Title
Preschooler’s ERPs of online/offline visualizations and embodiment theory

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN
1069-7977

Authors
D'Angiulli, Amedeo
Van Roon, Patricia

Publication Date
2014

Peer reviewed

Preschooler's ERPs of online/offline visualizations and embodiment theory

Patricia Van Roon (patriciavanroon@cmail.carleton.ca)
Department of Neuroscience, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

Amedeo D'Angiulli (amedeo_dangiulli@carleton.ca)
Department of Neuroscience & Institute of Interdisciplinary Studies, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

Abstract

We explored the relationships between a perceptual-attention task and a word-verification task using event-related potential (ERPs) in preschool children. Adopting an embodied multiple representation perspective, we set up the relationships between online (visual attention) and offline (mental imagery) simulation in the two tasks to test key aspects of abstract word acquisition. Online visualization of all word types, during visual selective attention, elicited early frontal and occipital activation (~ 100 ms). The extent of such activation was correlated with a higher occipital late component (800 ms) during offline visualization concurrent with processing more abstract/difficult words. Consistent with developmental vision-language interaction embodiment models, our results support the tenet that the transmission of word meanings by typically developing children may be intimately linked to the visual perceptual contexts in which words are learned.

Key words: embodied cognition; vision-language interaction; Event-Related Potentials; visual attention; word verification.

Visual mental imagery can be defined in terms of embodied cognition; a form of visual simulation or re-enactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind (Barsalou, 2008). A major challenge in the last 30 years has been the question of generated visual mental images and when their use becomes part of cognitive activity for action or language (Paivio, 2007). There is overwhelming evidence that the growth of the latter abilities are at the centre of human development and begin consolidating once they are formally practiced in school through language acquisition (Bornstein & Lamb, 2011). Thus, it seems plausible that there may be a narrow developmental period between infancy and formal schooling in which tasks valid for young children could be pragmatically used to capture correlates of offline simulation (such as visual mental imagery – i.e. sensory and perceptual reactivation) as clearly distinct from online simulation (readiness for action and planning) such as attention, prediction, or expectation on how things look and behave in the perceived world. The purpose of the present work was to explore the interplay between an embodied perceptual activity (goal-directed visual selective attention; VSAT) and an embodied socially-acquired cognitive activity (word-verification; WVT). To this end, we examined the neural correlates of the online and offline visualizations underlying such activities in preschool children. Here, we define "online simulation" as the dynamic processes of early automatic sensorimotor system activation during perception and action, while observing and attending to one or more objects (Gallese, 2009), whereas, we define "offline simulation" as the classic notion of mental imagery, that is, effortful, voluntary, goal-based late activation of larger systems including sensorimotor areas – i.e., ventral and/or dorsal visual pathways – (Decety & Grezes, 2006) specifically associated with re-enactment of previous experience and knowledge for the purpose of deep elaboration (DE) and planning.

Particularly relevant for differentiating which types of brain processes underlie different simulations and cognitive tasks are recent attempts to explain abstract concepts. Indeed, understanding the way we simulate the meaning of abstract words is critical for the evaluation of embodied language theories, according to which language is grounded in perception, action, and emotional systems (Scorolli et al., 2011). According to Paivio's (2007) dual coding theory (DCT), abstract words are represented in a linguistic system while concrete words would be represented in two systems, imagery and linguistic. Consistent with DCT, some embodied perspectives propose that perceptual simulations play an important role in highly image-able concepts or words (e.g., Dove, 2010). Different from DCT, however, would be that abstract words are explained as elaborations of sensorimotor experiences to abstract image-schemas and metaphors [i.e., "container" as and exemplification of the notion of "category"; or a concrete object (newspaper) standing for the action of giving news, Glenberg et al., 2008] whose format could nevertheless be amodal (e.g., symbols detached from perceptual and motor experience). Other embodied proposals, although sharing the tenet of multiple representations, differ from DCT by hypothesizing that not only concrete, but also abstract words are embodied. Notably, Barsalou et al. (2008)'s Perceptual Symbol System (PSS) theorizes that the linguistic system, instantiated in the left-hemisphere language areas, is involved mainly during shallow linguistic elaboration, whereas, deep elaboration involves language in a variety of tasks necessarily implying an offline simulation system comprising the bilateral posterior areas associated with mental imagery.

