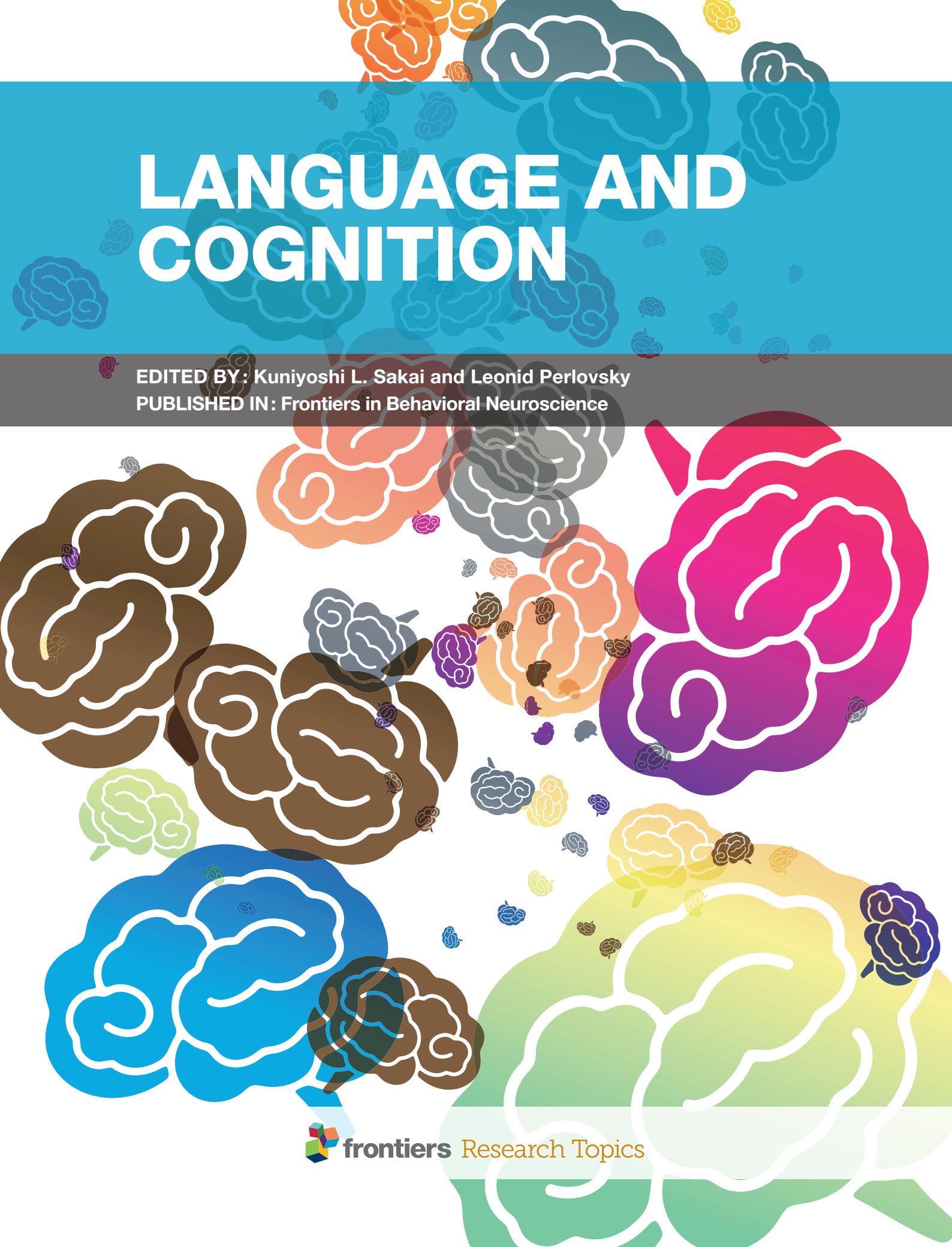
LANGUAGE AND COGNITION

EDITED BY: Kuniyoshi L. Sakai and Leonid Perlovsky
PUBLISHED IN: Frontiers in Behavioral Neuroscience



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88919-627-2
DOI 10.3389/978-2-88919-627-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

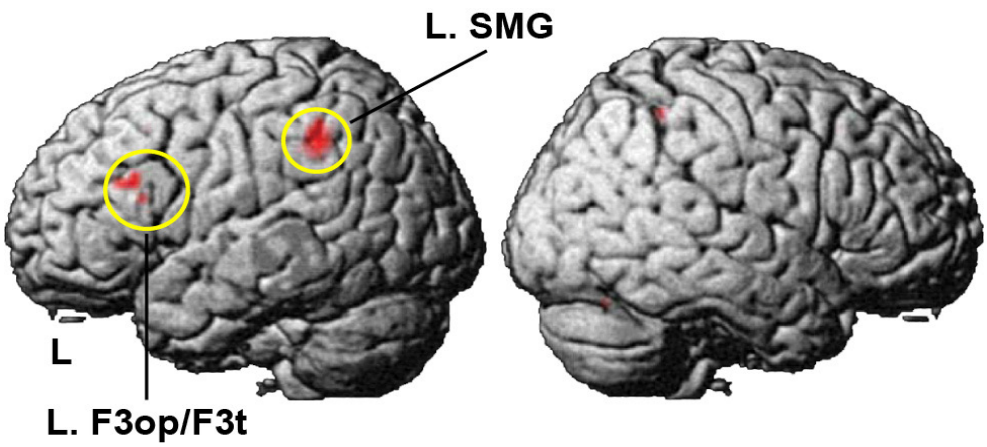
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

LANGUAGE AND COGNITION

Topic Editors:
Kuniyoshi L. Sakai, The University of Tokyo, Japan
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA



Functional evidence of a syntax-related network.
In a given sentence with syntactic structures, the greater depth of merged subtrees, but not the linear order of words, elicited significant activation in the pars opercularis and pars triangularis of the left inferior frontal gyrus (L. F3op/F3t), as well as the left supramarginal gyrus (L.SMG). Activations (red) are projected onto the left (L) and right lateral surfaces of a standard brain. Taken from Ohta, Fukui, and Sakai (2013; doi: 10.3389/fnbeh.2013.00204).

Interaction between language and cognition remains an unsolved scientific problem. What are the differences in neural mechanisms of language and cognition? Why do children acquire language by the age of six, while taking a lifetime to acquire cognition? What is the role of language and cognition in thinking? Is abstract cognition possible without language? Is language just a communication device, or is it fundamental in developing thoughts? Why are there no animals with human thinking but without human language? Combinations even among 100 words and 100 objects (multiple words can represent multiple objects) exceed the number of all the particles in the Universe, and it seems that no amount of experience would suffice to learn these associations. How does human brain overcome this difficulty?

