, 2010). Third, I review some empirical data from the domains of visual perception, categorization, and memory, that are difficult to reconcile with common assumptions in contemporary literature on language and thought, but are naturally accommodated by the label-feedback hypothesis. Finally, I briefly discuss the implications of taking an interactive-processing on the question of linguistic relativity.

THE FRAGILITY OF LINGUISTIC EFFECTS ON COGNITION AND PERCEPTION: A PARADOX?

One domain that has received a considerable amount of attention in the language and thought literature is that of putative effects of language on color categorization and color perception. Shortly after the posthumous publication of Benjamin Lee Whorf's essays (Whorf, 1956), the philosopher Max Black published a critique in which he commented on Whorf's now-famous passage: "We dissect nature along lines laid down by our native languages. Language is not simply a reporting device for experience but a

defining framework for it." (p. 213). Black remarked that Whorf's word-choice engendered confusion:

> "To dissect a frog is to destroy it, but talk about the rainbow leaves it unchanged. The case would be different if it could be shown that color vocabularies influence the perception of colors, but where is the evidence for that?" (Black, 1959, p. 231).

There is now a large and rapidly increasing number of findings showing just such effects: cross-linguistic differences in color vocabularies can cause differences in color categorization with concomitant effects on color memory and, indeed, color perception (Davies and Corbett, 1998; Davidoff et al., 1999; Roberson et al., 2005, 2008; Daoutis et al., 2006; Winawer et al., 2007; Thierry et al., 2009). For example, Winawer et al. (2007) presented English and Russian speakers with color swatches showing different shades of blue. Russian, unlike English, lexicalizes the category blue with two basic-level terms: "siniy" for darker blues and "goluboy" for lighter blues[1]. The subjects were asked to perform a simultaneous xAB task, deciding as quickly as possible whether a top color (x) exactly matched a color on its left (A) or on its right (B). The categorical relationship between the color x and the non-matching color was varied such that, for Russian speakers, the two colors were sometimes in the same lexical category and sometimes in different categories. All colors were in the "blue" category for English speakers. The results showed a categorical perception effect for Russian speakers only, as evidenced by slower reaction times (RTs) on within-category than between-category trials.

A possible mechanism by which cross-linguistic differences in categorical color perception can be produced is gradual perceptual warping caused by learning. On this account, long-term experience categorizing the color spectrum using language gradually warps the perceptual representations of color resulting in more similar representations of colors in the same category (i.e., those labeled by a common term) and/or less similar representations of colors grouped into distinct categories (i.e., those labeled by distinct terms). That is, learning and using words such as "siniy" and "goluboy" provides categorization practice that results in the gradual representational separation of the parts of the color spectrum to which the labels are applied. Different labeling patterns (using the generic term "blue") are therefore predicted to produce different patterns of discrimination across the color spectrum. This standard account of learned categorical perception (Goldstone, 1994, 1998; Goldstone and Barsalou, 1998) has been applied to the color domain, and as predicted, training individuals on a new color boundary can induce categorical perception (Ozgen and Davies, 2002).

On the perceptual learning account, once labels have provided sufficient categorization training for perceptual warping to occur, the warped perceptual space remains. And yet, a growing number of studies show that when participants are placed under conditions of verbal interference that is presumed to decrease the on-line influence of language, cross-linguistic differences seem to disappear. For example, Winawer et al. (2007) found that when Russian-speaking subjects were placed under verbal interference, within-category comparisons no longer took longer than between-category comparisons[2] (see also Roberson and Davidoff, 2000; Pilling et al., 2003; Gilbert et al., 2006; Drivonikou et al., 2007; Wiggett and Davies, 2008; cf. Witzel and Gegenfurtner, 2011). This bleaching effect of verbal interference is seen in other domains as well. For example, English and Indonesian-speaking monolinguals show memory patterns consistent with their language: better memory for different tenses in English than Indonesian, which does not require morphological tense markers (Boroditsky, 2003). The difference in memory between Indonesian and English speakers was attenuated with verbal interference.

Further evidence of the transient nature of effects of language on cognition comes from studies of the consequences of language impairments on putatively non-verbal processes (putative in the sense that if some cognitive process can be shown to be affected by language, is that process still non-verbal?). The logic as articulated by Goldstein (1924/1948) is that if language is involved in not only communicating thoughts but somehow "fixating" them, then language impairments should produce cognitive impairments. Indeed, as noted by Goldstein (see Noppeney and Wallesch, 2000 for review), individuals with aphasia appear to have a number of deficits that appear on their surface to have little to do with language. A particular difficulty is posed by categorization tasks requiring grouping on a particular dimension. In an effort to further distil this deficit, Cohen and colleagues concluded that "...aphasics have a defect in the analytical isolation of single features of concepts" (Cohen et al., 1980, 1981), yet are equal to control subjects "when judgment can be based on global comparison" (Cohen et al., 1980). In their examination of the anomic patient LEW, Davidoff and Roberson reached a similar conclusion, arguing that when a grouping task requires attention to one category while abstracting over others, LEW is "without names to assist the categorical solution." (Davidoff and Roberson, 2004, p. 166). In a recent study designed to examine the categorization-aphasia link more exhaustively, Lupyan and Mirman (under review) found that a group of patients with aphasia (selected on the basis of having varying levels of naming impairments) were specifically impaired on a categorization task requiring focusing on a specific dimension, e.g., selecting all the pictures of red objects from color images of familiar objects. The patients were selectively impaired on trials requiring categorizing by specific isolated dimensions, but had performance similar to controls on trials which required more global categorization such selecting objects typically found in a laundry room. Critically, the patients' impairment on this non-verbal task was best predicted by their performance on a standard confrontation naming test (PNT; Roach et al., 1996). Naming performance continued to predict categorization performance controlling for semantic impairments and general location

---

[1]There has been some confusion regarding the primacy of color terms such as "navy" in English. The crucial cross-linguistic difference here lies not so much in the frequency, ambiguity, or accessibility of the term "siniy" in the minds of Russian speakers versus the term "navy" in the minds of English speakers. Rather, the difference lies in the presence of a generic term "blue" in English and the lack of such a term in Russian. An inverse situation occurs in the domain of body part terms: The Russian word "ruka" (arm including the hand) has no corresponding generic term in English.

[2]Verbal interference actually *reversed* the usual categorical perception effect with within-category matching now taking less time than between-category matching (see also Gilbert et al., 2006 for a similar reversal). This odd pattern of results awaits an explanation.

of the lesion. These data do not suggest that successful categorization *depends* on an intact naming abilities, but that the two are intertwined such that naming impairments contribute to categorization impairments, particularly when the task requires isolating specific dimensions and cannot be accomplished through overall similarity[3].

Convergent evidence for the interactive relationship between language and categorization comes from a study in which I used verbal interference to attempt to simulate some of the categorization impairments that have been previously reported to be concomitant with naming impairments. Lupyan (2009) tested college undergraduates on an odd-one-out task in which participants were presented with triads of pictures or words and had to select the one that did not belong on some specific criterion, such as real-world size. On other trials, the task required selecting a picture or word that did not belong based on more thematic or functional relationship.

When tested with this task, the anomic patient LEW was selectively impaired in making size and color, but not function/thematic judgments (Experiment 7, Davidoff and Roberson, 2004). Healthy subjects undergoing verbal (but not visual) interference of the same type as used to bleach effects of language on color perception, showed a performance profile very similar to that of the anomic patient LEW.

THE PARADOX DISTILLED

The paradox then is this: if effects of language on perceptual processing are "Whorfian" in the sense of changing the underlying perceptual space (i.e., warping perception), then how can the space be "unwarped" so easily? Similarly, if language affects categorization by providing additional training opportunities, why would language impairments produce categorization impairments? In a recent debate hosted by *The Economist* on the proposition "The language we speak shapes how we think," Lila Gleitman remarked on the interpretation of the types of effects of language on color discussed above with the following observation:

> . . .here is the usual finding: "Disrupting people's ability to use language while they are making colour judgments eliminates the cross-linguistic differences." What is puzzling is why [Boroditsky] thinks this is a "pro" argument. In fact, it is the "con" argument, namely that the underlying structure and content of "thought" and "perception" are unaltered by palpable and general differences in language encoding (Gleitman, 2010).