Yet, most recently it has been argued that the language and situated simulation framework may neglect the social foundation of cognition. Indeed, according to the Words as social Tools (WAT) framework (Borghi et al., 2013), abstract word meanings depend more than concrete word meanings on the everyday experience of being exposed to language in social contexts. On this view, the difference between abstract and concrete words depends on the different acquisition modes, which can be perceptual, linguistic, or mixed, and which change with age and schooling. That is, developmentally abstract words are typically acquired later because it is more difficult to linguistically explain a word meaning than to point at its referent while labeling. Due to this complexity, the acquisition of abstract words typically requires longer and more intense social interactions, and often implies complex linguistic explanations with repetition. In contrast, the process by which young children learn concrete ("sight") words appears effortless and often occurs within a single episode of hearing the word spoken in context. As result, although both concrete and abstract words meanings are linked to sensorimotor and linguistic experience, WAT predicts that we rely more on language to understand abstract word meaning, whereas, we rely more on non-linguistic sensorimotor experience to grasp concrete word meaning (Borghi & Cimatti, 2010).

Embodied multiple representation theories, like WAT, neither assume that abstract and concrete words have different formats, nor assume elaboration from sensorimotor experience into amodal symbols. It postulates that abstract word meanings rely more on the embodied experience of exposure to language than concrete word meanings. The major difference between DCT and theories such as WAT is that, for DCT, abstract words rely on the verbal system alone, while WAT relies on both concrete and abstract words grounded in perception and action systems in addition to a linguistic system which plays a major role for abstract words. Thus, for example, "bicycle" would activate more perceptual and motor experiences than "justice", whereas "justice" would activate less perceptual and motor experiences but more linguistic experiences than "bicycle". That is, linguistic information is more relevant for abstract words, whereas, perception and action are necessary for concrete words.

Borghi et al. (2013) proposed a neuro-functional circuitry mechanism involved in tool-body assimilation as a preliminary computational model of WAT to explain how words may act as tools. Although this model focuses on words as tools for action and addresses components of visual perception and motor responses, it does not sufficiently address the issue of concrete versus abstract word processing. Moreover, the model does not go beyond a brain localization exercise because the actual underlying embodiment (i.e. the simulations which implement the functions) are not specified. The model seems geared towards localization via functional Magnetic Resonance Imaging (fMRI), which lacks the temporal resolution afforded by other more precise techniques such as event-related potentials (ERPs). Finally, the proposed neural circuitry is one of many possible alternative pathways, such as structures/connections self-contained in the visual ventral stream which play a major role in semantics and language processing but not tool use (Humphreys & Forde, 2001).

Given the issues just mentioned, the research questions concerning abstract versus concrete word comprehension, posed in the context of WAT, could perhaps be better framed using models that emphasize vision-language interaction. These types of embodied multiple representations theories stress that, even very early in the time course of processing, vision and language constrain each other and coordinate reciprocally to influence, guide or determine comprehension and semantic elaboration. According to these models, the individual contents of representations generated during language and visual processes may be integrated into salient chunks. This synthesis may occur through mapping between linguistic representations and visual object representations. Various candidate mechanisms have supported empirical evidence for mediating mapping including anticipatory eye movements, memory and attention processes, or online or offline simulations during comprehension and visual sentence-matching tasks (review in Huettig, Mishra & Olivers, 2012). These mechanisms may reflect the embodiment of the synthesis of language and vision, not "representations".

As noted by Mishra & Marmolejo-Ramos (2010), most of these accounts do not explicitly address the role of embodied simulations and the involvement of memory systems. Improving on the current proposal, these authors presented a highly dynamic interactive model which we will refer to as the Dynamic Interaction Vision-Language Approach (DIVLA) wherein "mental representations built during vision-language interaction affect both perception and action …at both a behavioural (events) and neurological (systems) level" (Mishra & Marmolejo-Ramos, 2010; p. 301). This latter model has similarities to WAT but diverges from it in that the visual component has primacy over the motor in terms of representations supporting the enacted or re-enacted (memory) simulations that mediate online and offline mapping between language and visual world.