Since the 19th century we know about involvement of Broca's and Wernicke's areas in language. What new knowledge of language and cognition areas has been found with

fMRI and other brain imaging methods? Every year we know more about their anatomical and functional/effective connectivity. What can be inferred about mechanisms of their interaction, and about their functions in language and cognition? Why does the human brain show hemispheric (i.e., left or right) dominance for some specific linguistic and cognitive processes? Is understanding of language and cognition processed in the same brain area, or are there differences in language-semantic and cognitive-semantic brain areas? Is the syntactic process related to the structure of our conceptual world?

Chomsky has suggested that language is separable from cognition. On the opposite, cognitive and construction linguistics emphasized a single mechanism of both. Neither has led to a computational theory so far. Evolutionary linguistics has emphasized evolution leading to a mechanism of language acquisition, yet proposed approaches also lead to incomputable complexity.

There are some more related issues in linguistics and language education as well. Which brain regions govern phonology, lexicon, semantics, and syntax systems, as well as their acquisitions? What are the differences in acquisition of the first and second languages? Which mechanisms of cognition are involved in reading and writing? Are different writing systems affect relations between language and cognition? Are there differences in language-cognition interactions among different language groups (such as Indo-European, Chinese, Japanese, Semitic) and types (different degrees of analytic-isolating, synthetic-inflected, fused, agglutinative features)? What can be learned from sign languages?

Rizzolatti and Arbib have proposed that language evolved on top of earlier mirror-neuron mechanism. Can this proposal answer the unknown questions about language and cognition? Can it explain mechanisms of language-cognition interaction? How does it relate to known brain areas and their interactions identified in brain imaging?

Emotional and conceptual contents of voice sounds in animals are fused. Evolution of human language has demanded splitting of emotional and conceptual contents and mechanisms, although language prosody still carries emotional content. Is it a dying-off remnant, or is it fundamental for interaction between language and cognition? If language and cognitive mechanisms differ, unifying these two contents requires motivation, hence emotions. What are these emotions? Can they be measured? Tonal languages use pitch contours for semantic contents, are there differences in language-cognition interaction among tonal and atonal languages? Are emotional differences among cultures exclusively cultural, or also depend on languages?

Interaction of language and cognition is thus full of mysteries, and we encouraged papers addressing any aspect of this topic.

Citation: Sakai, K. L., Perlovsky, L., eds. (2015). Language and Cognition. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-627-2

Table of Contents

frontiers in
BEHAVIORAL NEUROSCIENCE

EDITORIAL
published: 16 December 2014
doi: 10.3389/fnbeh.2014.00436

Language and Cognition

Leonid Perlovsky*[1]* and Kuniyoshi L. Sakai*[2]*

[1] *Department of Electrical Engineering, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA*
[2] *Department of Basic Science, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan*
**Correspondence: lperl@rcn.com; sakai@mind.c.u-tokyo.ac.jp*

Edited and reviewed by:
Nuno Sousa, University of Minho, Portugal

Keywords: language, cognition, brain, functional imaging, emotion

Interaction between language and cognition remains an unsolved scientific problem. What are the differences in neural mechanisms of language and cognition? Why do children acquire language by the age of six, while taking a lifetime to acquire cognition? What is the role of language and cognition in thinking? Is abstract cognition possible without language? Is language just a communication device, or is it fundamental in developing thoughts? Why are there no animals with human thinking but without human language? Combinations even among 100 words and 100 objects (multiple words can represent multiple objects) exceed the number of all the particles in the Universe, and it seems that no amount of experience would suffice to learn these associations. How does human brain overcome this difficulty?

Since the nineteenth century we know about involvement of Broca's and Wernicke's areas in language. What new knowledge about the brain regions responsible for language and cognition has been found with fMRI and other brain imaging methods? Every year we know more about their anatomical and functional/effective connectivity. What can be inferred about their interactions and functions in language and cognition? Why does the human brain show hemispheric (i.e., left or right) dominance for some specific linguistic and cognitive processes? Is linguistic and cognitive comprehension processed in the same or different regions? Do the syntactic processes affect the structure of our conceptual world?

Such issues regarding brain functions and mind have been increasingly drawing attention from various fields in recent years, and investigations that go beyond the boundaries of previous fields of study are becoming necessary. The need for study spanning the brain and the mind has given birth to a new discipline, such as cognitive neuroscience, neurolinguistics, biolinguistics, etc. We assume that mind is a part of brain function, and we tentatively define the mind as a combination of three main cognitive factors: perception, memory, and consciousness. Language is created by mind, yet, once uttered, words return to the mind, where they are understood. The cycle from the mind to the language and then from the language to the mind, is *recursive*, in that the language produced by the mind comes back to the mind once again. This recursiveness is important when considering the relationship between language and mind.

When viewed language and mind as a whole system, it is evident that the functions of language are part of the brain system at the same time as being involved in the workings of the mind. Moreover, information is exchanged between language and each of perception, memory, and consciousness in both directions. Namely, language is involved in both reciprocal and recursive information exchange with each element of the mind. Since language is tightly linked to the mind, it would be more natural to assume that language is a part of the mind than to think it is an entity which exits outside the mind. The study of language is, in essence, to understand a part of the "human" mind. The more we study the language used by humans, the more we will understand the structure of the mind.

Chomsky has suggested that language is separable from cognition (Berwick et al., 2013), and this notion has been well supported by functional imaging experiments in neuroscience (Sakai, 2005). On the opposite, cognitive and construction linguistics emphasized a single mechanism of both. Neither has led to a computational theory so far, but language is learned early in life with only limited cognitive understanding of the world (Perlovsky, 2009). Evolutionary linguistics has emphasized evolution leading to a mechanism of language acquisition, yet proposed approaches also lead to incomputable complexity. Papers in this volume report new knowledge on interacting language and cognition, still there remains more questions than answers.

In animals, emotional and conceptual contents of voice sounds are fused. Evolution of human language has demanded splitting of emotional and conceptual contents, as well as of their mechanisms, although language prosody still carries emotional content. Is it a dying-off remnant, or is it fundamental for interaction between language and cognition? If language and cognitive mechanisms differ, unifying these two contents requires motivation, hence emotions. What are these emotions? Can they be measured? If tonal languages use pitch contours for semantic contents, are there differences in language-cognition interaction among tonal and atonal languages? Are emotional differences among cultures exclusively cultural, or also depend on languages?

This volume introduces a broad range of research addressing these topics, including three opinion articles, one hypothesis and theory article, eight original research articles, and a pair of an opinion article and a general commentary article. Their summaries are as follows.

