Spontaneous Verbal Labeling in 24-month-old Infants

Manizeh Khan, Amy Geojo, Shanshan Wang, & Jesse Snedeker

Harvard University

Abstract word count: 147

Word count: 4556

Keywords: Phono-semantic priming; Language and thought; Implicit naming; Infants

Contact author:
Manizeh Khan
33 Kirkland Street
c/o Department of Psychology
Harvard University
Cambridge, MA 02138
USA

E-mail: khan@wjh.harvard.edu
Telephone: 857-919-7370

Abstract

Are linguistic representations spontaneously recruited for internal thought? The current study investigates this question in infants, to determine whether this propensity emerges early or is a late-acquired habit. Prior preferential-looking studies have found that infants show phonological and phono-semantic priming during word recognition. We employed a parallel *nonverbal* priming paradigm to explore whether linguistic representations are spontaneously activated in tasks that do not involve explicit labeling. In this task, 24-month-olds saw a series of unlabeled pictures. On critical trials, the prime (*cup*) was phonologically related to an unseen and unmentioned intermediary (*cat*), which was semantically related to the target (*dog*). Infants looked less at target pictures following phono-semantically related prime images compared to unrelated control images, indicating activation of the unheard phonological form of the prime label. This suggests that spontaneous verbal encoding emerges early in development, even in contexts where it serves no obvious communicative function.

1. Introduction

A central and controversial issue in cognitive science is the role of language in internal thought. Theorists as diverse as Mead (1934), Vygotsky (1934/1987), Whorf (1956) and Chomsky (2007) have proposed that language is the primary vehicle for internal thought. In contrast, others have argued that thought must occur over representations that are distinct from natural languages (Fodor, 1975; Jackendoff, 2002; Pinker, 1994). In some cases, the diverging opinions reflect differences in how the word "language" is defined—whether it picks out conceptual and semantic representations, syntax, phonology, or all of the above (see Jackendoff, 2002). Nevertheless, after definitional differences are sorted out and evolutionary claims are put aside, a clear empirical question remains: Does the specific form of the external language that we use to communicate play a role in our internal cognitive life? This question is often approached by finding languages with diverging syntactic and semantic categories and looking for differences in the performance of non-communicative tasks (see Gentner & Goldin-Meadow, 2003 for examples), but it can also be addressed by testing whether the phonological forms of linguistic expressions are active during nonlinguistic tasks.

There is no doubt that internal verbalization occurs. We have mental conversations with people who are absent, we silently coach ourselves during difficult tasks ("Eyes on the ball!"), and we use subvocal rehearsal as an short-term memory aid (Martínez Manrique & Vicente, 2010; Vicente & Martínez Manrique, 2011; Winsler, 2009). But several questions remain. How ubiquitous is verbal encoding: is it largely limited to strategic, metacognitive contexts like those above, or does it pervade our spontaneous thought? Does verbal encoding typically involve the activation of phonological representations or does internal thought use the semantic representations of language stripped free from their external trappings? And finally, when in development does spontaneous verbal encoding, outside of communicative contexts, first emerge? The present paper begins to explore this last question, by looking at the activation of verbal labels for visually-presented objects in 24-month-old infants.

While there is some evidence that adults spontaneously activate phonological representations in noncommunicative contexts, the data pattern is complex and ambiguous. In recognition memory tasks, the time spent looking at an object during the encoding phase is affected by name length of the object (Zelinsky & Murphy, 2000). Although this task had only a 2.5 second delay and could be solved visually, this pattern indicates that adults encoded the pictures verbally, perhaps so they could verbally rehearse their labels. Critically, the authors found no effect of name length in a parallel visual-search task which had no memory demands. There is, however, some evidence for the activation of linguistic form during visual search. Specifically, the presence of homophonous competitors results in interference (e.g., more looks to a baseball bat when searching for an animal bat) suggesting that the lexical label of the picture has been retrieved (Meyer, Belke, Telling, & Humphreys, 2007). Curiously, visual search tasks are not influenced by phonological overlap between the target and distractors, persons searching for “candy” do not linger on candles (Telling, 2008). This raises the possibility that the homophone effects either reflect activation at the lexical level (rather than at the phonological level) or depend on the participants’ metalinguistic awareness of homophony.

The origin of internal verbal representations has been a central theoretical issue in developmental psychology. Vygotsky (1934/1987) proposed that children initially use language solely as a social tool, with internal speech developing gradually from external dialog. The first step toward internalization occurs at around two or three when children begin talking aloud to themselves. This private speech becomes quieter, less frequent, more elliptical, and more covert during the late preschool and early elementary years (Berk, 1986; Winsler & Naglieri, 2003; see Winsler, 2009 for review). Researchers working within the Vygotskyian tradition interpret these changes in private speech as evidence for the emergence of internalized verbal thought (see e.g., Fernyhough, 2009).