This argument in one form or another has been invoked by a number of critics (Gleitman and Papafragou, 2005; Dessalegn and

Landau, 2008; Li et al., 2009). The reasoning seems to be that if linguistic influences on categorization and perception can be removed so easily (or conversely, appear after only a brief training period, e.g., Boroditsky, 2001; cf. January and Kako, 2007), then they must be superficial. Put another way, according to this critique, if an influence of language on, for example, color perception can be disrupted via a verbal manipulation, does this not mean that language was affecting a verbal process all along and therefore the effect is of language on language rather than language on perception? This rationale appears to rest on two assumptions: First, language is assumed to be a medium (a "transparent medium" even, H. Gleitman et al., 2004, p. 363). On this view, words *map onto* concepts, which are, by definition, independent of words (e.g., Gopnik, 2001; Snedeker and Gleitman, 2004; Gleitman and Papafragou, 2005). The second assumption is of a strict separation between verbal and non-verbal processing, and consequently between verbal and non-verbal representations. (This assumption is also evident in the "thinking for speaking" framework articulated by Slobin, 1996). Accepting these two assumptions, it is indeed puzzling how the sorts of effects of language on color categorization and perception discussed above can be simultaneously pervasive and fragile: if language alters concepts, should not these altered concepts persist regardless of how language is deployed on-line?

The *label-feedback hypothesis* is an attempt to reconcile this apparent paradox of how effects of language can be so vulnerable to interference while at the same time exerting apparently pervasive influence on basic perceptual processing (e.g., see Liu et al., 2009; Thierry et al., 2009; Mo et al., 2011 for effects of language on early visual processing in the domain of color perception). As I will argue, the reason these effects are sensitive to manipulations such as verbal interference is that many language exerts effects on perception by modulating ongoing perceptual processing on-line. This modulation, insofar as it is rapid and automatic, constitutes a change in the functional structure and content of "thought" referred to by Gleitman because language and thought are part of a distributed interactive system. As articulated by Whorf himself:

> Any activations [of the] processes and linkages [which constitute] the structure of a particular language. . . once incorporated into the brain [are] all linguistic patterning operations, and all entitled to be called thinking (Whorf, 1937, pp. 57–58 cited in Lee, 1996, p. 54).

A note of caution is in order: Viewing language as a part of an inherently interactive system with the capacity to augment processing in a range of non-linguistic tasks does not mean that performance on every task or representations of every concept are under linguistic control. Rather, the argument is that learning and using a system as ubiquitous as language has the potential to affect performance on a very wide range of tasks. A fruitful research strategy may be therefore to investigate what classes of seemingly non-verbal tasks are influenced by language (and which are not), and on what classes of tasks cross-linguistic differences yield consistent differences in performance. This point is expanded below in the Section "Implication of the Label-Feedback Hypothesis for 'Language and Thought' Research Program."

---

[3] Kemmerer et al. (2010) tested a large and very diverse group of brain-damaged patients on a battery of tasks including naming, word–picture matching, and attribute selection (e.g., deciding which picture depicts an action that is most tiring). The deficit profile was a complex one with patients showing virtually every pattern of dissociation between the tasks. Interestingly, naming performance was significantly correlated with performance on the picture-attribute task, but not at all with the picture-comparison task. It remains to be determined if these patterns of association reflect differences in the degree to which the tasks require selection of specific dimensions versus reliance on global association (Lupyan, 2009; Lupyan et al., under review; see also Sloutsky, 2010).

FROM PERCEPTION TO CATEGORIZATION TO VERBAL LABELS AND BACK AGAIN: THE LABEL-FEEDBACK HYPOTHESIS

Perceiving a stimulus as meaningful depends on (perhaps even requires) representing the stimulus in terms of a larger class. Consider that even a task as simple as deciding whether two "identical" objects, presented simultaneously in different locations are the "same" requires the observer to ignore that they are different by virtue of their positions. In the short-story *Funes the Memorius*, Borges describes a man incapable of categorization:

> "It was not only difficult for him to understand that the generic term dog embraced so many unlike specimens of differing sizes and different forms; he was disturbed by the fact that a dog at three-fourteen (seen in profile) should have the same name as the dog at three-fifteen (seen from the front)" (Borges, 1942/1999, p. 136).

Naming both of the above instances of dogs as a "dog" requires representing both as members of the same class – one which is associated with the label "dog." Clearly, naming depends on categorization. But does language, and the act of naming in particular, play an active role in the categorization process itself? In this section, I argue that names (verbal labels) play an active role in perception and categorization by selectively activating perceptual features that are diagnostic of the category being labeled. Critically, although this top-down augmentation of perceptual representations by language is likely to be in play, to some degree, even during passive vision, it can be up- or down-regulated through linguistic manipulations such as brief verbal training/verbal priming and verbal interference.

On the present view, categorization is the *process* by which detectably different (i.e., non-identical) stimuli come to be represented as identical, in some respect (see Lupyan et al., under review for discussion). Categorizing a stimulus thus involves changing its representation. However, placing two objects into the same category does not, logically, imply a change to their perceptual representations which on some accounts are impenetrable to the influence of conceptual categories (e.g., see Pylyshyn, 1999; Macpherson, 2012). In groundbreaking work, Goldstone and colleagues (Goldstone, 1994; Goldstone and Hendrickson, 2010 for review) showed that the categorization process alters perception itself. In a typical study, participants were trained to respond to items that parametrically vary on one or more dimensions with some belonging to "Category A" and others to "Category B" (Goldstone, 1994), or to discriminate between individuals belonging to a "club" and those not belonging (Goldstone et al., 2003). Following this training, visual discrimination ability is assessed (while controlling for effects of categorization from those of mere exposure[4]) and compared to visual discrimination prior to training or to discrimination following a control training task. A significant change in the perception of dimensions relevant to the categorization task suggests that categorization experience altered the visual appearance of the items being categorized. Rather than just being *mediated* by the category responses (i.e., participants judging two

stimuli as more similar by virtue of their belonging to the same category), the experience of categorization was found to warp perception, sensitizing some regions of perceptual space (e.g., those close to the category boundary; Goldstone, 1994). This warping effect affected the relationship between trained and novel stimuli– an effect argued by the authors to be incompatible with an effect of categorization on the decision process only (Goldstone et al., 2001).

Goldstone and colleagues' work on perceptual warping and learned categorical perception (e.g., Goldstone et al., 2001) provides a potential mechanism by which language may augment categorization. Because each act of naming is an act of categorization, learning to label some colors "green" and others "blue," provide a type of category-training which, over time, is expected to help pull apart the representations and resulting in decreased representational overlap between the two classes of stimuli. But how can one reconcile the perceptual warping process with the fragility of language-modulated effects outlined above?

The *label-feedback hypothesis* proposes that language produces *transient* modulation of ongoing perceptual (and higher-level) processing. In the case of color, this means that after learning that certain colors are called "green," the perceptual representations activated by a green-colored object become warped by top-down feedback as the verbal label "green" is co-activated. This results in a temporary warping of the perceptual space with greens pushed closer together and/or greens being dragged further from non-greens. Viewing a green object becomes a hybrid visuo-linguistic experience. Knowing that some colors are called green means that our everyday experiences of seeing become affected by the verbal term, which in turn makes the visual representation more categorical. This modulation can be increased – up-regulated – by activating the label to a greater than normal degree as when a participant hears a verbal label prior to seeing a visual display. Conversely, verbal interference is one way to down-regulate the activation of labels leading to reduced influences effect of language on "non-verbal" processing.

To illustrate how language can affect perceptual representations, consider a task in which subjects view briefly presented displays of the numerals 2 and 5, with several from each category presented simultaneously. The task is to attend to just the 5s and to press a button as soon as a small dot appears around one of the numerals. The more selectively participants can attend to the 5s, and just the 5s, the better they ought to perform. Before some trials, participants actually hear the word "five." This cue constitutes entirely redundant information because participants already know what they should do on each trial. The task of attending to the 5s remains constant for the entire 45-min experiment, thus the word "five" tells them nothing they do not already know. Yet, on the randomly intermixed trials on which they actually hear the word, participants respond more quickly (and, depending on the task, more accurately; Lupyan and Spivey, 2010b). This type of facilitation occurs even when the items are seen for only 100 ms, a time too brief to permit eye movements. Similar effects are obtained with more complex items such as pictures of chairs and tables. The linguistic facilitation is also transient. If too much time is allowed to elapse between the label and the onset of the display (more than ~1600 ms. in this case), no facilitation is seen. In fact, obtaining

---

[4]See Folstein et al. (2010) for a recent study of the role of mere exposure to exemplars on subsequent category learning.

such effects is only possible if hearing a word has a *transient* effect on visual processing; if the facilitation due to hearing a word carried through the entire experiment, the difference between the intermixed label and no-label trials would quickly vanish. Yet the difference persisted, in most cases through the entire experiment lasting for hundreds of trials (Lupyan and Spivey, 2010b) which was only possible if hearing a label affected perceptual processing in a transient, on-line manner.
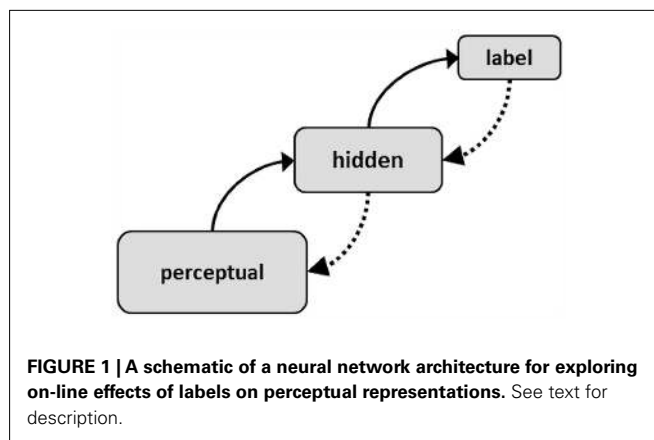
According to the label-feedback hypothesis, hearing the word "five" activates visual features corresponding to 5s, transiently moving the representations of 5s and 2s further apart, while making the perceptual representations of the various 5s on the screen more similar, and thereby easier to simultaneously attend. Notice that this task did not *require* identification or naming. Verbal labels were certainly not needed to see that 2s and 5s are perceptually different. Yet, overt language use – a hypothesized "up-regulation" of the linguistic modulation normally takes place during perception – had robust effects on perceptual processing.