First, Perlovsky (2013) introduces joint acquisition, dual hierarchy, and emotional prosody of language and cognition, such that emotional prosody may perform a fundamental function in connecting sounds and meanings of words. Vicario (2013) discusses about FOXP2 gene and language development, which might inform us about the origin of language. Perry and Lupyan

(2013) explain that language and thought are different but strongly interacting abilities, based on the online manipulation of linguistic activity.

Next, Ohta et al. (2013) propose computational principles of syntax in the regions specialized for language, thereby integrating theoretical linguistics and functional neuroimaging. Nagels et al. (2013b) present an fMRI study on the neural substrates of figurative language during natural speech perception. De La Cruz et al. (2013) show that finger counting helps cognitive robots to learn words. Straube et al. (2013) suggest that abstract information conveyed by speech and gesture may be processed independent of modality. Tilles and Fontanari (2013) examine reinforcement and inference in cross-situational word learning. Nagels et al. (2013a) indicate the role of semantic abstractness and perceptual category in processing speech accompanied by gestures. Zhong et al. (2013) study a self-organizing pre-symbolic neural model representing sensorimotor information. Shuai and Gong (2013) analyze temporal relationships between top-down and bottom-up processing in lexical tone perception. Vicario and Rumiati (2013) demonstrate how notions of left and right affect processing of trading verbs.

We end the volume with a highly-popular discussion on the role of open access publications in linguistics, contributed by Haspelmath (2013) and Bragazzi (2013).

REFERENCES

Berwick, R. C., Friederici, A. D., Chomsky, N., and Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends Cogn. Sci.* 17, 89–98. doi: 10.1016/j.tics.2012.12.002

Bragazzi, N. L. (2013). The importance of open access publishing in the field of Linguistics for spreading scholarly knowledge and preserving languages diversity in the era of the economic financial crisis. *Front. Behav. Neurosci.* 7:91. doi: 10.3389/fnbeh.2013.00091

De La Cruz, V. M., Di Nuovo, A., Di Nuovo, S., and Cangelosi, A. (2013). Making fingers and words count in a cognitive robot. *Front. Behav. Neurosci.* 7:13. doi: 10.3389/fnbeh.2014.00013

Haspelmath, M. (2013). Why open-access publication should be nonprofit—a view from the field of theoretical language science. *Front. Behav. Neurosci.* 7:57. doi: 10.3389/fnbeh.2013.00057

Nagels, A., Chatterjee, A., Kircher, T., and Straube, B. (2013a). The role of semantic abstractness and perceptual category in processing speech accompanied by gestures. *Front. Behav. Neurosci.* 7:181. doi: 10.3389/fnbeh.2013.00181

Nagels, A., Kauschke, C., Schrauf, J., Whitney, C., Straube, B., and Kircher, T. (2013b). Neural substrates of figurative language during natural speech perception: an fMRI study. *Front. Behav. Neurosci.* 7:121. doi: 10.3389/fnbeh.2013.00121

Ohta, S., Fukui, N., and Sakai, K. L. (2013). Computational principles of syntax in the regions specialized for language: integrating theoretical linguistics and functional neuroimaging. *Front. Behav. Neurosci.* 7:204. doi: 10.3389/fnbeh.2013.00204

Perlovsky, L. (2013). Language and cognition—joint acquisition, dual hierarchy, and emotional prosody. *Front. Behav. Neurosci.* 7:123. doi: 10.3389/fnbeh.2013.00123

Perlovsky, L. I. (2009). Language and cognition. *Neural Netw.* 22, 247–257. doi: 10.1016/j.neunet.2009.03.007

Perry, L. K., and Lupyan, G. (2013). What the online manipulation of linguistic activity can tell us about language and thought. *Front. Behav. Neurosci.* 7:122. doi: 10.3389/fnbeh.2013.00122

Sakai, K. L. (2005). Language acquisition and brain development. *Science* 310, 815–819. doi: 10.1126/science.1113530

Shuai, L., and Gong, T. (2013). Temporal relation between top-down and bottom-up processing in lexical tone perception. *Front. Behav. Neurosci.* 7:97. doi: 10.3389/fnbeh.2014.00097

Straube, B., He, Y., Steines, M., Gebhardt, H., Kircher, T., Sammerand, G., et al. (2013). Supramodal neural processing of abstract information conveyed by speech and gesture. *Front. Behav. Neurosci.* 7:120. doi: 10.3389/fnbeh.2013.00120

Tilles, P. F. C., and Fontanari, J. F. (2013). Reinforcement and inference in cross-situational word learning. *Front. Behav. Neurosci.* 7:163. doi: 10.3389/fnbeh.2013.00163

Vicario, C. M. (2013). FOXP2 gene and language development: the molecular substrate of the gestural-origin theory of speech? *Front. Behav. Neurosci.* 7:99. doi: 10.3389/fnbeh.2013.00099

Vicario, C. M., and Rumiati, R. I. (2013). Left-right compatibility in the processing of trading verbs. *Front. Behav. Neurosci.* 7:16. doi: 10.3389/fnbeh.2014.00016

Zhong, J., Cangelosi, A., and Wermter, S. (2013). Toward a self-organizing pre-symbolic neural model representing sensorimotor primitives. *Front. Behav. Neurosci.* 7:22. doi: 10.3389/fnbeh.2014.00022

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 November 2014; accepted: 01 December 2014; published online: 16 December 2014.

*Citation: Perlovsky L and Sakai KL (2014) Language and Cognition. Front. Behav. Neurosci. **8**:436. doi: 10.3389/fnbeh.2014.00436*

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

OPINION ARTICLE
published: 19 September 2013
doi: 10.3389/fnbeh.2013.00123

Language and cognition—joint acquisition, dual hierarchy, and emotional prosody

Leonid Perlovsky*

The AFRL and Athinoula A. Martinos Center for Biomedical Imaging, Harvard University, Charlestown, MA, USA
**Correspondence: lperl@rcn.com*

Edited by:
Kuniyoshi L. Sakai, The University of Tokyo, Japan

Keywords: language, cognition, acquisition, dual hierarchy, prosody, emotion

FUNCTION OF LANGUAGE AND COGNITION IN THINKING

Do we think with language, or is it just a communication device used for expression of completed thoughts? What is a difference between language and cognition? Chomsky (1995) suggested that these two abilities are separate and independent. Cognitive linguistics emphasizes a single mechanism for both (Croft and Cruse, 2004). Evolutionary linguistics considers the process of transferring language from one generation to the next one (Cangelosi and Parisi, 2002; Christiansen and Kirby, 2003; Hurford, 2008). This process is a "bottleneck" that forms the language. Brighton et al. (2005) demonstrated emergence of compositional language due to this bottleneck. Still, none of these approaches resulted in a computational theory explaining how humans acquire language and cognition. Here I discuss a computational model overcoming previous difficulties and based on a hypothesis that language and cognition are two separate and closely integrated abilities. I identify their functions and discuss why human thinking ability requires both language and cognition.