However, there is little direct evidence that internal verbalization is absent before the rise and fall of private speech. Unlike adults, young children do not report thinking in words (Flavell, Green, Flavell, & Grossman, 1997; Manfra & Winsler, 2006), but this could reflect a failure to understand the question or introspect, rather than the absence of internal verbalization (Flavell & Wong, 2009). In memory studies with pictoral stimuli, preschoolers make errors based on visual similarity but do not to show effects of word length, phonological similarity or verbal suppression, suggesting that they do not use internal verbalization even in those tasks where adults rely on it (Conrad, 1971; Ford & Silber, 1994; Hayes & Schulze, 1977; Hitch & Halliday, 1983; Hitch, Halliday, Dodd, & Littler, 1989; but see Henry, Tuner, Smith, & Leather, 2000; Hulme, Silvester, Smith, & Muir, 1986). This pattern, however, could reflect a failure to use verbal rehearsal or retrieval strategies, rather than the absence of verbal encoding (Hitch & Halliday, 1983; Johnston & Conning, 1990: Nairne, 1990). In fact, in picture memory studies 5-year-olds have been observed making silent lip movements as the pictures are presented, suggesting that they are retrieving their verbal labels (Locke & Fehr, 1970; but see Flavell, Beach, & Chinsky, 1966).

In this paper, we explore a fundamentally different hypothesis about the development of internal verbalization, one that is rooted in psycholinguistic research and the information-processing tradition. Psycholinguists construe language as a series of linked representations (phonological, lexical, syntactic and semantic) which are constructed during comprehension and production (see e.g., Alario, Costa, Ferreira, & Pickering, 2006; Altmann, 2001; Snedeker, 2009). Language development involves, among other things, acquiring connections between levels, such as the mappings between phonological word forms and the concept that a word expresses (Jackendoff, 2002). Once a connection has been acquiried, activation at one level of representation can result in activation of the linked representation, as it must for successful word comprehension or production. Covert speech occurs when lexical and phonological representations are constructed without giving rise to articulatory plans (Indefrey & Levelt, 2004). On this construal, the representational basis for internal verbalization is available as soon as language acquisition begins, raising the possibility that older infants might already think in words. To explore this prediction, we looked for evidence of phonological activation in 24-month-olds who were passively viewing pictures.

The present study builds directly on experiments by Mani and Plunkett (2010; 2011; Mani, 2010) which show that infant word recognition is sensitive to phonological and phono-semantic priming. For example, in the phonological priming study, infants were shown a prime image (e.g. a picture of a cup), unaccompanied by any explicit label, followed by split screen with two images (a cat and a house), one of which was explicitly labeled ("cat"). Infants' looking times to the target image (cat) were affected by whether or not the label is phonologically related to the name of the previous silent image: 18-month-olds show phonological facilitation, while 24-month-olds show phonological interference.

For these effects to emerge, infants had to activate the label of the prime image, which was never spoken aloud. Thus some form of internal verbal activity occurred. Mani and Plunkett (2010) interpret this activity as implicit naming, and draw a parallel to the homophone effects observed in adults by Meyer and colleagues (2007). However, this task is different from Meyer's in a fundamental way: the unlabeled prime is embedded in an overt word-recognition task, rather than a non-linguistic search task. Throughout the experiment infants heard an alternating stream of labeled targets and unlabeled primes. Consequently, infants may have generated labels for primes because they expected to hear these words or because they understood this task as a labeling game and were trying to play along. In fact, in adults unlabeled pictures produce phonological interference when they are embedded in a linguistic task, like picture naming (Meyer & Damian, 2007; Morsella & Miozzo, 2002), but not when they are embedded in a nonlinguistic task like visual search (Telling, 2008; see also Zelinsky & Murphy, 2000). Thus the prior preferential-looking studies do not answer the core question that motivates the present experiment: Do infants verbalize their experiences in non-linguistic contexts, or do they do so only in expectation of hearing or producing external labels?