In other studies, my colleagues and I have shown that hearing similarly redundant words can improve performance in a pop-out visual search (Lupyan, 2008a) and improves search efficiency in more difficult search tasks (Lupyan, 2007). Hearing a label can even make an invisible object visible. Lupyan and Spivey (2010a) showed that hearing a spoken label increased visual sensitivity (i.e., increased the d' in a simple object detection task: simply hearing a label enabled participants to detect the presence of briefly presented masked objects which were otherwise invisible (see also Ward and Lupyan, 2011 who showed that hearing labels can make visible stimuli suppressed through continuous flash suppression).

A SIMPLE MODEL OF ON-LINE LINGUISTIC EFFECTS ON PERCEPTUAL REPRESENTATIONS
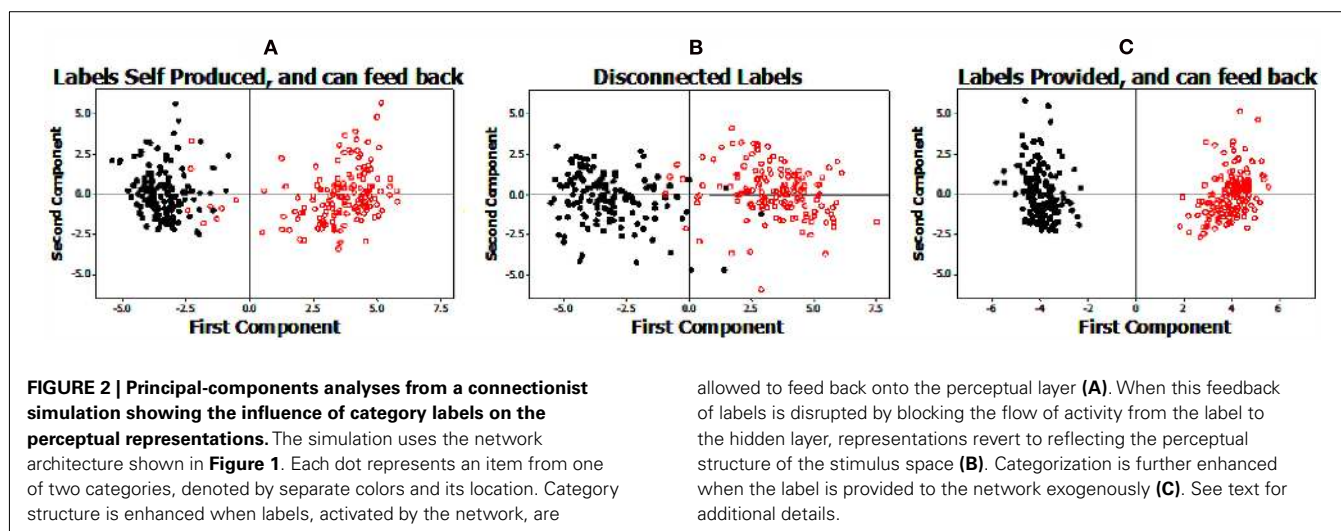
A simple model implementing the idea of labels as modulators of lower-level representations is shown in **Figure 1**. The model is implemented as a fully recurrent neural network (Rumelhart et al., 1986). Solid lines denote feedforward connections and dashed lines denote feedback connections. In this implementation, the perceptual layer is provided with a feature-based input of a current object. The model is trained on two categories instantiated as a distortion from one of two category prototypes (for a more detailed description, see Lupyan, in press). Let us arbitrarily call one category "chairs" and the other "tables." During training, the model learns to produce names, e.g., to produce the label "chair" given one of the chairs, and comprehend names: given the label "chair," it activates properties characteristic of chairs. Due to the one-to-many mapping between-category labels and category exemplars the network cannot know which particular object is being referred to when presented with just the category label. It is this one-to-many mapping that allows the network to generalize and make inferences to un-seen properties. Because some properties (e.g., having a back) are more closely correlated with category membership than other properties (e.g., being brown) the category labels become more strongly associated with properties that are typical or diagnostic of the denoted category, and dissociated from properties that are not diagnostic of the category.

Following this training, we can examine what happens to representations of category exemplars when the label is allowed to

FIGURE 1 | A schematic of a neural network architecture for exploring on-line effects of labels on perceptual representations. See text for description.

feed back on the activity in the perceptual layers. **Figure 2** shows a principal-components analysis (PCA) of the perceptual representations of exemplars from two categories learned in the context of labels. In **Figure 2A**, the label is endogenous to the network. The network produces the label itself in response to the perceptual input, and the label is then allowed to feedback to affect the visual representations. This corresponds to what is hypothesized to occur in the default case: perceptual representations are modulated on-line by verbal labels via top-down feedback. In **Figure 2B** the labels are prevented from affecting the representations on-line by disabling the name-to-hidden-layer connections. The category separation observed in this PCA plot is due entirely from bottom-up perceptual differences between the two categories. This situation is logically equivalent to a verbal interference condition (although in reality, label activations are only one kind of top-down influences affecting visual processing). In **Figure 2C** the labels are provided exogenously to the network along with the perceptual input. This case is equivalent to the label trials in the experiment described above (Lupyan and Spivey, 2010b). Much clearer category separation is observed. Insofar as correct categorization depends on representing similarities between exemplars, it is facilitated by the influence of labels. There is a cost to this enhanced categorization. The more categorical representations produced by the labels are beneficial for categorization-type tasks, but reduce accuracy in the representation of the idiosyncratic properties of individual exemplars. Indeed, when participants are shown pictures of chairs and tables and are asked to label some of them with the category labels ("chair" and "table"), they show poorer subsequent recognition of items that they labeled (Lupyan, 2008b).

The simple model shown in **Figure 1** can be extended to help understand how label-feedback may affect performance in categorization tasks such as those requiring the isolation of specific dimensions – impaired in aphasia and under conditions of verbal interference. Feedback from the activation of a dimensional label such as "size" or "color" is predicted to have the same kind of cohering effect – facilitating the grouping of objects by their dimensions. This role of labels in realigning representations is one way to explain the facilitatory effect of labels in relational reasoning (Kotovsky and Gentner, 1996; Ratterman and Gentner, 1998; Gentner and Loewenstein, 2002).

FIGURE 2 | Principal-components analyses from a connectionist simulation showing the influence of category labels on the perceptual representations. The simulation uses the network architecture shown in **Figure 1**. Each dot represents an item from one of two categories, denoted by separate colors and its location. Category structure is enhanced when labels, activated by the network, are allowed to feed back onto the perceptual layer **(A)**. When this feedback of labels is disrupted by blocking the flow of activity from the label to the hidden layer, representations revert to reflecting the perceptual structure of the stimulus space **(B)**. Categorization is further enhanced when the label is provided to the network exogenously **(C)**. See text for additional details.

ON-LINE VERSUS SUSTAINED EFFECTS OF LABELS ON PERCEPTION AND COGNITION

The demonstrations of the effects of labels on perceptual processes discussed above focused on transient effects such as those produced by overtly hearing a category name. Finding that language influences visual processing, but only in the few seconds immediately after we hear a word, while curious, is clearly of limited theoretical import. The key assumption in such experiments (e.g., Lupyan, 2007, 2008a; Lupyan and Spivey, 2010a,b) is that overt presentation of labels (or, as shown by Lupyan and Swingley, in press, language production in the form of self-directed speech) can exaggerate what is hypothesized to be the normal on-line influence of language on task performance. Verbal interference, on this view, is a comparable *down*-regulation of language. Such manipulations can shed light on the "normal" function played by language in cognition and perception. In this section I briefly review some findings suggesting that perceptual processes are influenced rapidly and automatically by language. That is, the normal state in adults is closer to **Figure 2A** in which automatically activated labels modulated perceptual representations, than **Figure 2B** in which perceptual representations mapped onto category labels, but were impermeable to linguistic feedback.