Among fundamental mechanisms of cognition are mental representations, memories of objects and events (Perlovsky, 2001, 2006a). The surrounding world is understood by matching mental representations to patterns in sensor signals. However, mathematical modeling of this process since the 1950s met with difficulties. The first difficulty is related to a need to consider combinations of sensor signals, objects, and events. The number of combinations is very large and even a limited number of signals or objects form a very large number of combinations, exceeding all interactions of all elementary particles in a lifetime of the Universe (Perlovsky, 1998). This is known as combinatorial complexity, CC. This difficulty in modeling the mind has been overcome by dynamic logic (Perlovsky, 2001, 2006a,b, 2007a; Perlovsky et al., 2011). Whereas classical logic considers static statements such as "this is a chair," dynamic logic models processes from vague to crisp representations. These processes do not need to consider combinations, an initial vague state of a "chair" matches any object in the field of view, and at the end of the process it matches the chair actually present, without CC.

The second difficulty is similar still even more complex. It is related to the fact that "events" and "situations" in the world do not necessarily exist "ready for cognition." There are many combinations of percepts and objects, a near infinity, events and situations important for understanding and learning have to be separated from those that are just random collections of meaningless percepts or random objects (Perlovsky and Ilin, 2012). Events and situations recognized by non-human animals are very limited compared to human abilities to differentiate events in the world. Human cognitive abilities acquire their power due to language. Language is "easier" to learn than cognitive representations. Language representations: words, phrases exist in the surrounding language "ready made," created during millennia of cultural evolution. Therefore, language could be learned without much real-life experience; only interactions with language speakers are required. Every child learns language early in life before acquiring full cognitive understanding of events and their cognitive meanings. Thus, language is learned early in life with only limited cognitive understanding of the world (Perlovsky, 2009a, 2012c). Cognitive representations of situations and abstract concepts initially exist in vague states. Throughout the rest of life, language guides acquisition of cognitive representations from experience. Vague cognitive representations become more crisp and concrete. Thinking involves both language and cognition, and as we discuss later thinking about abstract ideas usually involves language more than cognition, not too different from thinking by children.

THE DUAL HIERARCHY

Cognitive representations are organized in mind in an approximate hierarchy (Grossberg, 1988) from sensor-motor percepts near "bottom," to objects "higher up," to situations, and to still more abstract cognitive representations. Language representations are organized in a parallel hierarchy from sounds, and words for objects and situations, to phrases, and to more abstract language representations. Our previous discussion can be described by an integrated mathematical model of language and cognition forming a dual hierarchy (Perlovsky, 2009a), as illustrated in **Figure 1**. Neural evidence suggests that the hierarchy is approximate, not as definite as shown in this figure.

Hierarchical organization of cognition and related brain structures are reviewed in (Badre, 2008). In particular, anterior-posterior axis corresponds to a gradient of abstract-concrete cortex functions. Hierarchical organization of language functions is also well established. However, hierarchical organization of language does not correspond to a particular spatial axis in the brain, it is distributed (Price, 2012). Therefore, the dual hierarchy in **Figure 1** is a functional hierarchy not organized along a spatial axis in the brain as in this figure. A fundamental aspect of acquiring

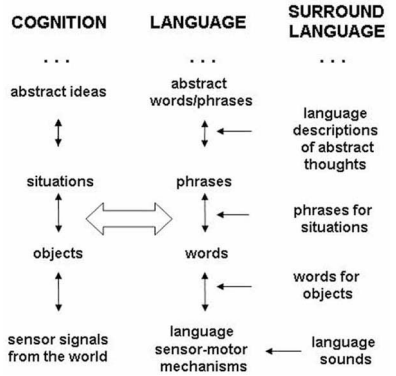


FIGURE 1 | The dual hierarchy. Language and cognition are organized into approximate dual hierarchy. Learning language is grounded in the surrounding language throughout the hierarchy. Cognitive hierarchy is grounded in experience only at the very "bottom."

mental representations is interaction between higher and lower layer representations (top and bottom layers). In this interaction a lower layer representations are organized in more abstract and general concept-representations at a higher layer. These interactions are referred to as bottom-up and top-down signals (BU and TD) indicated in **Figure 1** by vertical arrows.

Mathematical model of the dual hierarchy is described in Perlovsky (2009a, 2012c) and Perlovsky and Ilin (2010, 2012). This model explains many facts about thinking, language, and cognition, which has remained unexplainable and would be considered mysteries, if not so commonplace.

The dual model makes a number of experimentally testable predictions. (1) It explains functions of language and cognition in thinking: cognitive representations model surrounding world, relations between objects, events, and abstract concepts. Language stores culturally accumulated knowledge about the world, yet language is not directly connected to objects, events, and situations in the world. Language guides acquisition of cognitive representations from random percepts and experiences, according to what is considered worth learning and understanding in culture. Events that are not described in language are likely not even noticed or perceived in cognition. (2) Whereas language is acquired early in life, acquiring cognition takes a lifetime. The reason is that language representations exist in surrounding language "ready-made," acquisition of language requires only interaction with language speakers, but does not require much life experience. Cognition on the opposite requires life experience. (3) This is the reason why abstract words excite only language regions of brain, whereas concrete words excite also cognitive regions (Binder et al., 2005). The dual model predicts that abstract concepts are often understood as word descriptions, but not in terms of objects, events, and relations among them. (4) This model explains why language is acquired early in life, whereas cognition takes a lifetime. It also explains why children can acquire the entire hierarchy of language including abstract words without experience necessary for understanding them. (5) Since dynamic logic is the basic mechanism for learning language and cognitive representations, the dual model suggests that language representations become crisp after language is learned (5–7 years of age), however, cognitive representations may remain vague for much longer; the vagueness is exactly the meaning of "continuing learning," this takes longer for more abstract and less used concepts. (6) The dual model gives mathematical description of the recursion mechanism (Perlovsky and Ilin, 2012). Whereas Hauser et al. (2002) postulate that recursion is a fundamental mechanism in cognition and language, the dual model suggests that recursion is not fundamental, hierarchy is a mechanism of recursion.