The current study addresses this question by looking for phono-semantic priming in 24-month-olds in a non-linguistic task in which objects are never labeled for the child and the child is never asked to produce a label. In phono-semantic priming, the prime is phonologically related to an intermediary that is semantically related to the target; the intermediary is never explicitly presented during the task (e.g. *cup-cat-dog*). The phenomenon of phono-semantic priming has been observed both in adult and child language processing (Huang & Snedeker, 2011; Marslen-Wilson, 1987; Yee & Sedivy, 2006), and has been observed in 24-month-old infants using the preferential-looking priming paradigm described above (Mani, 2010; Mani, Durrant, & Floccia, 2012). Phono-semantic priming is an attractive tool for examining implicit phonological activation because phono-semantic relations are more difficult for participants to notice than phonological relationships, and therefore less available for strategic processes. Nevertheless, phono-semantic priming still relies on activating the phonological label of the prime.

The present study used a preferential-looking paradigm modeled on Mani (2010); infants saw a prime image followed by a split-screen of two images, one of which was phono-semantically related to the prime on half the trials. There was, however, one critical difference in our paradigm: none of the pictures (targets or primes) were labeled for the child. If infants' free-viewing of the split-screen differs based on whether or not one of the images is phono-semantically related to the previous image, this will provide strong evidence that phonological labels are spontaneously implicitly activated when infants encounter familiar visual objects. Such a finding would be consistent with the view that linguistic representations are recruited for internal thought, even in the early stages of language acquisition.

2. Method

2.1. Participants

Twenty four 24-month-olds from English speaking households participated in this study. An additional four infants were tested but could not be included in the final dataset: one infant did not complete the experiment, and three were excluded due to technical failures or experimenter error.

2.2. Materials

Ten prime-target-distractor triplets of color images were compiled (see Appendix for a complete item list). In each triplet, the prime was phono-semantically related to the target; for example, *cup* is phono-semantically related to *dog* because the label "cup" shares its phonological onset with "cat", and cats are semantically related to dogs. The intermediary or subprime, in this case *cat*, is never mentioned nor visually presented during the task. The distractor item was phonologically, semantically and phono-semantically unrelated to both the prime and the target. An unrelated control version of each item was created by shuffling the prime and target images between items, such that they were no longer phono-semantically related (and were also not phonologically or semantically related). Distractors were yoked to the target image during this shuffling. Since nothing is labeled during this task, an image is deemed the target simply by virtue of having been phono-semantically related to the prime in the original item triplets that were created. Each participant saw each item in either the related-prime or the unrelated-prime condition. Four experimental lists were created such that, for every item, related-prime and unrelated-prime conditions and target position on screen was counterbalanced across participants.

The items used in this study were adapted from Mani (2010), with some substitutions due to differences between British and American English. The words that were relevant to the task (primes, targets, distractors and subprimes) were, on average, produced by 75% of 24-month-olds, in the norming study for the MacArthur-Bates Communicative Development Inventory (Dale & Fenson, 1996). Further, following the preferential-looking task, participants' knowledge of the words relevant to the task was confirmed by asking the caregiver to complete a vocabulary checklist, which indicated that these words were comprehended by an average of 97 percent of the infants tested (range across items = 88 percent – 100 percent). While caregivers completed this checklist, infants were asked to name the prime images that had been presented during the study. Critically, this naming task occurred after the preferential looking task. Images were presented in a booklet and infants were only prompted with questions such as "what is this?" or "can you help me name this?". Five infants did not complete this task because they were too shy or tired. The pictures were correctly named by an average of 92% of the infants who completed this task (range = 84% – 100% by picture).

In order to rule out the possibility that direct semantic or visual similarity between the related-prime and the target could account for differences from the unrelated control condition, we collected adult ratings for these items on Amazon Mechanical Turk. For every item, the semantic relatedness between the related-prime and the target, the unrelated-prime and the target, and the related-subprime and the target were assessed with ten adults rating each pair on a scale from 1-7 (unrelated to highly related). The visual similarity between the related-prime and the target, and the unrelated-prime and the target, were also assessed. These ratings confirmed that neither the related and unrelated primes were semantically related to the targets; mean ratings were 1.54 and 1.33 respectively, and these ratings were not significantly different from each other, $t(9)=0.96$, $p>.3$. Further, in both cases, the semantic relatedness ratings were significantly lower than for the subprime-target pairs (M=5.61, *t*s>23, *p*s<.001). The visual similarity ratings did not significantly differ between the related and unrelated prime conditions, mean ratings were 1.43 and 1.51 respectively, $t(9)=0.34$, $p>.7$.