Consider a task in which an observer is presented with two stimuli and needs to determine, as quickly as possible, whether they are visually identical. Naturally, the more subtle the differences, the more difficult the judgment. Consider now the letter pairs B-b and B-p. The letters in each pair are visually equidistant, but *conceptually* B-b are more similar than B-p. Despite this conceptual difference, reaction times (RTs) for B-p and B-b judgments are equivalent when the two letters are presented simultaneously (Lupyan, 2008a). However, when the second letter is presented $\geq$150 ms. after the first (with the first still present on the screen), B-b judgments become more difficult than B-p judgments (Lupyan et al., 2010b). We claimed this occurs because during this delay, the representation of the first letter becomes augmented by its conceptual category, rendering "B" more similar to "b" and more distinct from "p." This effect is further enhanced when subjects actually *hear* the letter name (Lupyan, 2008a), i.e., up-regulating language appears to exaggerate the categorical perception effect.

Although these results show basic perception to be dynamically influenced by conceptual categories, the results do not directly address the role played specifically by the category names. This question is beginning to be addressed using the work described below.

Using fMRI, Tan et al. (2008) showed that in a same-different color discrimination task, similar to the simultaneous condition of the B-p task described above, Wernicke's area (posterior part of BA 22) showed greater activity for easy-to-name versus hard-to-name colors suggesting its automatic activation in this non-verbal task. Although the authors attempted to interpret the selective activity in terms of the effects of language on visual discrimination, clearly, no such causal attribution of the neural activity can be made; its activity may be consistent with activation of color names, but does not indicate that this activity affects visual processing. On the current account, such causal effects are exactly what is expected, with category effects in vision emerging (in some part) due to activation of category *names*. One way to test this prediction is by disrupting the activity and measuring its outcome. In a recent study, we administered TMS to Wernicke's area while participants performed the B-p/B-b same-different task (Lupyan et al., in preparation). Insofar as slower responses to B-b relative to B-p are the result of label-feedback, disrupting this activity should eliminate the RT difference between B-p and B-b stimuli. The results showed that an inhibitory stimulation regime completely eliminated the RT difference between responding "different" to B-p and B-b letter pairs. Control stimulation to the vertex had no effect. To my knowledge, no theory of visual processing classifies Wernicke's area (posterior superior temporal gyrus) as "visual." That disruption of activity in this region alters behavioral responses on a visual task supports the hypothesis that the effects of conceptual categories (here, letter categories) on visual processing are subserved in part by a classic language area, stimulation of which possibly disrupts its usual modulation of neighboring posterior regions of the ventral visual pathway.

The transient effects of labels on perception described above may be special cases of normally occurring top-down modulations of vision by linguistic, contextual and other "cognitive" factors. An example of such modulations of a more sustained nature can be

seen when one examines the role of meaningfulness in vision. As might be expected, it is easier to recognize and discriminate meaningful entities than meaningless ones. For example, it takes about 200 ms. longer to recognize that the items in the pair ꒒/꒒ are physically identical than it does to make the same judgment for ꓑ/ꓑ or ᖯ/ᖯ (Lupyan, 2008a). The stimuli ᖯ and ᖯ differ in meaningfulness, of course, but they also differ in familiarity. We simply have more experience processing ᖯs as compared with ꓑs. In a very simple study, Lupyan and Spivey (2008) used a visual search task in which participants were asked to search for a � among ᒪs (or vice-versa). The stimuli were meaningless and perceptually novel. Some participants were explicitly told at the start of the experiment that the shapes should be thought of as rotated 2 and 5s. This simple instruction dramatically improved overall RTs and led to shallower search slopes, indicating more efficient visual processing. The effect of construing a stimulus as meaningful (and in this case, associating it with a named category) produced a sustained effect in the sense that once induced, the facilitation persists, an effect reminiscent of the well-known hidden Dalmatian in a piecemeal image, which once known to be present in the image, cannot be "un-seen" (Gregory, 1970; see also Porter, 1954). Arguably, such effects are *also* on-line effects (see also Bentin and Golland, 2002). The degree to which such conceptual effects on visual processing are truly linguistic requires further investigation and neurostimulation techniques such as TMS and tDCS will potentially prove useful (Lupyan et al., 2010a).

These results potentially inform the findings of cross-linguistic differences in early ERPs in response to changing colors. Thierry et al. (2009) found that Greek speakers who, like Russian speakers, have separate words for light and dark blues, showed a greater visual mismatch negativity – an early component showing condition-differences starting at ∼160 ms that has been used to index automatic, and arguably preattentive change detection – when presented with color changes that spanned the lexical boundary. The authors found some differences in the P1 component as well. On the one hand, such differences in early visual processing may be viewed as consequences of long-term perceptual warping produced by language (or perhaps other cultural factors). This account however, would be at a loss to explain why in other studies verbal interference can eliminate cross-linguistic differences on behavioral measures of categorical color perception. An alternative account is that viewing colors automatically activates their names that warp perceptual representations on-line. The observed effects on early perception are thus evidence not of a permanent change in bottom-up processing, but rather of a sustained top-down modulation possibly induced by activation of the color names during the task[5].

THE NEURAL PLAUSIBILITY OF LANGUAGE-MODULATED PERCEPTION

Understanding the word "chair" is clearly a more complex process than detecting the presence of a shape in a visual display or

determining which of two color swatches matches a third. How can a complex "high-level" process influence low-level and much more rapid processes such as simple detection? This would indeed be puzzling if the brain were a feedforward system. It is not. Neural processing is intrinsically interactive (Mesulam, 1998; Freeman, 2001). As eloquently argued in a prescient paper by Churchland et al. (1994), the brain is only grossly hierarchical: sensory input signals are only a part of what drives "sensory" neurons, processing stages are not like assembly line productions, and later processing can influence earlier processing (p. 59). This view has in recent years received overwhelming support (e.g., Mumford, 1992; Rao and Ballard, 1999; Lamme and Roelfsema, 2000; Foxe and Simpson, 2002; Reynolds and Chelazzi, 2004; Gilbert and Sigman, 2007; Kveraga et al., 2007; Mesulam, 2008; Koivisto et al., 2011).

To give two examples from vision of gross violations of hierarchical processing: (1) the "late" prefrontal areas of cortex can at times respond to the presence of a visual stimulus *before* early visual cortex (V2; Lamme and Roelfsema, 2000 for review). (2) The well-known classical receive fields of V1 neurons showing orientation tuning appear to be dynamically reshaped by horizontal and top-down processes. Within 100 ms. after stimulus onset, V1 neurons are re-tuned from reflecting simple orientation features, to representing figure/ground relationships over a much larger visual angle (Olshausen et al., 1993; Lamme et al., 1999).

Effects of verbal labels on vision can be seen as embodying a similar, but more complex type of perceptual modulation as the reshaping of V1 receptive fields. Although the neural loci of these effects are at present unknown, one possibility is that processing an object name initiates a volley of feedback activity to object-selective regions of cortex such as IT (Logothetis and Sheinberg, 1996), producing a predictive signal or "head start" to the visual system (Kveraga et al., 2007; Esterman and Yantis, 2008; Puri and Wojciulik, 2008). On several theories of attention (e.g., biased competition theory of Desimone and Duncan, 1995), these predictive signals would enable neurons that respond to the named object to gain a competitive advantage (see also Vecera and Farah, 1994; Kramer et al., 1997; Deco and Lee, 2002; Kravitz and Behrmann, 2008). Given feedback from object-selective cortical regions, winning objects can bias earlier spatial regions of visual cortex.

LABELS AND STIMULUS TYPICALITY

The two-dimensional projection of the perceptual representations shown in **Figure 2** hides an interesting interaction between labels and stimulus typicality. Not surprisingly, the network shows basic typicality effects. The correct category label is more quickly and/or strongly activated when the network is presented with a more typical item (i.e., an item having more typical values on dimensions learned by the network to be important). The somewhat counterintuitive consequence is that it is these already typical items that are most affected by labels: the items tend to become even more typical as the network fills in undefined or unknown features with category-typical values. The atypical exemplars (i.e., instances on the periphery of the category), although having the most *potential* to be affected by the label, interact with the label more weakly than the more central exemplars. One can visualize this effect using a magnet metaphor: an object positioned far from a magnet can

---

[5]The authors did not test whether linguistic manipulations such as verbal interference reduce or eliminate the cross-linguistic difference in the visual mismatch negativity, although in a commentary they admit that this would be a natural followup (Athanasopoulos et al., 2009).

be moved a greater distance than an object positioned close to the magnet, but because the magnetic field drops off rapidly with increasing distance, the object farther away is being pulled only weakly and may not move at all. Such a mechanism has similarities to the perceptual magnet effect in perception of phonemes (Kuhl, 1994) and the attractor field model in visual perception (e.g., Tanaka and Corneille, 2007).