Extending WAT and DIVLA, we explored the relationships between the event-related potentials (ERPs) correlates underlying a word-verification task and those underlying a perceptual-attention task in young children who had not fully developed a complete sophisticated linguistic system.

Hypotheses & Predictions Given that abstract words do not have a specific object or entity as referent, many of them may be acquired linguistically, i.e., listening to other people explaining their content to us, rather than perceptually. This may be due to the differing degrees of complexity during acquisition: learning to use a word such as "bicycle" is simpler than a word like "justice" and the use of linguistic labels may be crucial for consolidating experiences in social communication context as diverse as those related to the notion of "justice." Therefore, ERPs should show differences known to reflect (1) perceptual processing at parietal sites from about 100-200 ms for concrete words in both tasks, (2) semantic processing and imagery (Sitnikova, et al., 2008) at occipital sites at about 600-800 ms (i.e., the late positive potential) for abstract words in the WVT 5&6, and (3) in line with the WAT hypothesis, abstract and concrete words should both evoke activations grounded in perception and action, therefore, we should observe increased positive amplitudes in the early ERPs for concrete words (i.e. level 4) as compared to abstract words (level 5&6) where the opposite pattern should occur for late ERP amplitudes. Specifically, for the VSAT online processing, the occipital-parietal region (O-P) should show more significant effects than the central-temporal (C-T) which, in turn, should show more significant effects than frontal (F) sites (i.e. O-P > C-T > F). However, the WVT offline processing should show more significant effects at O-P sites than at F sites which, in turn, should show more significant effects than C-T sites (i.e. O-P > F > C-T).

These findings would provide preliminary converging evidence that both online and offline visualizations support word meaning acquisition and implicate overlapping visual pathways and networks. Consequently, development of word comprehension could rely heavily on non-linguistic, visual processes, regardless of whether the word is concrete or abstract. If our expectations are confirmed, the present study would support a developmental derivation of the DIVLA argument (Mishra & Marmolejo-Ramos, 2010) that word meaning transmission in typically developing children may be intimately linked to the visual perceptual contexts in which words are acquired, learned, and remembered.

Materials and Methods

In both tasks, children were tested in a sound-proof electromagnetically shielded EEG booth. Each child was positioned in front of the computer so that the eyes were approximately 58 cm from the screen. Children were monitored visually and aurally. In addition to being widely recognized as the most child-friendly neuroimaging technique currently available, ERPs are ideal because their high temporal resolution allows direct examination of the visual word processing tested in this study. ERPs continuously measure brain processing of visual or auditory stimulus presentations and responses.

Participants Thirteen normal healthy children, ages 4.5-6.5 years (4 female, mean=5.10 yrs, SD=0.75). The final sample of 13 children was obtained after exclusion of four female participants due to excessive EEG artefacts or inadequate performance level (accuracy<75%) in the cognitive task. All participants were Caucasian children with normal or corrected-to-normal vision and no hearing or other known sensory or cognitive impairments. The children lived in the same neighbourhood corresponding to the same geographic area for the daycare centre they attended. All children were from middle-high SES.

All participants scored within 0.5 standard deviations of the mean on the following standardized age-normed control measures: *Behaviour Rating Inventory of Executive Function®–Preschool Version (BRIEF®-P)*from Psychological Assessment Resources (PAR) Inc. (Gioia, Espy, & Isquith, 2009); and the *Child Behaviour Checklist for Ages 1½-5 (CBCL/1½-5)*(Achenbach, 2009). In addition to the above measures, the participants exceeded expectations in the *Early Development Instrument* (Janus et al., 2007) in all developmental domains.

Behavioural Tasks For the *visual selective attention task (VSAT)*, a target stimulus (duck image) and standard stimulus (turtle image) were presented on a computer monitor for 500 ms, followed by a 500 ms ISI. Each child pushed the button when the duck appeared and was asked not to push the button if any other image appeared. The duck was shown 25% of the time and the turtle 75% of the time. Each child completed 12 practice trials followed by an experiment session of 150 randomly presented triggers. Each session lasted about 5 minutes. See Figure 1A.