2.3. Procedure

Participants sat on their caregiver's lap on a chair approximately 6 feet from the projector screen; caregivers were asked to keep their eyes closed for the duration of the task. A video camera immediately below the screen recorded the child's face so that their eye movements could be coded. Audio that was played through speakers behind the screen was also recorded to the videotape. The design of the experiment was modeled on Mani (2010). Each participant viewed 5 trials in the related-prime condition and 5 trials in the unrelated-prime (control) condition with trial order randomized. The sequence of events within a trial is depicted in Figure 1. First, the participant saw the prime image presented in the center of the screen for 1500ms with no sound. Next, a blank screen appeared for 200ms, followed by the target and the distractor image. These images appeared onscreen side by side for 2550ms, with left-right order of the two pictures counterbalanced across items and experimental lists. 50ms after the target and distractor images appeared, a pre-recorded attention-getter in a female voice was played, such as “oh” or “ah”. Finally, the participant saw a blank screen until the experimenter judged s/he was looking at the center of the screen, at which point the prime image for the next trial was presented. The related and unrelated prime conditions only differed with respect to the pairing of prime and target images.

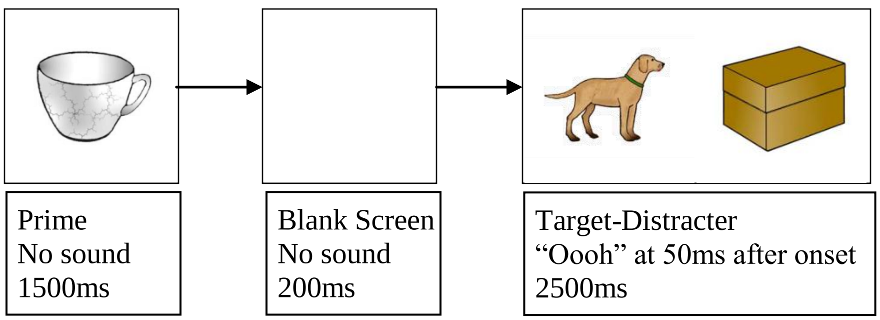

Figure 1. Schematic of a trial in related-prime condition (prime *cup*, subprime *cat*, and target item *dog*). In the unrelated-prime condition, the sequence remains the same except for the prime image, which would be replaced with a picture of a house.

Following the preferential-looking task, caregivers completed the vocabulary checklist. During this time, infants were shown the prime images again and asked to name them.

2.4. Coding

Eye movements were coded from the videotape using frame-by-frame viewing (33ms per frame). Fixations during the target-distractor screen were coded as left, right, center or away. Frames where the participant’s pupils were not clearly visible were coded as track loss. Coding began 50ms after the target and distractor images appeared on screen, at the point when the attention-getter was heard, and continued for 2000ms. Trials were excluded if the participant had not looked at the prime image or if there was more than 50% track loss during the critical time window (3.3% of trials). No participants were excluded for showing a side bias (> 80% looking to either the left or right side of the screen across the ten trials).

3. Results

For each frame between 50ms and 2050ms from target-distractor image onset, we noted whether the participant was looking at the target for each trial (see Figure 2). Frames where the participant was not looking at either the target or the distractor were excluded from the analyses (5% of related-prime condition frames, 4% of unrelated-prime condition frames).