(7) Another mystery of human-cognition, not addressed by cognitive or language theories, is basic human irrationality. This has been widely discussed and experimentally demonstrated following discoveries of Tversky and Kahneman (1974), leading to the 2002 Nobel Prize. According to the dual hierarchy model, the "irrationality" originates from the dichotomy between cognition and language. Language is crisp and conscious while cognition might be vague and ignored when making decisions. Yet, collective wisdom accumulated in language may not be properly adapted to one's personal circumstances, and therefore be irrational in a concrete situation. In the 12th century Maimonides wrote that Adam was expelled from paradise because he refused original thinking using his own cognitive models, but ate from the tree of knowledge and acquired collective wisdom of language (Levine and Perlovsky, 2008).

EMOTIONAL PROSODY AND ITS COGNITIVE FUNCTION

The dual model implies connections between language and cognitive representations, indicated by a wide horizontal arrow in **Figure 1**. These neural connections have to be developed and maintained. This requires motivation, in other words, emotions. These emotions must be in addition to utilitarian meanings of words, otherwise only practically useful words would be connected to their cognitive meanings. Also these emotions must "flow" from language to cognition, so that language is able to perform its cognitive function of guiding acquisition of cognitive representations, organizing experience according to cultural contents of language. These emotions therefore must be contained in language sounds, before cognitive contents are acquired.

This requirement of emotionality of language sounds is surprising and contradictory to assumed direction of evolution of language. Evolution of the language ability required rewiring of human brain in the direction of freeing vocalization from uncontrollable emotions (Deacon, 1997; Perlovsky, 2009b). Yet, the dual model requires that language sounds be emotional. Emotionality of human voice is most pronounced in songs (Perlovsky, 2010, 2012a,d, 2013b). Emotions of everyday speech are low, unless affectivity is specifically intended. We may not notice emotions in everyday "non-affective" speech. Nevertheless, this emotionality is important for developing the cognitive part of the dual model. If language is highly emotional, speakers are passionate about what they say, however, evolving new meanings might be slow, emotional ties of sounds to old meanings might be "too strong." If language is low-emotional, new words are easy to create, however, motivation to develop the cognitive part of the dual model might be low, the real-world meaning of language sound might be lost. Cultural values might be lost as well. Indeed languages differ in how strong are emotional connections between sounds and meanings. This leads to cultural differences. Thus, the dual model leads to Emotional Sapir-Whorf Hypothesis (Perlovsky, 2007b, 2009b, 2012b). Strength

of emotional connections between sound and meaning depends on language inflections. In particular, after English lost most of its inflections, it became a low emotional language, powerful for science and engineering. At the same time English is losing autonomous connections to cultural values that used to be partially inherent in language sounds. Fast change of cultural values during recent past is usually attributed to progress in thinking, whereas effects of change in emotionality of language sounds have not been noticed.

Emotional prosody can be important for overcoming cognitive dissonance. Cognitive dissonance is a discomfort due to holding contradictory cognitions (Festinger, 1957; Harmon-Jones et al., 2009). It is resolved by discarding contradictions. If a new word contradicts existing knowledge its meaning might be discarded. Emotional prosody as well as songs could be fundamental mechanisms that overcome cognitive dissonance and enable keeping new contradictory knowledge (Masataka and Perlovsky, 2012; Perlovsky, 2013a).

CONCLUSION AND EXPERIMENTAL PREDICTIONS

This article advances a hypothesis about functions of language and cognition in thinking, and possible model of their interactions. This is the only computable model explaining a number of mysteries about language and cognition and overcoming computational difficulties. It makes a number of predictions that could be experimentally tested, including the following: cognitive representations model the world, while language representations only model language; abstract cognitive representations can only be acquired due to language; abstract cognition is more clearly represented in language whereas cognitive representations may remain vague throughout life.

ACKNOWLEDGMENTS

I am thankful for discussions with my colleagues, Michel Cabanac and Nobuo Masataka.

REFERENCES

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends Cogn. Sci.* 12, 193–200. doi: 10.1016/j.tics.2008.02.004

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., and Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *J. Cogn. Neurosci.* 17, 1–13. doi: 10.1162/0898929054021102

Brighton, H., Smith, K., and Kirby, S. (2005). Language as an evolutionary system. *Phys. Life Rev.* 2, 177–226. doi: 10.1016/j.plrev.2005.06.001

Cangelosi, A., and Parisi, D. (eds.). (2002). *Simulating the Evolution of Language.* London: Springer. doi: 10.1007/978-1-4471-0663-0

Chomsky, N. (1995). *The Minimalist Program.* Cambridge: MIT Press.

Christiansen, M. H., and Kirby, S. (2003). *Language Evolution.* New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199244843.001.0001

Croft, W., and Cruse, D. A. (2004). *Cognitive Linguistics.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511803864

Deacon, T. W. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain.* New York, NY: Norton.

Festinger, L. (1957). *A Theory of Cognitive Dissonance.* Stanford, CA: Stanford University Press.

Grossberg, S. (1988). *Neural Networks and Natural Intelligence.* Cambridge: MIT Press.

Harmon-Jones, E., Amodio, D. M., and Harmon-Jones, C. (2009). "Action-based model of dissonance: a review, integration, and expansion of conceptions of cognitive conflict," in *Adv Exp Soc Psychol 41,* ed M. P. Zanna (Burlington, ON: Academic Press), 119–166.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569

Hurford, J. (2008). "The evolution of human communication and language," in *Sociobiology of Communication: an Interdisciplinary Perspective*, eds P. D'Ettorre and D. Hughes (New York, NY: Oxford University Press), 249–264.

Levine, D. S., and Perlovsky, L. I. (2008). Neuroscientific insights on biblical myths: simplifying heuristics versus careful thinking: scientific analysis of millennial spiritual issues. *Zygon J. Sci. Relig.* 43, 797–821. doi: 10.1111/j.1467-9744.2008.00961.x

Masataka, N., and Perlovsky, L. I. (2012). The efficacy of musical emotions provoked by Mozart's music for the reconciliation of cognitive dissonance. *Sci. Rep.* 2, 694. doi: 10.1038/srep00694

Perlovsky, L. I. (1998). Conundrum of combinatorial complexity. *IEEE Trans. PAMI* 20, 666–670. doi: 10.1109/34.683784

Perlovsky, L. I. (2001). *Neural Networks and Intellect: Using Model-Based Concepts.* New York, NY: Oxford University Press.