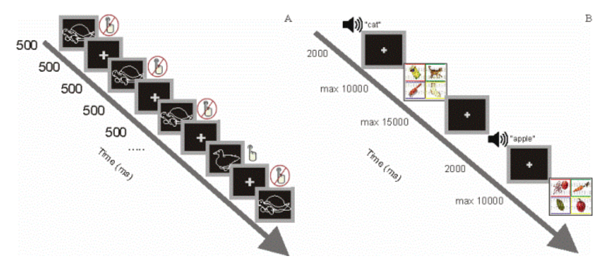

Figure 1. Event timing. A. VSAT stimulus timing and inter-trial fixation cross. B. WVT word presentation, picture array, and inter-trial fixation cross.

The *word-verification task (WVT)* was a computerized version of the standard Peabody Picture Vocabulary Test-Third Edition (PPVT-III), which measures receptive vocabulary and word comprehension (Dunn & Dunn, 1997). The test includes a total of 19 vocabulary sets consisting of 24 target words. Each target word is first presented over insert phones (70 dBHL) followed by an array of four colour images presented together on the monitor and surrounded by a coloured rectangular frame. The coloured frames correspond to the coloured buttons on the response pad. Each child was asked to decide which one of the four pictures on the computer monitor best represented the target word. A pre-recorded English speaking female voice presented the words recorded at a rate of 250 Hz (Figure 1B). The PPVT-III stimulus sets are arranged in order of increasing difficulty (i.e. from concrete to more abstract and complex, please note that difficulty is not a confound but it is how abstractness has been traditionally operationalized for behavioral measurement, see Paivio, 2007) so that the task can be calibrated to the child's vocabulary level as assessed by the norm-based critical range. Importantly, the PPVT administration is set up so that order of presentation and concreteness (abstractness) level are not systematically related. Correlations between the PPVT and kindergarten language comprehension are very strong (*median r* > .65, see Dunn & Dunn, 2007). The PPVT is an accepted measure of semantic elaboration. Performance accuracy rates (> 75%) for the behavioural tasks ensure that the children perform each task according to the instructions. All participants completed Levels 4, 5 and 6, going from concrete to more abstract words. Preliminary RT analysis, however, showed Level 5 did not significantly differ from Level 6; therefore, these levels were collapsed in subsequent ERP analyses.

EEG artefact reduction Trials contaminated by excessive peak-to-peak deflection (i.e. > 100 μV or < -100 μV) due to non-stereotypical or paroxysmal noise at non-ocular electrode sites were manually excluded prior to principal component analysis (PCA) decomposition and ocular correction. Ocular correction was assessed using the Brain Electric Source Analysis (BESA v.5.4.28) Surrogate Model (BR_Brain Regions_LR.bsa). PCA was used to analyze the ocular data within the continuous data file and provide a set of components that represented the variance in the EEG correlated to the eye movements which was subsequently corrected. The proportion of rejected trials was less than 10 percent after artefact correction and removal.

Electrophysiological measures Electrophysiological signals were amplified (gain of 10; Range of +/-200 μV; Accuracy 29 nV/LSB) and low pass filtered at 200 Hz via SynAmps2 and a SCAN™ 4.3 EEG system with a sampling rate of 1000 Hz. Acquisition filters were single-pole Butterworth, 6 dB per octave, 3 dB down at 500 Hz. All electrodes were referenced to a separate nose tip reference electrode and all data were re-referenced to a common average reference.

The EEG data were analyzed and averaged using BESA protocols. Offline averaging was performed such that ERPs were averaged separately for each stimulus type and condition for each electrode with an epoch of -200 ms prestimulus to 1000 ms post-stimulus and baseline corrected from -200 to 0 ms. For the VSAT, ERPs to targets (i.e. ducks) included the individual trials that followed a correct response immediately after target presentation. Averaging was also performed for the standards (i.e. turtle). For the WVT, ERPs were averaged around the auditory word presentation and a second average around the visual image presentation followed by a correct response.