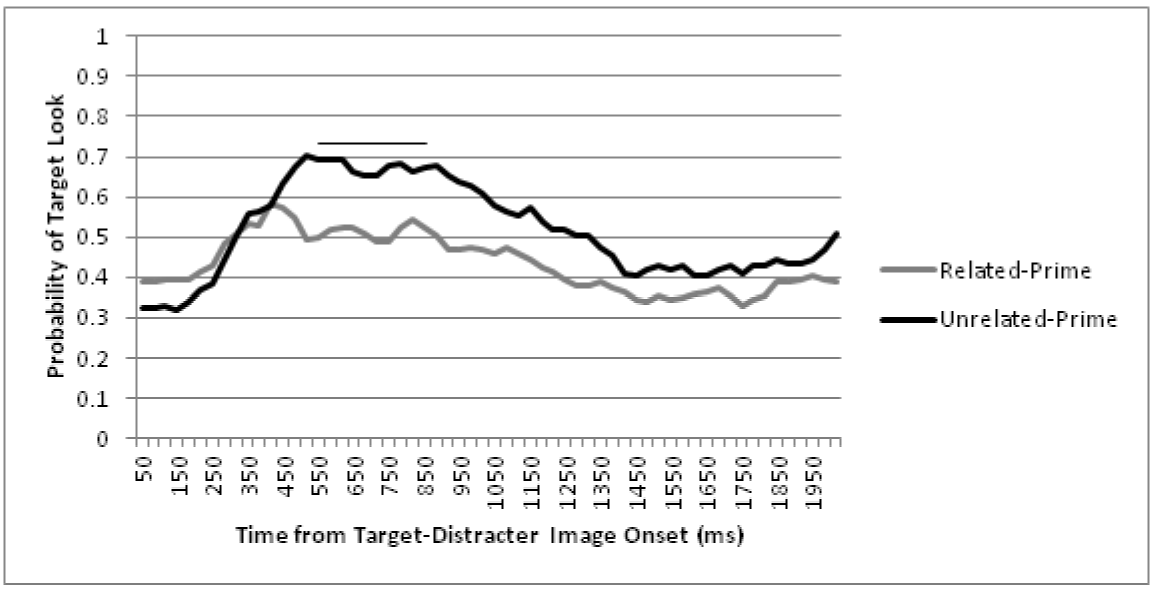

Figure 2. Probability of looking at the target (rather than the distractor) over time. The solid horizontal bar depicts the time region where infants were more likely to look at the target when it was unrelated to the prime ($p < .05$, using parametric cluster test, see text).

Based on the prior findings from the word comprehension task (Mani, 2010), we expected that phono-semantic priming would result in reduced looks to the target in the related-prime condition. However, because of the open-ended nature of our task, we did not have an a priori hypothesis about when this priming effect would emerge. Since our participants were not directed to look at either of the images, their fixation patterns cannot be predicted on the basis of the word comprehension studies. Instead their fixations reflect processes that occur spontaneously during passive viewing (e.g., object recognition or spontaneous verbal labeling). These processes could be synchronized to the onset of the visual stimuli in multiple ways. Thus, we needed a statistical approach which would allow us to compare looking patterns in the related-prime and unrelated-prime conditions across the entire time period when the images were present.

We adopted a strategy for identifying time windows with significant priming effects that allowed for flexibility while maintaining the probability of Type I error at $p<.05$. Specifically, we used a non-parametric cluster-size statistical test following Maris and Oostenveld (2007), which permits testing each time point without inflating the likelihood of generating a false positive. For each time point, we performed a logistic mixed-effects regression on fixations to the target, with prime condition as a fixed effect and subjects and items as random effects, implemented using the lme4 library (Bates, & Maechler, 2010) in R, version 2.11.1 (R Development Core Team, 2010). Time points were grouped into larger windows by identifying clusters of adjacent time points with significant prime condition effects ($p <.05$, two-tailed), and the z statistic for each time point within the cluster was summed to yield a summary statistic for the cluster. To determine the probability of observing a cluster of that size by chance, we conducted 1000 simulations where the condition labels for each trial were randomly assigned. For each of the 1000 sets of re-shuffled data, we implemented the analysis described above and saved the summary statistic for the largest cluster that was identified. A cluster of time points from the original data was considered to show a significant effect of prime condition if its summed z score was greater than the summary statistic of the largest cluster found in 95% of the re-shuffled simulations. Thus the total probability of a false positive is kept below .05. The *p*-values reported here reflect the proportion of re-shuffled simulations that found clusters with summary statistics as large or larger than the reported cluster.

Using this technique, we found fewer looks to the target in the related-prime condition from 550-850ms following target-distractor picture onset ($p<.05$). To test for weaker but long-lasting priming effects, we repeated the analysis described above with a relaxed alpha of .2 for each time point, while maintaining an overall alpha of .05 for each cluster. In this analysis, we observed an effect of prime condition in the 517-1383ms time window ($p<.01$), again reflecting an inhibitory priming effect, consistent with the phono-semantic priming results from Mani (2010).

4. Discussion

This study explored whether infants engage in spontaneously verbal encoding in a non-communicative and non-linguistic task. Infants saw a stream of pictures, none of which were labeled. Nevertheless, their looking patterns on the test trials were influenced by the names of the objects that they had previously seen. Specifically, when the label of the prime object was phonologically related to a close semantic associate of the target, infants were less likely to look at this object. Since phono-semantic priming depends upon the phonological form of the prime, this result indicates that infants internally generated the labels for the visually-presented images. Thus, early in development, language is active, even outside of communicative contexts or linguistic tasks, suggesting it may play a role in internal thought.

These results support the hypothesis we derived from the psycholinguistic model: since internal verbalization involves a subset of the processes involved in language production, it should emerge early in language acquisition. Our findings also provide additional evidence that the 24-month-old infants have a lexicon with the architectural features required to support phono-semantic priming (e.g., phonological neighborhood structure and a close coupling of phonological and semantic representations), confirming the findings from word-recognition tasks (Mani et al., 2012; Mani, 2010).