Effects of typicality turn out to be quite pervasive: In visual tasks, up-regulating the effect of labels through overt presentation of the label benefits typical category members more than atypical ones. Effects of labels on perceptual processing appear to be stronger for more typical exemplars. For instance, the effect of hearing a label is strong for a numeral in a typical font (5), compared to when it was rendered in a less typical font (⌐; Lupyan, 2007; Lupyan and Spivey, 2010b). In the recognition memory task described above (Lupyan, 2008b) it labeling the typical exemplars led to poorer memory whereas labeling atypical exemplars did not. As a further demonstration that processing an item in the context of its name activates a more typical representation, consider the following two results:

(1) In Experiment 6 of Lupyan (2008b), participants were asked to rate pictures of chairs and lamps on typicality (from very typical to very atypical). The pictures were presented, one at a time, followed by a prompt with the rating scale. The text of the prompt either mentioned the name of the category by name ("chair"/"lamp") or did not (a within-subject manipulation). Participants were instructed to always rate the object's typicality with respect to its category. That is, the task was the same regardless of how the prompt was worded. Yet, participants were more likely to rate the same pictures as more typical when asked, "How typical was that chair" than "How typical was that object," rating the already typical objects *more* typical when referred to by their name (**Figure 3**).

(2) Categories like chair, although comprising concrete objects, are rather fuzzy and do not have formal definitions. In contrast, categories like triangle, *can* be formally defined (Armstrong et al., 1983). All triangles are three-sided polygons and all three-sided polygons are triangles. When queried, all tested participants (18/18) correctly stated this formal definition. When tested on a speeded recognition task, participants showed a typicality/canonicality effect, being faster to recognize isosceles than scalene triangles. This effect, however, was obtained only on trials when participants were cued with the word "triangle." When, on randomly intermixed trials, participants were cued with the phrase "three sides," they were equally fast to recognize isosceles and scalene triangles. According to the label-feedback hypothesis, the category label "triangle" activates a more typical triangle, which in this case appears to correspond to an isosceles/equilateral triangle with a horizontal base. One interesting prediction is that if the label tends to activate a canonical triangle, then referring to a non-canonical triangle explicitly with the word "triangle" may actually alter judgments of its physical properties. To test this prediction, participants were asked to estimate the angle of triangles with a prompt that asked to either estimate the angle of "this triangle" or of "this three-sided figure" (with the instruction

varying between-subjects). Participants in the *triangle* condition over-estimated the angle more than participants in the *three-sided* condition (**Figure 4**) – possibly caused by a contrast effect between the activated canonical (non-rotated) triangle and the rotated triangle being judged. This difference persisted for the entire length of the experiment (about 150 trials; Lupyan, 2011; Lupyan et al., in preparation).

THE ROLE OF VERBAL LABELS IN THE LEARNING OF NOVEL CATEGORIES

The learning of categories is in principle separable from the learning of their names. A child, for example, can have a conceptual category of "dog" (such that different dogs are reliably classified as being the same kinds of thing) without having a name for the category. In practice, however, the two processes are intimately linked. Not only does conceptual development
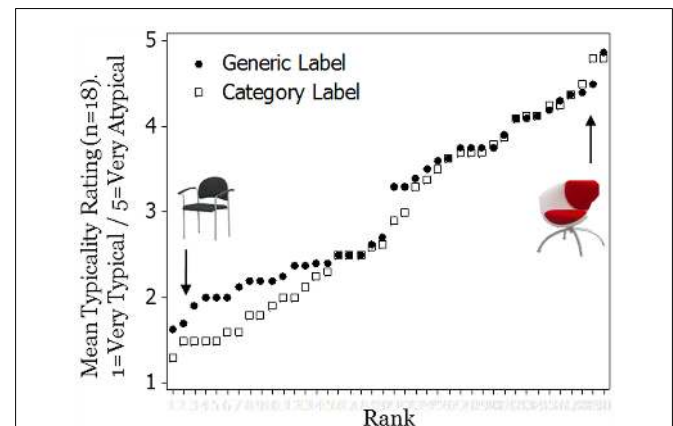
FIGURE 3 | A comparison of typicality ratings of chairs and lamps when the prompt includes the category name ("chair" or "lamp") and when it includes a generic referent ("object"; Lupyan, 2008b, Experiment 6). The x-axis is the average typicality rank of each picture from very typical to very atypical.
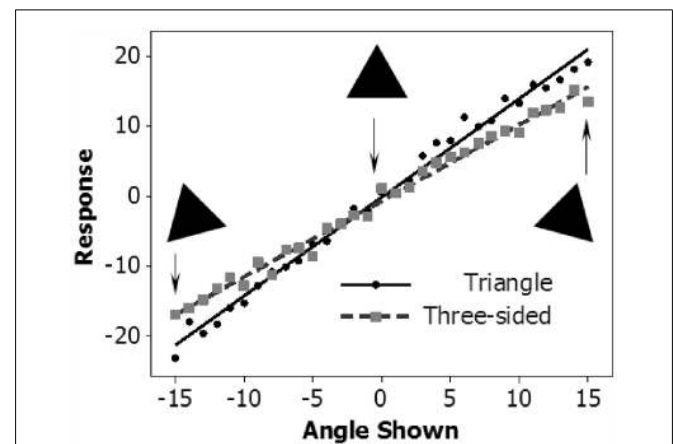
FIGURE 4 | A comparison of judgments of the base angle relative to the horizontal of triangles called "triangles" and the same figures called "three-sided shapes" (Lupyan, 2011).

shape verbal development (e.g., Snedeker and Gleitman, 2004), but verbal learning impacts conceptual development (Waxman and Markow, 1995; Gumperz and Levinson, 1996; Levinson, 1997; Spelke and Tsivkin, 2001; Gentner and Goldin-Meadow, 2003; Yoshida and Smith, 2005; Lupyan et al., 2007). The idea that language shapes concepts has two implications. The first is that it is of course *through* language that we learn much of what we know. This is often seen as trivial, as when Devitt and Sterelny wrote, apparently without irony, that "the only respect in which language clearly and obviously does influence thought turns out to be rather banal: language provides us with most of our concepts" (Devitt and Strelny, 1987, p. 178)[6]. The second implication is that the very use of words may facilitate, or in some cases enable, the ability to impose categories on the external world. Do category names actually facilitate the learning of novel categories?

In a study designed to answer this question, Lupyan et al. (2007) compared the ability of participants to learn categories that were labeled to the learning of the same categories without names. The basic task required participants to learn to classify 16 "aliens" into those that ought to be approached and those to be avoided, responding with the appropriate direction of motion (approach/escape). The category distinction involved subtle differences in the configuration of the "head" and "body" of the creatures. On each training trial, one of the 16 aliens appeared in the center of the screen and had to be categorized by moving a character in a spacesuit (the "explorer") toward or away from the alien, with auditory feedback marking the response as correct or not. In the *label* conditions, a printed or auditory label (the nonsense terms, "leebish" and "grecious") appeared next to the alien; in the *no-label* condition, the alien remained on the screen by itself. All the participants received the same number of categorization trials and saw the aliens for exactly the same duration; the only difference between the groups was the presence of the category labels that followed each response. The labels, being perfectly predictive of the behavioral responses, constituted entirely redundant information.

The results showed that participants in the label conditions learned to classify the aliens much faster than those in the no-label conditions. When the labels were replaced with equally redundant and easily learned non-linguistic and non-referential cues (corresponding to where the alien lived), the cues failed to facilitate categorization. After completing the category-training phase during which participants in both groups eventually reached ceiling performance, their knowledge of the categories was tested in a speeded categorization task using a combination of previously categorized and novel aliens, presented without any feedback or labels. Results showed that those who learned the categories in the presence of labels retained their category knowledge throughout the testing phase. Those who learned the categories without labels showed a decrease in accuracy over time. Thus, learning named categories appears to be easier than learning unnamed categories. More than just learning to map words onto pre-existing concepts

[6]It is probably too obvious to mention, but this function of language is far from banal. Consider that in the absence of language, much of what humans need to learn to survive would have to be learned through slow and dangerous trial and error (Harnad, 2005). It is not an exaggeration to claim that without the ability to learn through language human culture would not exist (Deacon, 1997).

(cf. Li and Gleitman, 2002; Snedeker and Gleitman, 2004), words appear to facilitate the categorization process itself.

IMPLICATIONS OF THE LABEL-FEEDBACK HYPOTHESIS FOR THE "LANGUAGE AND THOUGHT" RESEARCH PROGRAM

Most work investigating the relationship between language, cognition, and perception has assumed that verbal and non-verbal representations are fundamentally distinct and the goal of the "language and thought" research program is to understand whether and how linguistic representations affect non-linguistic representations (Wolff and Holmes, 2011). On such a view, information communicated or encoded via language comprises what is essentially a separate "verbal" modality or channel (Paivio, 1986). Linguistic effects are ascribed either to language influencing "deep" non-verbal processes which ought to not be affected by verbal interference or acquired language deficits, or else hinge on high-level processes that combine verbal and non-verbal input in some way (e.g., Roberson and Davidoff, 2000; Pilling et al., 2003; Dessalegn and Landau, 2008; Mitterer et al., 2009). Neither proposed mechanism can explain how language can have pervasive effects on perceptual processing that are nevertheless permeable to linguistic manipulations such as verbal interference – the paradox outlined above.