Perlovsky, L. I. (2006a). Toward physics of the mind: concepts, emotions, consciousness, and symbols. *Phys. Life Rev.* 3, 22–55. doi: 10.1016/j.plrev.2005.11.003

Perlovsky, L. I. (2006b). Fuzzy dynamic logic. *New Math. Nat. Comput.* 2, 43–55. doi: 10.1142/S1793005706000300

Perlovsky, L. I. (2007a). "Neural dynamic logic of consciousness: the knowledge instinct," in *Neurodynamics of Higher-Level Cognition and Consciousness,* eds L. I. Perlovsky and R. Kozma (Heidelberg: Springer Verlag), 73–108. doi: 10.1007/978-3-540-73267-9_5

Perlovsky, L. I. (2007b). Evolution of languages, consciousness, and cultures. *IEEE Comput. Intell. Mag.* 2, 25–39. doi: 10.1109/MCI.2007.385364

Perlovsky, L. I. (2009a). Language and cognition. *Neural Netw.* 22, 247–257. doi: 10.1016/j.neunet.2009.03.007

Perlovsky, L. I. (2009b). Language and emotions: emotional sapir-whorf hypothesis. *Neural Netw.* 22, 518–526. doi: 10.1016/j.neunet.2009.06.034

Perlovsky, L. I. (2010). Musical emotions: functions, origin, evolution. *Phys. Life Rev.* 7, 2–27. doi: 10.1016/j.plrev.2009.11.001

Perlovsky, L. I. (2012a). Cognitive function, origin, and evolution of musical emotions. *Music. Sci.* 16, 185–199; doi: 10.1177/1029864912448327

Perlovsky, L. I. (2012b). Emotionality of languages affects evolution of cultures. *Rev. Psychol. Front.* 1, 1–13.

Perlovsky, L. I. (2012c). Brain: conscious and unconscious mechanisms of cognition, emotions, and language. *Brain Sci.* 2, 790–834.

Perlovsky, L. I. (2012d). Cognitive function of music part, I. *Interdisc. Sci. Rev.* 7, 129–142.

Perlovsky, L. I. (2013a). A challenge to human evolution – cognitive dissonance. *Front. Psychol.* 4:179. doi: 10.3389/fpsyg.2013.00179

Perlovsky, L. I. (2013b). Cognitive function of music part II. *Interdisc. Sci. Rev.* 38, 149–173.

Perlovsky, L. I., Deming, R. W., and Ilin, R. (2011). *Emotional Cognitive Neural Algorithms with Engineering Applications. Dynamic Logic: from vague to crisp*. Heidelberg: Springer. doi: 10.1007/978-3-642-22830-8

Perlovsky, L. I., and Ilin, R. (2010). Neurally and mathematically motivated architecture for language and thought. *Open Neuroimag. J.* 4, 70–80.

Perlovsky, L. I., and Ilin, R. (2012). Mathematical model of grounded symbols: perceptual symbol system. *J. Behav. Brain Sci.* 2, 195–220. doi: 10.4236/jbbs.2012.22024

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 816–847. doi: 10.1016/j.neuroimage.2012.04.062

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124

Received: 26 August 2013; accepted: 02 September 2013; published online: 19 September 2013.

Citation: Perlovsky L (2013) Language and cognition—joint acquisition, dual hierarchy, and emotional prosody. Front. Behav. Neurosci. 7:123. doi: 10.3389/fnbeh.2013.00123

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

OPINION ARTICLE
published: 05 August 2013
doi: 10.3389/fnbeh.2013.00099

FOXP2 gene and language development: the molecular substrate of the gestural-origin theory of speech?

Carmelo M. Vicario *

School of Psychology, The University of Queensland, Brisbane, QLD, Australia
**Correspondence: uqcvicar@uq.edu.au*

Edited by:
Kuniyoshi L. Sakai, The University of Tokyo, Japan

The view that language evolved from a primarily gestural mode of communication has its roots on the 18th-century philosophers speculations (Vico, 1953/1744; de Condillac, 1971/1756).

Over time, these philosophical thoughts have gradually got consistence, thanks to the research in the field of Psychology and Neuroscience, which has provided exciting evidence in support of the so-called gestural origin-theory of language (Corballis, 2002). This theory recognizes to the gestures a precise role in language development. In particular, in terms of evolution it has been suggested that spoken language evolves from an ancient communication system using arm gestures. Accordingly, it has been suggested that gestures of the mouth might have been added to the manual system to form a combined manuofacial gestural system (Corballis, 2002; Gentilucci and Corbalis, 2006).

The literature on Mirror Neurons has provided a strong support to the gestural-origin theory thanks to the evidence of a close relationship between arm (Gallese et al., 1996; Rizzolatti et al., 1996) and mouth actions (Ferrari et al., 2003) in the brain of non-human primates. In particular, it has been shown that the area F5 of the monkey premotor cortex includes also a class of neurons that discharge when the animal grasps an object with the either the hand or the mouth (Rizzolatti et al., 1988). However, although a recent meta-analysis of 125 human fMRI studies (Molenberghs et al., 2012) identified a core network of human brain regions that possess mirror properties associated with action observation and execution, there is also a literature that challenges the existence of these neurons in humans. In fact, direct evidence for the existence of mirror neurons in humans is still lacking (Dinstein et al., 2008; Hickok, 2009). On the other hand, some studies provide data against it. For example, the study of Lingnau et al. (2009) failed in finding adaptation for motor acts that were first executed and then observed (as found for executed motor acts, when these were preceded by execution or observation of the same motor act) in the brain areas that are typically considered as endowed of mirror properties. This implies that the link between motor gestures and spoken language is not necessarily mediated by neurons provided of "mirror" properties.

Clues in support of the gestural-origins theory of the language are also provided by the research on humans. For example, grasping larger objects (Gentilucci et al., 2001) and bringing them to the mouth (Gentilucci et al., 2004) induces selective increases in parameters of lip kinematics and voice spectra of syllables pronounced simultaneously with action execution. Moreover, it has been reported that repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation (Gentilucci et al., 2006). This suggests that Broca's area is probably involved in the simultaneous control of gestures and word pronunciation.

Neuroimaging studies on humans provide a further support to this direct link between gestures and verbal language. Sakai et al. (2005) used functional Magnetic Resonance Imaging to examine hemispheric dominance during the processing of signed and spoken sentences. Their study was provided of two conditions: (i) the sign condition with sentence stimuli in Japanese Sign Language (JSL) in which were tested Deaf signers of JSL, hearing bilinguals (children of Deaf adults, CODA) of JSL and Japanese (JPN); (ii) the speech condition in which were tested hearing monolinguals (Mono) of JPN with auditory JPN stimuli alone (AUD), or with an audiovisual presentation of JPN and JSL stimuli (A and V). The authors found that the ventral part of the left inferior frontal gyrus (F3t/F3O) showed no main effects of modality condition, providing evidence in support of the existence on a common area for the processing of linguistic information from both signed and spoken sentences. Moreover, it has been recently documented the common involvement of the left area 7A in the superior parietal lobule while performing a sequenced button presses task or a sequence of different syllables repetition task (Heim et al., 2012). These data demonstrate the existence of a common cortical module in the area 7A while sequencing vocal gestures and hand motor actions.