Results

Behavioural data Descriptive analysis on the behavioural data show that VSAT accuracy was very high (M = 89.85%; SD = 5.03) and relatively rapid (*mean RTs* = 681.90 ms, SD = 129.96. Age correlated significantly with accuracy ($r_{(13)}$ = 0.61, $p < 0.05$) but not with RTs. On the longer WVT task, children took about 3.0 s (SD = 0.61) to respond after picture array presentation. Similarly to the VSAT, accuracy was relatively high as all children completed successfully between 2 and 6 target word set and most children were within ±1SD from the mean in terms of age-normed standard scores. There was no correlation between age and performance accuracy or RTs on the WVT test.

ERP data ANOVA Contrasts between mean amplitudes were conducted for the time windows of interest and taking into account standard deviations and standard errors of the baseline mean across the entire ERP epoch; the Simes-Bonferroni procedure was used to correct for multiple comparisons.

VSAT grand averages (Figure 2) of the ERP were calculated for the duck and turtle events, respectively. At 100ms, both duck and turtle ERPs showed significant high bilateral response in the posterior area suggesting visual processing of the images. At 200 ms, strong frontal activity occurred, perhaps indicating working memory to determine the course of action. At 700 ms, bilateral frontal activity coinciding with the button press followed. Focused contrasts revealed the strongest bilateral effects in the midline electrodes. The largest significant differences occurred between attended (duck) and unattended (turtle) waveforms in the midline electrodes between 300 and 500 ms in the occipital electrode, between 400 and 600 ms and 600 and 800 ms in the parietal electrode, and between 500 and 1000 ms in the frontal electrode. All contrasts were $t_{(13)} = 2.27$, $p < 0.05$. On average, the largest peak amplitude detected in the examined time windows was 2.84 μV (relative to baseline activity estimated $Z = 2.60$, $p < 0.01$).

WVT grand average (Figure 3) ERPs for level 4 and combined level 5&6 show early co-activation of opposite polarity in right anterior and bilateral posterior electrodes from 100 to 200 ms. After 300 ms, the activity localized to the right side across the centro-parietal and temporal electrodes. To verify deep elaboration (DE) after hearing the target word, ERPs split of correct WVT trials were averaged. Fz was the most representative electrode between 400 and 900 ms, and showed a significant difference between the waveforms of correctly identified words although the direction of the effect is reversed at about half of this interval, ($t_{13} = 4.23$, $p < 0.01$).

To analyze word-verification events of concrete and abstract WVT word sets, concrete words were included only from the first set (Level 4) and abstract words were included from the very last successfully completed sets for each participant (Levels 5&6). The most important and significant differences were found at the right occipital electrode (O2). As predicted, there was a higher positive activation which occurred concurrently with concrete words in the first 100 ms of the WVT task ($t_{(13)} = 6.52$; $p <$ 0.0001). Conversely, there was higher positive activation which occurred concurrently with abstract words in the later 750-850 ms interval ($t_{(13)} = 7.71$; $p < 0.0001$).

Control analysis Multiple regression analyses showed no significant relationships between ERPs and control variables, indicating that findings are not due to confounds of these other factors.

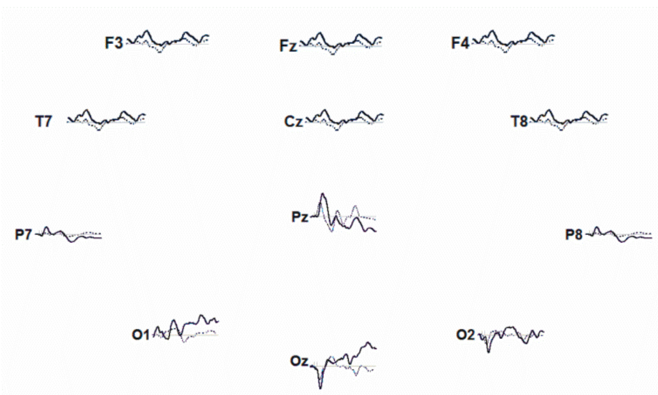

Figure 2. VSAT showing ERPs for all recorded channels. Negative is plotted upwards. Solid line is target (correct response to duck) and dashed line is standard (turtle). Scale is -5 μV with negative plotted upwards.