The label-feedback hypothesis provides a way of resolving the paradox. Effects of language can indeed run "deep" in the sense of affecting low-level processes (e.g., Thierry et al., 2009) – the very processes claimed by Gleitman (2010) to be impervious to language. Such effects of language on, e.g., color perception need not arise from language somehow permanently warping perceptual space. Thinking of these effects as occurring on-line explains why they can be modulated by verbal factors such as overt language use and verbal interference. Framing effects of language as occurring on-line does not render them superficial, strategic, or necessarily under voluntary control (Lupyan and Spivey, 2010a,b; Lupyan et al., 2010b). On this formulation, the distinction between verbal and non-verbal representations becomes moot, just as taking seriously the pervasiveness of top-down effects in perception renders moot the distinction between "earlier" and "later" cortical areas (Gilbert and Sigman, 2007).

To return to the case of linguistic effects on color perception: On the present view, a visual representation of a color, e.g., blue, becomes rapidly modulated by the activation of the word "blue," a process that can be exaggerated by exogenous presentation of the label and attenuated by manipulations such as verbal interference. Thus, although the *bottom-up* processing of color is likely to be independent of language and identical in speakers of different languages, the *top-down* effects in which language takes part are dependent on the word-color associations to which the speakers have been exposed, and will thus be correspondingly different between speakers who possess a generic term "blue" and those who do not. Such modulations occur as the label becomes active (over the course of a few 100 ms). There is nothing mysterious about this process: it is simply the consequence of the idea that visual representations involved in making even the simplest visual decisions are augmented by feedback higher-level, and typically more anterior brain regions. Feedback from language-based activations such as the activation of the word "green" on seeing green color patches can be seen as one form of such top-down influence.

Although color processing has been a popular testing ground for exploring effects of language[7], the label-feedback hypothesis has a broader relevance. At stake is the question of whether and to what degree perception of familiar objects is continuously augmented by the labels that become co-active with perceptual representations of these objects. This means that once a label is learned, it can potentially modulate subsequent processing (visual and otherwise) of objects to which the label refers. Indeed, the benefits of names in learning novel categories (Lupyan et al., 2007), may derive, at least in part, from the labels' effect on perceptual processing of the exemplars (see also Lupyan and Thompson-Schill, 2012). Lexicalization patterns differ substantially between languages (e.g., Bowerman and Choi, 2001; Lucy and Gaskins, 2001; Majid et al., 2007; Evans and Levinson, 2009). Accordingly, speakers of different languages end up with different patterns of associations between labels and external objects, resulting in different top-down effects of language on ongoing "non-verbal" processing in speakers of different languages.

The label-feedback hypothesis as presented here does not claim to be relevant to all effects labeled as "Whorfian" in the literature. The most direct application is to the processes of categorization and object perception. The hypothesis does not predict that any differences in the grammar of language translate to meaningful differences in "thought." A pervasive additional source of confusion in the language and thought literature that I have not discussed here relates to predicting the consequences that a particular linguistic difference should have on a particular putatively non-linguistic task. Consider, for example, the observation that English verbs highlight the manner of motion (e.g., walk, run, hop) leaving the path as an option, while Spanish verbs highlight the path of motion (e.g., entrar, pasar) leaving the manner as an option (Talmy, 1988). Does the priority of manner information in English mean that English speakers should have better memory for manner than Spanish speakers? Perhaps, but one might just as easily predict the opposite pattern: Spanish speakers ought to have better memory for manner information because, when it is mentioned, it is more unexpected and thus more salient (cf. Gennari et al., 2002; Papafragou et al., 2008). Progress in this area appears to require a firmer marriage between memory researchers and psycholinguists.

More generally, rather than attempting to decide whether a given representation comprises a verbal or visual "code" (e.g., Dessalegn and Landau, 2008), on the current proposal, it may be more productive to measure the degree to which performance on *specific tasks* is being modulated by language, modulated differently by different languages, or is truly independent of any experimental manipulations that can be termed linguistic. On this account, the central question is not "do speakers of different languages have different color concepts" but rather "how does language affect the perceptual representations of color brought to bear on a given task." Much of the literature in the language and thought arena holds an implicit (and sometimes explicit) assumption that there exists such things as *the* concept of a dog, or *the* concept of green-ness. On this assumption, accepting that the concept of green-ness is influenced by language creates the expectation that one should observe those linguistic effects on any task that taps into that singular color concept. Failure to observe these effects is then used by as an argument against linguistic relativity or language-mediated vision. On an alternative view, however, conceptual representations are dynamic assemblies that are a function of prior knowledge as well as current task demands (Casasanto and Lupyan, 2011; Lupyan et al., under review; see also Prinz, 2004). There is therefore no single concept of green-ness. Rather, the influence of language on ongoing cognitive and perceptual processing may be present in some tasks and non-existent in others. For example, given the categorical nature of linguistic reference, one prediction is that effects of language ought to become stronger in tasks that require or promote categorization and weaker in tasks that discourage it (e.g., realistic drawing, remembering exact spatial locations, judging a continuously varying motion trajectory). By understanding how language may augment specific cognitive and perceptual processes, we can make predictions about the kinds of tasks should or should not be influenced by language broadly construed and by differences between languages.

CONCLUSION

I have argued that a pervasive source of theoretical confusion regarding effects of language on cognition and perception stems from a failure to appreciate the degree to which virtually all cognitive and perceptual acts reflect interactive-processing, combining bottom-up and top-down sources of information. An effect of language on how we perceive the rainbow does not require it to alter the responses of photoreceptors. A deep and persistent effect of language on object concepts does not require it to alter conceptual "cores" (indeed, the very existence of such conceptual cores is debatable, Barsalou, 1987; Prinz, 2004; Casasanto and Lupyan, 2011).

Our perception of rainbows, dogs, and everything in between is a product of both their physical properties and top-down processes. The idea that words affect ongoing cognitive and perceptual processes via top-down feedback provides a useful way for thinking about the interaction of language with other processes. In its present form, the label-feedback hypothesis is merely a sketch, but as evidenced by some of the studies reviewed in this paper, this framework provides a powerful intuition pump for generating testable predictions. The label-feedback hypothesis is broadly consistent with what we know about neural mechanisms of perception and categorization, although its neural underpinnings remain almost completely unexplored. The next step is to understand these mechanisms.

---

[7]Witzel and Gegenfurtner (2011) present a cogent argument that most recent investigations of categorical color perception have made incorrect assumptions regarding psychophysical distances in the CIE color space, such that color pairs claimed to be equally spaced in psychophysical space may not be, rendering many of the claims made by these studies difficult to interpret.

REFERENCES

Armstrong, S. L., Gleitman, L. R., and Gleitman, H. (1983). What some concepts might not be. *Cognition* 13, 263–308.

Athanasopoulos, P., Wiggett, A., Dering, B., Kuipers, J.-R., and Thierry, G. (2009). The Whorfian mind: electrophysiological evidence that language shapes perception. *Commun. Integr. Biol.* 2, 332–334.

Barsalou, L. W. (1987). "The instability of graded structure: implications for the nature of concepts," in *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, ed. U. Neisser (Cambridge: Cambridge University Press), 101–140.

Bentin, S., and Golland, Y. (2002). Meaningful processing of meaningless stimuli: the influence of perceptual experience on early visual processing of faces. *Cognition* 86, B1–B14.

Black, M. (1959). Linguistic relativity: the views of Benjamin Lee Whorf. *Philos. Rev.* 68, 228–238.

Borges, J. L. (1942/1999). *Collected Fictions*. Penguin (New York, NY: Non-Classics).

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

Boroditsky, L. (2010). "How the languages we speak shape the ways we think: the FAQs," in *The Cambridge Handbook of Psycholinguistics*, eds M. J. Spivey, M. Joanisse, and K. McRae (Cambridge: Cambridge University Press).

Boroditsky, L. (2003). "Linguistic relativity," in *Encyclopedia of Cognitive Science*, ed. L. Nadel (London: Macmillan), 917–922.

Bowerman, M., and Choi, S. (2001). "Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories," in *Language Acquisition and Conceptual Development*, eds M. Bowerman and S. C. Levinson (Cambridge: Cambridge University Press), 475–511.

Carruthers, P. (2002). The cognitive functions of language. *Behav. Brain Sci.* 25, 657–674.

Casasanto, D. (2008). Who's afraid of the big bad whorf? Crosslinguistic differences in temporal language and thought. *Lang. Learn.* 58, 63–79.