Finally, a support to the gestural-origin theory of language originates from the study of human infants. For example, Fogel and Hannan (1985) provided evidence of gesture-vocalization synchrony in 2-and 3-months-old human infants. Word comprehension in children between 8 and 10 months and word productions between 11 and 13 months are typically accompanied by deictic gestures (Volterra et al., 1979; Bates and Snyder, 1987). Deictic gestures (referring to an object or location) are particularly important since they allow reference to grow from the immediate context toward abstraction by helping infants understand the link between symbols and referents (De Villiers Rader and Zukow-Goldring, 2010). Moreover, deictic gestures seem able to predict linguistic development in both typical and atypical human populations across many cultures (Iverson and Goldin-Meadow, 2005). All these studies suggest that gestures provide a foundation for each new stage in early linguistic development.

A recent discovery in the field of genetics seems providing new insights in support of the gestural-origin theory.

In particular, evidence suggests that the *FOXP2* gene, located on the human chromosome 7 (Fisher et al., 1998), could be the molecular substrate linking speech with gesture. In fact, this gene is involved not only in speech production and comprehension but also in gesture coordination.

In an early work Gopnik (1990) argued that the *FOXP2* gene is involved in the development of morphosyntax. For this motivation this gene has been identified more broadly as the "grammar gene" (Pinker, 1994). However, a subsequent investigation suggested that the core deficit associated with the abnormal expression of this gene is one of articulation, with grammatical impairment as secondary outcome (Watkins et al., 2002). Thus, it was proposed (Corballis, 2004) that this gene may play a role in the incorporation of vocal articulation, but have little to do with grammar itself. In support of this suggestion it has been reported that *FOXP2* shows overlapping expression patterns within brains of zebra finches and fetal human brains, particularly in subcortical regions that play important roles in sensorimotor integration and coordinated movements important for vocalization and speech (Teramitsu et al., 2004). Moreover, recent studies on humans extend the role of *FOXP2* gene to the coordination of upper limb movements. For example, the recent study of Peter et al. (2011) found an influence of the *FOXP2* gene on several language processing tasks such as nonword repetition, real word reading efficiency, rapid oral reading. Interestingly, they documented an effect of this gene also on rapid motor sequencing ability which also included finger movements.

Another recent work (Wilcke et al., 2012) has shown that the Single Nucleus Polymorphism (rs12533005) of the *FOXP2* gene can be associated with congenital dyslexia. Interestingly, the difficulty with sequential finger movements is another type of deficit which may characterize reading disorders (Tiffin-Richards et al., 2004). Furthermore, *FOXP2* mutations seem to account for the childhood apraxia of speech (CAS) (MacDermot et al., 2005; Laffin et al., 2012), which is characterized by problems in saying sounds, syllables, and words. Peter (2012) recently described the CAS *FOXP2* phenotype in multi-generational families as characterized not only by deficits in sequential processing at the level of alternating oral motor movements, which is consistent with the traditional CAS definition as a motor programing disorder, but also by deficits in sequential hand movements.

All these studies provide suggestive evidence that the *FOXP2* gene might be the possible molecular substrate linking gestures with verbal language. However, the research in support of this hypothesis is still limited, although there are promising fields of investigation. For example, it would be interesting to assess the impact of the *FOXP2* gene polymorphism on linguistic and manual skills in healthy adults. This investigation not only would provide a further support to the molecular substrate hypothesis for the gestural-origin theory of speech, but it could have also practical implications for developmental and educational psychology, as it might allow an early assessment of the risk for dyslexia and/or dysgraphia in childhood individuals.

Other potential issues worthy of investigation might refer to the study of the expression of *FOXP2* in individuals with special linguistic and/or manual skills (e.g., polyglot people, painters of talent); the influence played by particular socio-environmental factors on its expression, which in turn might influence linguistic and/or manual skills of healthy individuals.

Finally, it would be intriguing to valuate whether the *FOXP2* genetic variations influence the resilience of linguistic and/or manual functions in patients affected by stroke.

REFERENCES

Bates, E., and Snyder, L. S. (1987). "The cognitive hypothesis in language development," in *Infant Performance and Experience: New Findings with the Ordinal Scales,* eds E. Ina, C. Uzgiris, and E. J. McVicker Hunt (Urbana, IL: University of Illinois Press), 168–204.

Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.

Corballis, M. C. (2004). FOXP2 and the mirror system. *Trends Cogn. Sci.* 8, 95–96. doi: 10.1016/j.tics.2004.01.007

de Condillac, E. B. (1971/1756). *An Essay on the Origin of Human Knowledge: Being a Supplement to Mr. Locke's Essay on the Human Understanding (A facsimile reproduction of the 1756 translation by T. Nugent of Condillac's 1747 essay)*. Gainesville, FL: Scholars' Facsimiles and Reprints.

De Villiers Rader, N., and Zukow-Goldring, P. (2010). How the hands control attention during early word learning. *Gesture* 10, 202–221. doi: 10.1075/gest.10.2-3.05rad

Dinstein, I., Thomas, C., Behrmann, M., and Heeger, D. J. (2008). A mirror up to nature. *Curr. Biol.* 18, R13–R18. doi: 10.1016/j.cub.2008.01.044

Ferrari, P. F., Gallese, V., Rizzolatti, G., and Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur. J. Neurosci.* 17, 1703–1714. doi: 10.1046/j.1460-9568.2003.02601.x

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., and Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature Genet.* 18, 168–70. doi: 10.1038/ng0298-168

Fogel, A., and Hannan, T. E. (1985). Manual actions of nine- to fifteen-week-old human infants during face-to-face interaction with their mothers. *Child Dev.* 56, 1271–1279.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609. doi: 10.1093/brain/119.2.593

Gentilucci, M., Benuzzi, F., Gangitano, M., and Grimaldi, S. (2001). Grasp with hand and mouth: a kinematic study on healthy subjects. *J. Neurophysiol.* 86, 1685–1699.