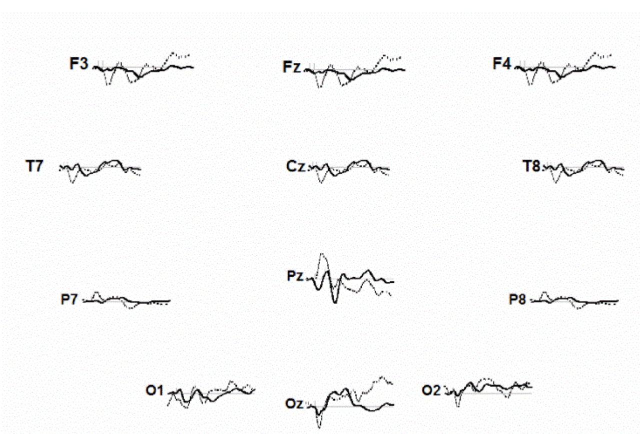

Figure 3: The WVT concrete (level 4) and abstract (level 5&6) visual task. The concrete line is solid and abstract is dashed. Scale is -5 μV with negative plotted upwards.

A final analysis of 100 ms bins over the 0-699 ms period for each task (Berman and Friedman, 1995) examined the maximum activations and show that the overall VSAT pattern of O-P>C-T>F matches our predication whereas the WVT pattern P>O>T>C-F was not as predicted and shows much greater parietal activation for abstract words.

Discussion

Our main findings show that, during word verification, early ERP activity for concrete (easier, less effortful) words is higher than abstract (more difficult, more effortful) words. However, the opposite pattern occurs for late ERP amplitudes. In addition, behavioural performance measures (i.e. accuracy and reaction times) as well as ERP activity show that children carry out the tasks in a similar way. Specifically, the visual components of each task produce an immediate, strong early response from occipital electrodes followed by a response from frontal electrodes.

The latter results suggest that children recruit neural dynamics that activate concomitantly with visual perceptual processing (i.e. online visualization). We then tested whether a positive potential was elicited in the same cortical area during a word comprehension task when, some milliseconds after hearing a word, the children processed its meaning for a subsequent discriminatory response (i.e. selecting the correct picture of the named object among three other possibilities). The results suggest that verifying the meaning of abstract words relies on neurocognitive processes associated with offline visualization, while verifying the meaning of concrete words relies on processes associated with online visualization. Visualization results indicate a common substrate in the visual system, not only in the early stages of deep elaboration (comprehension) but also in the later ones.

In both tasks, ERPs show involvement of frontal motor processes at 700-750 ms; however, this may reflect the button press in response to the VSAT picture selection, whereas for the WVT, this would likely indicate reasoning or problem solving for abstract word processing. The tasks reflect aspects of word comprehension in children that, to our knowledge, have not been adequately investigated by embodiment research. This study highlights new aspects of embodiment that should be integrated into the corpus of evidence and data already accumulated.

One possible interpretation of these findings is that, as originally proposed by Vygotsky (1962), words are social tools in the sense that they regulate inter-subjective communication by driving people's attention and visualization (online/offline simulations) in a direction leading to the sharing of intended experience and knowledge. Words evoke images of the world that can be exchanged among interlocutors for different types of social communication; language in this respect is not unique. Images, pictures, gestures, and non-verbal behaviours are all embodied instruments of social communication constantly used, for example, in learning activities carried out by children. Consider the enormous role of picture books for literacy in Western culture. These data show that vision-language integration in the case of single-word processing works in much the same way whether words are concrete or abstract. However, these findings also suggest that more visual information is needed in processing abstract rather than concrete words, as a result of how the former ones are acquired. Presumably, this involves a richer, more complex visualization, more cognitive control and more effortful elaboration (Borghi et al., 2011). Future considerations for the present findings would be to examine the effects of verbs (actions) and the role of nouns (objects) and, perhaps, extend this to sentence processing.

In conclusion, this study provides neurophysiological evidence underlying attention and language comprehension in young children that, to our knowledge, has not yet been addressed. Overall, the findings support the view that social transmission of word meaning in children depends greatly on the attended visual experiences in the contexts in which words are acquired, learned, and remembered.