Casasanto, D., and Lupyan, G. (2011). Ad hoc cognition. *Presented at the Annual Conference of the Cognitive Science Society*, Boston, MA.

Cassirer, E. (1962). *An Essay on Man: An Introduction to a Philosophy of Human Culture*. New Haven, CT: Yale University Press.

Churchland, P. S., Ramachandran, V., and Sejnowski, T. J. (1994). "A critique of pure vision," in *Large-Scale Neuronal Theories of the Brain*, eds C. Koch and J. L. Davis (Cambridge, MA: The MIT Press), 23–60.

Clark, A. (1998). "Magic words: how language augments human computation," in *Language and Thought: Interdisciplinary Themes*, eds P. Carruthers and J. Boucher (New York, NY: Cambridge University Press), 162–183.

Cohen, R., Kelter, S., and Woll, G. (1980). Analytical competence and language impairment in aphasia. *Brain Lang.* 10, 331–347.

Cohen, R., Woll, G., Walter, W., and Ehrenstein, H. (1981). Recognition deficits resulting from focussed attention in aphasia. *Psychol. Res.* 43, 391–405.

Daoutis, C. A., Franklin, A., Riddett, A., Clifford, A., and Davies, I. R. L. (2006). Categorical effects in children's colour search: a cross-linguistic comparison. *Br. J. Dev. Psychol.* 24, 373–400.

Davidoff, J., Davies, I. R. L., and Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature* 398, 203–204.

Davidoff, J., and Roberson, D. (2004). Preserved thematic and impaired taxonomic categorisation: a case study. *Lang. Cogn. Process.* 19, 137–174.

Davies, I. R. L., and Corbett, G. (1998). A cross-cultural study of color-grouping: tests of the perceptual-physiology account of color universals. *Ethos* 26, 338–360.

Deacon, T. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain*. London: The Penguin Press.

Deco, G., and Lee, T. S. (2002). A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing* 44, 775–781.

Dennett, D. C. (1994). "The role of language in intelligence," in *What is Intelligence? The Darwin College Lectures*, ed. J. Khalfa (Cambridge: Cambridge University Press).

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual-attention. *Annu. Rev. Neurosci.* 18, 193–222.

Dessalegn, B., and Landau, B. (2008). More than meets the eye: the role of language in binding and maintaining feature conjunctions. *Psychol. Sci.* 19, 189–195.

Devitt, M., and Strelny, K. (1987). *Language and Reality: An Introduction to the Philosophy of Language*. Cambridge, MA: MIT Press.

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., and Davies, I. R. L. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1097–1102.

Esterman, M., and Yantis, S. (2008). Category expectation modulates object-selective cortical activity. *J. Vis.* 8, 555a.

Evans, N., and Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429.

Fausey, C. M., and Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychon. Bull. Rev.* 18, 150–157.

Folstein, J. R., Gauthier, I., and Palmeri, T. J. (2010). Mere exposure alters category learning of novel objects. *Front. Psychol.* 1:40. doi:10.3389/fpsyg.2010.00040

Foxe, J. J., and Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans – a framework for defining "early" visual processing. *Exp. Brain Res.* 142, 139–150.

Freeman, W. J. (2001). *How Brains Make Up Their Minds*, 1st Edn. New York, NY: Columbia University Press.

Gennari, S. P., Sloman, S. A., Malt, B. C., and Fitch, W. T. (2002). Motion events in language and cognition. *Cognition* 83, 49–79.

Gentner, D., and Goldin-Meadow, S. (2003). *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, MA: MIT Press.

Gentner, D., and Loewenstein, J. (2002). "Relational language and relational thought," in *Language, Literacy, and Cognitive Development*, eds J. Byrnes and E. Amsel (Mahwah, NJ: LEA), 87–120.

Gilbert, A. L., Regier, T., Kay, P., and Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proc. Natl. Acad. Sci. U.S.A.* 103, 489–494.

Gilbert, C. D., and Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron* 54, 677–696.

Gleitman, H., Fridlund, A. J., and Reisberg, D. (2004). *Psychology*, 6th Edn. New York: Norton and Company.

Gleitman, L. (2010). *Economist Debates: Language: This House Believes that the Language We Speak Shapes How We Think*. Available at: http://www.economist.com/debate/days/view/632 [retrieved September 1, 2011].

Gleitman, L., and Papafragou, A. (2005). "Language and thought," in *Cambridge Handbook of Thinking and Reasoning*, eds K. Holyoak and B. Morrison (Cambridge: Cambridge University Press), 633–661.

Goldstein, K. (1924/1948). *Language and Language Disturbances*. New York: Grune and Stratton.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *J. Exp. Psychol. Gen.* 123, 178–200.

Goldstone, R. L. (1998). Perceptual learning. *Ann. Rev. Psychol.* 49, 585–612.

Goldstone, R. L., and Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition* 65, 231–262.

Goldstone, R. L., and Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisc. Rev. Cogn. Sci.* 1, 69–78.

Goldstone, R. L., Lippa, Y., and Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition* 78, 27–43.

Goldstone, R. L., Steyvers, M., and Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Mem. Cognit.* 31, 169–180.

Gopnik, A. (2001). "Theories, language, and culture: whorf without wincing," in *Language Acquisition and Conceptual Development*, eds M. Bowerman and S. C. Levinson (Cambridge: Cambridge University Press), 45–69.

Gregory, R. L. (1970). *The Intelligent Eye*. New York: McGraw-Hill.

Gumperz, J. J., and Levinson, S. C. (1996). *Rethinking Linguistic Relativity*. Cambridge: Cambridge University Press.

Harnad, S. (2005). "Cognition is categorization," in *Handbook of Categorization in Cognitive Science*, eds H. Cohen and C. Lefebvre (San Diego, CA: Elsevier Science), 20–45.

James, W. (1890). *Principles of Psychology*, Vol. 1. New York: Holt.

January, D., and Kako, E. (2007). Re-evaluating evidence for linguistic relativity: reply to Boroditsky (2001). *Cognition* 104, 417–426.

Kemmerer, D. (2010). "How words capture visual experience: the perspective from cognitive neuroscience," in *Words and the Mind: How Words Capture Human Experience*, 1st Edn, eds B. Malt and P. Wolff (New York, NY: Oxford University Press), 289–329.

Kemmerer, D., Rudrauf, D., Manzel, K., and Tranel, D. (2010). Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex*. doi:10.1016/j.cortex.2010.11.001. [Epub ahead of print].

Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., and Salminen-Vaparanta, N. (2011). Recurrent processing in V1/V2 contributes to categorization of natural scenes. *J. Neurosci.* 31, 2488–2492.

Kotovsky, L., and Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Dev.* 67, 2797–2822.

Kramer, A. F., Weber, T. A., and Watson, S. E. (1997). Object-based attentional selection – grouped arrays or spatially invariant representations?: comment on Vecera and Farah (1994). *J. Exp. Psychol. Gen.* 126, 3–13.

Kravitz, D. J., and Behrmann, M. (2008). The space of an object: object attention alters the spatial gradient in the surround. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 298–309.

Kuhl, P. (1994). Learning and representation in speech and language. *Curr. Opin. Neurobiol.* 4, 812–822.

Kveraga, K., Ghuman, A. S., and Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain Cogn.* 65, 145–168.

Lamme, V. A. F., Rodriguez-Rodriguez, V., and Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cereb. Cortex* 9, 406–413.

Lamme, V. A. F., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579.

Lee, P. (1996). *The Whorf Theory Complex: A Critical Reconstruction.* Philadelphia, PA: John Benjamins Pub Co.

Levinson, S. C. (1997). "From outer to inner space: linguistic categories and non-linguistic thinking," in *Language and Conceptualization*, eds J. Nuyts and E. Pederson (Cambridge: Cambridge University Press), 13–45.

Li, P., Dunham, Y., and Carey, S. (2009). Of substance: the nature of language effects on entity construal. *Cogn. Psychol.* 58, 487–524.

Li, P., and Gleitman, L. (2002). Turning the tables: language and spatial reasoning. *Cognition* 83, 265–294.

Liu, Q., Li, H., Campos, J. L., Wang, Q., Zhang, Y., Qiu, J., Zhang, Q., and Sun, H. J. (2009). The N2pc component in ERP and the lateralization effect of language on color perception. *Neurosci. Lett.* 454, 58–61.

Loewenstein, J., and Gentner, D. (2005). Relational language and the development of relational mapping. *Cogn. Psychol.* 50, 315–353.

Logothetis, N. K., and Sheinberg, D. L. (1996). Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621.

Lucy, J. A., and Gaskins, S. (2001). "Grammatical categories and the development of classification preferences: a comparative approach," in *Language Acquisition and Conceptual Development*, eds M. Bowerman and S. C. Levinson (Cambridge: Cambridge University Press), 257–283.

Lupyan, G. (2007). "Reuniting categories, language, and perception," in *Twenty-Nineth Annual Meeting of the Cognitive Science Society*, eds D. S. McNamara and J. G. Trafton (Austin, TX: Cognitive Science Society), 1247–1252.