There are several possible paths from phono-semantic priming to the observed decrease in target looking in this task. It is worth noting that studies using a similar paradigm by Mani and colleagues have also found a decrease in looking to targets following phonologically and phono-semantically related primes (Mani, 2010; Mani & Plunkett, 2011; c.f. Mani et al., 2012). In particular, they find a decrease in target looking in studies with 24-month-old infants where there is only a single phoneme overlap between the prime and subprime or target, which matches the conditions of the current study. In their studies, Mani & Plunkett (2011) and Mani (2010) suggest that the phonological activation of the prime label ultimately inhibits activation of the target label, delaying word recognition. However, it is unclear that inhibition of the target label should necessarily lead to a decrease in looking in the current study, which does not involve word recognition. Another possibility is that accessing the prime label actually facilitated processing of the target object, such that infants were able to retrieve the label more quickly and look away sooner. This explanation is consistent with fixation patterns in Zelinsky & Murphy's (2000) recognition memory task, where adults looked less at objects which required less time to verbalize.

While these findings challenge Vygotsky's specific claims about the age at which internal speech emerges (1934/1987), they lend support to his broad proposal that external symbols play a central role in internal mental life. Researchers in the Vygotskian tradition have argued that internal verbal representations emerge as private speech declines between the ages of 4 and 7

(Winsler, 2009). There are two ways in which this theory could be modified to account for our findings. First, one could retain the premise that private speech is a gateway into internal speech and draw a theoretical distinction between spontaneous verbal activity, which is present in infancy, and true internal speech. This would require further development of the theory and a re-examination of prior research which has used phonological representation as an index of internal speech (Al-Namlah, Fernyhough, & Mein, 2006). Second, one could seek another explanation for the disappearance of private speech such as the acquisition of taboos about talking to oneself (Duncan & Tarulli, 2009).

Our data raise new questions about how verbal encoding changes during development. In adults, the phonological encoding of pictoral stimuli is well-established in short-term memory tasks and language production tasks, but the results from nonlinguistic visual-search tasks are inconsistent. For example, Meyer et al. (2007) found that adults looked longer at distractors when they were homophonous with the target, suggesting that participants were accessing the image labels. On the other hand, Telling (2008) failed to find any effects of partial phonological overlap between target and distractor labels in a range of similar tasks, suggesting that adults may not be activating phonological labels during visual search. In another visual search experiment, Zelinsky and Murphy (2000) did not find effects of verbal activation in adult participants. In contrast, our results demonstrate that infants implicitly name the objects that they see in a nonlingustic task with no memory component. We have considered three explanations for this difference.

First, the free-viewing task could be more sensitive to the effects of implicit verbal labeling than the visual-search task. While more research is needed to evaluate this possibility, our initial studies suggest that it is wrong. We have found that adults show the same pattern of effects in a free-viewing task as they do in the visual-search task: greater looking to homophonous targets but not to phonologically-related words (Khan, Fitts, & Snedeker, 2012).

Second, over development we may gain the ability to limit the activation of linguistic representations to contexts where they are useful (short-term memory tasks) or necessary (word production tasks). This shift could be part of a broader developmental change in the ability to use goals to control cognitive processes, linked to the maturation of prefrontal cortex and the development of executive functions (see e.g., Zelazo, Carlson, & Kesek, 2008).

Finally, perhaps adults, like infants, spontaneously activate linguistic labels of objects they see, but this activation is undetectable with current paradigms because the adult lexicon is so much larger than the child's. A typical 24-month-old produces about 300 words (Dale & Fenson, 1996), a typical adult knows about 60,000 (e.g., Miller, 1996; Nagy & Herman, 1987; Pinker, 1994). As a result, a given word will have many more phonological neighbors for an adult than for a 2-year-old, thus the priming effect for any given phonological neighbor may be diminished. This hypothesis could explain why adults in visual-search tasks show effects of homophony (Meyer et al., 2007) but not phonological overlap (Telling, 2008): homophones necessarily have a greater degree of phonological overlap than non-homophonous pairs.

The use of spontaneous verbal encoding in infancy is relevant to our broader understanding of the relationship between language and thought. The influence of language on thought is often studied by looking at how speakers of different languages perform on what appear to be non-linguistic cognitive tasks. When differences between populations are found they are often interpreted as evidence that language affects the concepts that are available or salient to the speaker (see Gentner & Goldin-Meadow, 2003 for examples). Our findings suggest another possible explanation: linguistic representations may become active in many tasks which do not explicitly require language comprehension or production. When this happens, differences in the language itself may influence performance, even if speakers of both languages have access to the same conceptual representations. This account correctly predicts that verbal interference can disrupt cross-linguistic cognitive differences (see e.g., Winawer et al., 2007) and that a bilingual's pattern of performance will often depend on the language in which they are tested (Boroditsky, Ham, & Ramscar, 2002; Barner, Iganaki, & Li, 2009). Such verbal interference effects have not been explored by developmental psychologists, but the present results suggest that they may begin very early in life.