References

Achenbach, T. M. (2009). The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications. Burlington, VT: University of Vermont Research Center for Children, Youth and Families.

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

Barsalou, L. W., Santos, A., Simmons, W.K., & Wilson, C.D. (2008). Language and simulation in conceptual processing. In M. De Vega, A.M. Glenberg, & A.C. Graesser, A. (Eds.). *Symbols, embodiment, and meaning*. Oxford: Oxford University Press.

Berman, S., & Friedman, D. (1995). The development of selective attention as reflected by event-related brain potentials. Journal of Experimental Child Psychology, 59(1), 1–31. doi:10.1006/jecp.1995.1001Borghi, A.M., & Cimatti, F. (2010). Embodied cognition and beyond: acting and sensing the body. *Neuropsychologia, 48(3),* 763-73. doi: 10.1016/j.neuropsychologia.2009.10.029.

Borghi, A.M., Flumini, A., Cimatti, F., Marocco, D., & Scorolli, C. (2011). Manipulating objects and telling words: a study on concrete and abstract words acquisition. *Frontiers in Psychology, 2*(15), 1-14. doi: 10.3389/fpsyg.2011.00015. eCollection 2011.

Borghi, A.M., Scorolli, C., Caligiore, D., Baldassarre, G., & Tummolini, L. (2013). The embodied mind extended: using words as social tools. *Frontiers in Psychology, 4*(214), 1-10. doi: 10.3389/fpsyg.2013.00214. eCollection 2013.

Bornstein, M. H., & Lamb, M. E. (Eds.) (2011). *Developmental Science: An Advanced Textbook*. New York, NY: Psychology Press.

Decety, J., & Grèzes, J. (2006). The power of simulation: Imagining one's own and other's behaviour. *Brain Research, 1079,* 4-14.

Dove, G.(2010).An additional heterogeneity hypothesis. *Behavioural and Brain Sciences, 33,* 209–210.

Dunn, L. M., & Dunn, L. M. (1997). Peabody Picture Vocabulary Test–Third Edition. Circle Pines, MN: AGS Publishing.

Dunn, L. M., & Dunn, D. M. (2007). The Peabody Picture Vocabulary Test, Fourth Edition. Bloomington, MN: NCS Pearson, Inc.

Gallese, V. (2009). Motor abstraction: a neuroscientific account of how action goals and intentions are mapped and understood. *Psychological Research, 73,* 486-98.

Gioia, G. A., Espy, K. A., & Isquith P. K. (2003). Behaviour Rating Inventory of Executive Function®-Preschool Version (BRIEF®-P).

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., & Buccino, G. (2008). Processing abstract language modulates motor system activity. *Quarterly Journal of Experimental Psychology, 61,* 905–919.

Humphreys, G.W., & Forde, E.M. (2001).Hierarchies, similarity, and interactivity in object recognition: "category-specific" neuropsychological deficits. *Behavioural Brain Sciences, 24(3),* 453-76, discussion 476-509.

Huettig, F., Mishra, R.K., & Olivers, C.N. (2012). Mechanisms and representations of language-mediated visual attention. *Frontiers in Psychology, 2,* 394. doi: 10.3389/fpsyg.2011.00394.

Janus, M., Brinkman, S., Duku, E., Hertzman, C., Santos, R., Sayers, M. & Schroeder, J. (2007). The Early Development Instrument: A Population-based measure for communities. A handbook on development, properties, and use. Offord Centre for Child Studies, Hamilton, ON.

Mishra, R.K., & Marmolejo-Ramos, F. (2010). On the mental representations originating during the interaction between language and vision. *Cognitive Processing, 11(4),* 295-305, doi: 10.1007/s10339-010-0363-y.

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach.* Mahwah, NJ: Erlbaum.

Scorolli, C., Binkofski, F., Buccino, G., Nicoletti, R., Riggio, L., & Borghi, A. M. (2011). Abstract and concrete sentences, embodiment, and languages. *Frontiers in Psychology, 2,* 227.

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience, 20,* 2037-2057.

Vygotsky, L. S. (1962). *Thought and language.* Cambridge, UK: MIT Press.