Lupyan, G. (2008a). The conceptual grouping effect: categories matter (and named categories matter more). *Cognition* 108, 566–577.

Lupyan, G. (2008b). From chair to "chair:" a representational shift account of object labeling effects on memory. *J. Exp. Psychol. Gen.* 137, 348–369.

Lupyan, G. (2009). Extracommunicative functions of language: verbal interference causes selective categorization impairments. *Psychon. Bull. Rev.* 16, 711–718.

Lupyan, G. (2011). Representations of basic geometric shapes are created ad-hoc. concepts, actions, and objects workshop. *Presented at the Concepts, Actions, and Objects Workshop*, Rovereto.

Lupyan, G. (in press). "What do words do? Towards a theory of language-augmented thought," in *The Psychology of Learning and Motivation*, Vol. 57, ed. B. H. Ross.

Lupyan, G., Mirman, D., Hamilton, R. H., and Thompson-Schill, S. L. (2010a). Linking language, cognitive control, and categorization: evidence from aphasia and transcranial direct current stimulation. *Presented at the Second Conference on the Neurobiology of Language*, San Diego, CA.

Lupyan, G., Thompson-Schill, S. L., and Swingley, D. (2010b). Conceptual penetration of visual processing. *Psychol. Sci.* 21, 682–691.

Lupyan, G., Rakison, D. H., and McClelland, J. L. (2007). Language is not just for talking: labels facilitate learning of novel categories. *Psychol. Sci.* 18, 1077–1082.

Lupyan, G., and Spivey, M. J. (2008). Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Curr. Biol.* 18, R410–R412.

Lupyan, G., and Spivey, M. J. (2010a). Making the invisible visible: auditory cues facilitate visual object detection. *PLoS ONE* 5, e11452. doi:10.1371/journal.pone.0011452

Lupyan, G., and Spivey, M. J. (2010b). Redundant spoken labels facilitate perception of multiple items. *Atten. Percept. Psychophys.* 72, 2236–2253. doi:10.3758/APP.72.8.2236

Lupyan, G., and Swingley, D. (in press). Self-directed speech affects visual processing. *Q. J. Exp. Psychol.*

Lupyan, G., and Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *J. Exp. Psychol. Gen.* 141, 170–186.

Macpherson, F. (2012). Cognitive penetration of colour experience: rethinking the issue in light of an indirect mechanism. *Philos. Phenomenol. Res.* 84, 24–62.

Majid, A., Gullberg, M., van Staden, M., and Bowerman, M. (2007). How similar are semantic categories in closely related languages? A comparison of cutting and breaking in four Germanic languages. *Cogn. Linguist.* 18, 179–194.

McClelland, J. L., and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception 1. An account of basic findings. *Psychol. Rev.* 88, 375–407.

Mesulam, M. M. (1998). From sensation to cognition. *Brain* 121, 1013–1052.

Mesulam, M. M. (2008). Representation, inference, and transcendent encoding in neurocognitive networks of the human brain. *Ann. Neurol.* 64, 367–378.

Meteyard, L., Bahrami, B., and Vigliocco, G. (2007). Motion detection and motion verbs – language affects low-level visual perception. *Psychol. Sci.* 18, 1007–1013.

Mitterer, H., Horschig, J. M., Musseler, J., and Majid, A. (2009). The influence of memory on perception: it's not what things look like, it's what you call them. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1557–1562.

Mo, L., Xu, G., Kay, P., and Tan, L.-H. (2011). Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color.

Proc. Natl. Acad. Sci. U.S.A. 108, 14026–14030.

Mumford, D. (1992). On the computational architecture of the neocortex II. The role of cortico-cortical loops. *Biol. Cybern.* 66, 251.

Noppeney, U., and Wallesch, C. W. (2000). Language and cognition – Kurt Goldstein's theory of semantics. *Brain Cogn.* 44, 367–386.

Olshausen, B. A., Anderson, C. H., and Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* 13, 4700–4719.

Ozgen, E., and Davies, I. R. L. (2002). Acquisition of categorical color perception: a perceptual learning approach to the linguistic relativity hypothesis. *J. Exp. Psychol. Gen.* 131, 477–493.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach.* New York: Oxford University Press.

Papafragou, A., Hulbert, J., and Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition* 108, 155–184.

Pilling, M., Wiggett, A., Ozgen, E., and Davies, I. R. L. (2003). Is color "categorical perception" really perceptual? *Mem. Cognit.* 31, 538–551.

Porter, P. B. (1954). Another picture puzzle. *Am. J. Psychol.* 67, 550–551.

Prinz, J. J. (2004). *Furnishing the Mind: Concepts and their Perceptual Basis.* The MIT Press.

Puri, A. M., and Wojciulik, E. (2008). Expectation both helps and hinders object perception. *Vision Res.* 48, 589–597.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* 22, 341–365.

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. *Nat. Neurosci.* 2, 79–87.

Ratterman, M. J., and Gentner, D. (1998). "The effect of language on similarity: the use of relational symbols improves young children's performance on a mapping task," in *Advances in Analogy Research: Integration of Theory and Data From the Cognitive, Computational and Neural Sciences*, eds K. Holyoak, D. Gentner and B. Kokinov (Sophia: New Bulgarian University).

Reynolds, J. H., and Chelazzi, L. (2004). Attentional modulation of visual processing. *Annu. Rev. Neurosci.* 27, 611–47.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., and Brecher, A. (1996). The Philadelphia naming test: scoring and rationale. *Am. J. Speech Lang. Pathol.* 24, 121–133.

Roberson, D., and Davidoff, J. (2000). The categorical perception of colors and facial expressions: the effect of verbal interference. *Mem. Cognit.* 28, 977–986.

Roberson, D., Davidoff, J., Davies, I. R. L., and Shapiro, L. R. (2005). Color categories: evidence for the cultural relativity hypothesis. *Cogn. Psychol.* 50, 378–411.

Roberson, D., Pak, H., and Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: evidence from Korean. *Cognition* 107, 752–762.

Rumelhart, D. E., McClelland, J. L., and the PDP Research Group. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vols 1 and 2. Cambridge, MA: MIT Press.

Slobin, D. (1996). "From 'thought and language' to 'thinking for speaking,'" in *Rethinking Linguistic Relativity*, eds J. J. Gumperz and S. C. Levinson (Cambridge: Cambridge University Press), 70–96.

Sloutsky, V. M. (2010). From perceptual categories to concepts: what develops? *Cogn. Sci.* 34, 1244–1286.

Snedeker, J., and Gleitman, L. (2004). "Why is it hard to label our concepts?" in *Weaving a Lexicon*, eds D. G. Hall and S. R. Waxman (Cambridge, MA: The MIT Press), 257–294.

Spelke, E. S. (2003). "What makes us smart? Core knowledge and natural language," in *Language in Mind: Advances in the Study of Language and Thought*, eds D. Gentner and S. Goldin-Meadow (Cambridge, MA: MIT Press), 277–311.

Spelke, E. S., and Tsivkin, S. (2001). "Initial knowledge and conceptual change: space and number," in *Language Acquisition and Conceptual Development* (Cambridge Cambridge University Press), 475–511.

Talmy, L. (1988). Force dynamics in language and cognition. *Cogn. Sci.* 12, 49–100.

Tan, L. H., Chan, A. H. D., Kay, P., Khong, P.-L., Yip, L. K. C., and Luke, K.-K. (2008). Language affects patterns of brain activation associated with perceptual decision. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4004–4009.

Tanaka, J. W., and Corneille, O. (2007). Typicality effects in face and object perception: further evidence for the attractor field model. *Percept. Psychophys.* 69, 619–627.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., and Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4567–4570.

Vecera, S. P., and Farah, M. J. (1994). Does visual-attention select objects or locations. *J. Exp. Psychol. Gen.* 123, 146–160.

Vygotsky, L. (1962). *Thought and Language.* Cambridge, MA: MIT Press.

Ward, E. J., and Lupyan, G. (2011). Linguistic penetration of suppressed visual representations. *Presented at the Vision Sciences Society*, Naples, FL.

Waxman, S. R., and Markow, D. B. (1995). Words as invitations to form categories: evidence from 12- to 13-month-old infants. *Cogn. Psychol.* 29, 257–302.

Whorf, B. L. (1956). *Language, Thought, and Reality.* Cambridge, MA: MIT Press.

Wiggett, A., and Davies, I. R. L. (2008). The effect of Stroop interference on the categorical perception of color. *Mem. Cognit.* 36, 231.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7780–7785.

Witzel, C., and Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *J. Vis.* 11, 16.

Wolff, P., and Holmes, K. (2011). Linguistic relativity. *Wiley Interdiscip. Rev. Cogn. Sci.* 2, 253–265.

Yoshida, H., and Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychol. Sci.* 16, 90–95.