Acknowledgements

We thank Carlyn Friedberg for her assistance with data collection and Hugh Rabagliati for comments and discussion, in particular with respect to data analysis. This work was supported by funding from the National Science Foundation (BCS-0921012 to JS), the Harvard College Research Program (to SW), and the Natural Sciences and Engineering Research Council of Canada (to MK).

References

Al-Namlah, A. S., Fernyhough, C., & Meins, E. (2006). Sociocultural influences on the development of verbal mediation: Private speech and phonological recoding in Saudi Arabian and British samples. *Developmental Psychology*, *42*, 117–131.

Alario, F., Costa, A., Ferreira, V. S., & Pickering, M. J. (2006). Architectures, representations and processes of language production. *Language And Cognitive Processes, 21*(7-8), 777-789.

Altmann, G.T.M. (2001). The language machine: Psycholinguistics in review. *British Journal of Psychology*, *92*, 129–170.

Barner, D., Inagaki, S., & Li, P. (2009). Language, thought, and real nouns. *Cognition*, *111*, 329-344.

Berk, L. E. (1986). Relationship of elementary school children's private speech to behavioral accompaniment to task, attention, and task performance. *Developmental Psychology*, *22*, 671–680.

Boroditsky, L., Ham, W. & Ramscar, M. (2002). What is universal about event perception? Comparing English and Indonesian speakers. *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*. Fairfax, VA.

Chomsky, N. (2007). Of Minds and Language. *Biolinguistics*, 1, 9-27.

Conrad, R. (1971). The chronology of the development of covert speech in children. *Developmental Psychology*, *5*, 398–405.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28*, 125–127.

Duncan, R., & Tarulli, D. (2009). On the persistence of private speech: Empirical and theoretical considerations. In A. Winsler, C. Fernyhough, I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 176-187). New York, NY US: Cambridge University Press.

Gentner, D., & Goldin-Meadow, S. (2003). *Language in mind*. Cambridge, MA: MIT Press.

Huang, Y. T., & Snedeker, J. (2011). Cascading activation across levels of representation in children's lexical processing. *Journal of Child Language*, *38*, 644–661.

Fernyhough, C. (2009). Dialogic thinking. In A. Winsler, C. Fernyhough, I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 42-52). New York, NY US: Cambridge University Press.

Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development, 37,* 283–299.

Flavell, J. H., Green, F. L., Flavell, E. R., & Grossman, J. B. (1997). The development of children's knowledge about inner speech. *Child Development*, *68,* 39.

Flavell, J. H., & Wong, A. A. (2009). Young children's knowledge about overt and covert private speech. In A. Winsler, C. Fernyhough, I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 143-149). New York, NY US: Cambridge University Press.

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

Ford, S., & Silber, K. P. (1994). Working memory in children: A developmental approach to the phonological coding of pictorial material. *British Journal of Developmental Psychology*, *12,* 165–175.

Hayes, D. S., & Schulze, S. A. (1977). Visual encoding in preschoolers' serial retention. *Child Development*, *48*, 1066-1070.

Henry, L. A., Turner, J. E., Smith, P. T., & Leather, C. (2000). Modality effects and the development of the word length effect in children. *Memory*, *8*, 1–17.

Hitch, G. J., & Halliday, M. S. (1983). Working memory in children. *Philosophical Transactions of the Royal Society of London B, 302*, 325–340.

Hitch, G. J., Halliday, M. S., Dodd, A., & Littler, J. E. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology*, *7,* 347–362.

Hulme, C., Silvester, J., Smith, S., & Muir, C. (1986). The effects of word length on memory for pictures: Evidence for speech coding in young children. *Journal of Experimental Child Psychology*, *41*, 61–75.

Indefrey, P. P., & Levelt, W. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92,* 101-144.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Johnston, R. S., & Conning, A. (1990). The effects of overt and covert rehearsal on the emergence of the phonological similarity effect in 5-year-old children. *British Journal of Developmental Psychology*, *8*, 411–418.

Khan, M., Fitts, W.S., & Snedeker, J. (2012). *Do adults spontaneously recruit linguistic labels during visual object processing?* Manuscript in preparation.

Manfra, L., & Winsler, A. (2006). Preschool children's awareness of private speech. *International Journal of Behavioral Development (30)*, 537-549.

Mani, N. (2010). When Cup primes Dog: Phono-semantic priming in the toddler lexicon. *Proceedings of the 34th Annual Boston University Conference on Language Development* (pp. 291–302). Boston: Cascadilla Press.

Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: evidence for implicit naming in 18-month-olds. *Psychological science*, *21*, 908–913.

Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*, *121*, 196–206.

Mani, N., Durrant, S., & Floccia, C. (2012). Activation of phonological and semantic codes in toddlers. *Journal of Memory and Language*, *66*, 612–622.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–90.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.

Martínez Manrique, F., & Vicente, A. (2010). " What the …!": The role of inner speech in conscious thought. *Journal of Consciousness Studies*, *17*(9), 141–167.

Mead, G.H. (1934). *Mind, self, and society*. Chicago: University of Chicago Press.

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, *14,* 710–716.

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition*, *35*(3), 494–503.

Miller, G.A. (1996). *The science of words*. New York: W.H. Freeman.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 555–563.

Nagy, W.E. & Herman, P.A. (1987). Breadth and depth of vocabulary knowledge: implications for acquisition and instruction. In M.G. McKeown & M.E. Curtis (Eds.) *The nature of vocabulary acquisition*. Hillsdale, NJ: Erlbaum.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269.

Pinker, S. (1994). *The language instinct.* New York: HarperCollins.

Snedeker, J. (2009).Children's sentence processing. In E. Bavin (Ed.), *The Handbook of Child Language* (pp. 331-338). Cambridge: Cambridge University Press.

Telling, A. L. (2008). *Semantic and phonological context effects in visual search*. (Unpublished doctoral dissertation). University of Birmingham, Birmingham.

Vicente, A., & Martinez Manrique, F. (2011). Inner Speech: Nature and Functions. *Philosophy Compass*, *6*, 209–219.

Vygotsky, L.S. (1987). Thinking and speech (N. Minick, Trans.). In R.W. Rieber & A.S. Carton (Eds.), *The collected works of L.S. Vygotsky: Vol. 1. Problems of general psychology* (pp. 37-285). New York: Plenum. (Original work published 1934)

Whorf, B. L. (1956). Science and linguistics. In J. B. Carroll (Ed.), *Language, thought and reality: selected writings of Benjamin Lee Whorf*. Cambridge; MA: MIT Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 7780–7785.

Winsler, A. (2009). Still talking to ourselves after all these years: A review of current research on private speech. In A. Winsler, C. Fernyhough, I. Montero (Eds.), *Private speech, executive functioning, and the development of verbal self-regulation* (pp. 3-41). New York, NY US: Cambridge University Press.

Winsler, A., & Naglieri, J. (2003). Overt and covert verbal problem-solving strategies: developmental trends in use, awareness, and relations with task performance in children aged 5 to 17. *Child Development*, *74*, 659–78.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology. Learning, memory, and cognition*, *32*(1), 1–14.

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood. In C. Nelson & M. Luciana (Eds), *Handbook of Developmental Cognitive Neuroscience* (2nd Ed.) (pp. 553-574). Cambridge, MA: MIT Press.

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science*, *11*, 125–131.

Appendix

Target	Distractor	Related Prime (SR, VS)	Related Subprime (SR)	Unrelated Prime (SR, VS)
apple	pen	ball (1.9, 2.6)	banana (5.8)	cup (1.3, 1.1)
bread	phone	key (1.0, 1.2)	cake (5.9)	sheep (1.2, 2.1)
bus	tree	carrot (1.3, 1.7)	car (5.9)	duck (1.1, 1.7)
coat	foot	house (2.4, 1.1)	hat (5.2)	ball (1.1, 1.4)
cow	flower	hand (1.2, 1.4)	horse (5.6)	book (1.0, 1.6)
dog	box	cup (1.5, 1.2)	cat (5.7)	house (2.8, 2.0)
sock	plane	sheep (1.8, 2.1)	shoe (5.8)	key (1.0, 1.1)
table	brush	cheese (1.8, 1.2)	chair (5.7)	carrot (1.5, 1.1)
train	spoon	book (1.5, 1.2)	boat (5.4)	hand (1.2, 1.0)
window	bib	duck (1.0, 1.4)	door (5.1)	cheese (1.1, 1.2)

SR indicates the average rating for semantic relatedness to the target and VS indicates the average rating for visual similarity to the target.