Gentilucci, M., Bernardis, P., Crisi, G., and Dalla Volta, R. (2006). Repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation. *J. Cogn. Neurosci.* 18, 1059–1074. doi: 10.1162/jocn.2006.18.7.1059

Gentilucci, M., and Corbalis, M. C. (2006). From manual gesture to speech: a gradual transition. *Neurosci. Biobehav. Rev.* 30, 949–960. doi: 10.1016/j.neubiorev.2006.02.004

Gentilucci, M., Santunione, P., Roy, A. C., and Stefanini, S. (2004). Execution and observation of bringing a fruit to the mouth affect syllable pronunciation. *Eur. J. Neurosci.* 19, 190–202. doi: 10.1111/j.1460-9568.2004.03104.x

Gopnik, M. (1990). Feature-blind grammar and dysphasia. *Nature* 344, 715. doi: 10.1038/344715a0

Heim, S., Amunts, K., Hensel, T., Grande, M., Huber, W., Binkofski, F., et al. (2012). The role of human parietal area 7A as a link between sequencing in hand actions and in overt speech production. *Front. Psychol.* 3:534. doi: 10.3389/fpsyg.2012.00534

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J. Cogn. Neurosci.* 21, 1229–1243. doi: 10.1162/jocn.2009.21189

Iverson, J. M., and Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychol. Sci.* 16, 367–371. doi: 10.1111/j.0956-7976.2005.01542.x

Laffin, J. J., Raca, G., Jackson, C. A., Strand, E. A., Jakielski, K. J., and Shriberg, L. D. (2012). Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridization. *Genet. Med.* 14, 928–936. doi: 10.1038/gim.2012.72

Lingnau, A. Gesierich, B., and Caramazza, A. (2009). Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proc. Natl. Acad. Sci. U.S.A*. 106, 9925–9930. doi: 10.1073/pnas.0902262106

MacDermot, K. D., Bonora, E., Sykes, N., Coupe, A. M., Lai, C. S., Vernes, S. C., et al. (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am. J. Hum. Genet.* 76, 1074–1080. doi: 10.1086/430841

Molenberghs, P., Cunnington, R., and Mattingley, J. B. (2012). Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. *Neurosci. Biobehav. Rev*. 36, 341–349. doi: 10.1016/j.neubiorev.2011.07.004

Peter, B. (2012). Oral and hand movement speeds are associated with expressive language ability in children with speech sound disorder. *J. Psycholinguist Res.* 41, 455–474. doi: 10.1007/s10936-012-9199-1

Peter, B., Raskind, W. H., Matsushita, M., Lisowski, M., Vu, T., Berninger, V. W., et al. (2011). Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. *J. Neurodev. Disord.* 3, 39–49. doi: 10.1007/s11689-010-9065-0

Pinker, S. (1994). *The Language Instinct.* New York, NY: Morrow.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., and Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp. Brain. Res*. 71, 491–507. doi: 10.1007/BF00248742

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain. Res.* 3, 131–141. doi: 10.1016/0926-6410(95)00038-0

Sakai, K. L., Tatsuno, Y., Suzuki, K., Kimura, H., and Ichida, Y. (2005). Sign and speech: amodal commonality in left hemisphere dominance for comprehension of sentences. *Brain* 128, 1407–1417. doi: 10.1093/brain/awh465

Teramitsu, I., Kudo, L. C., London, S. E., Geschwind, D. H., and White, S. A. (2004). Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *J. Neurosci.* 24, 3152–3163. doi: 10.1523/JNEUROSCI.5589-03.2004

Tiffin-Richards, M. C., Hasselhorn, M., Richards, M. L., Banaschewski, T., and Rothenberger, A. (2004). Time reproduction in finger tapping tasks by children with attention-deficit hyperactivity disorder and/or dyslexia. *Dyslexia* 10, 299–315. doi: 10.1002/dys.281

Vico, G. B. (1953/1744). *La Scienza Nuova*. Bari: Laterza.

Volterra, V., Bates, E., Benigni, L., Bretherton, I., and Camaioni, L. (1979). "First words in language and action: a qualitative look," in *The Emergence of Symbols: Cognition and Communication in Infancy,* eds E. Bates, L. Benigni, I. Bretherton, L. Camaioni, and V. Volterra (New York, NY: Academic Press), 141–222.

Watkins, K. E., Dronkers, N. F., and Vargha-Khadem, F. (2002). Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. *Brain* 125, 452–464. doi: 10.1093/brain/awf058

Wilcke, A., Ligges, C., Burkhardt, J., Alexander, M., Wolf, C., Quente, et al. (2012). Imaging genetics of FOXP2 in dyslexia. *Eur. Hum. Genet.* 20, 224–229. doi: 10.1038/ejhg.2011.160

Received: 09 July 2013; accepted: 18 July 2013; published online: 05 August 2013.

Citation: Vicario CM (2013) FOXP2 gene and language development: the molecular substrate of the gestural-origin theory of speech? Front. Behav. Neurosci. 7:99. doi: 10.3389/fnbeh.2013.00099

frontiers in
BEHAVIORAL NEUROSCIENCE

OPINION ARTICLE
published: 17 September 2013
doi: 10.3389/fnbeh.2013.00122

What the online manipulation of linguistic activity can tell us about language and thought

Lynn K. Perry* and Gary Lupyan

Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA
**Correspondence: lkperry@wisc.edu*

Edited by:
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA

Keywords: verbal interference, transcranial direct current stimulation (tDCS), language and thought, linguistic relativity, labeling

Questions about the relationship between language and thought have long fascinated psychologists, philosophers, and the general public. One specific question is the extent to which verbal labels causally impact cognitive processes—how does calling an object by a particular name influence the way people categorize it; how does knowing words for mental states influence our reasoning about the minds of others; how does learning and using words like *left* influence our navigation behavior? One way to learn how the words we use to label objects, mental states, or locations affect our thoughts is to increase or decrease the ease with which we can use these words and observe outcomes of these manipulations on "non-linguistic" tasks. For example, if the word *left* enables us to remember which way to turn, preventing its activation might be expected to disrupt navigation. Manipulating the labeling process (and the engagement of language more broadly) is therefore very useful in exploring how language influences cognition. In this paper, we review two methodologies for implementing linguistic manipulations: verbal interference and transcranial direct current stimulation (tDCS), and discuss what we can learn about the role of language in cognitive processes from this line of research.

VERBAL INTERFERENCE

The primary method for manipulating linguistic activity is the use of verbal interference. The logic of verbal interference is well summarized by Winawer et al. (2007) in the context of studying the effects of language on color perception:

> "[I]f linguistic processes play an active, online role in perceptual tasks, then a verbal dual task, but not a non-linguistic dual task, should diminish the goluboy/siniy [light blue / dark blue] category advantage found in Russian speakers," (Winawer et al., 2007, p. 7781).

Winawer et al. reported that Russian speakers appear to perceive a larger perceptual distinction between light and dark blues, consistent with a lexical difference between these categories in Russian. One possibility is that these differences stem from long-term perceptual learning caused by years of distinguishing colors in one language (e.g., Özgen and Davies, 2002). An alternative is that the cross-linguistic perceptual differences arise from online top-down influences of language (e.g., Fonteneau and Davidoff, 2007; Lupyan, 2008). If true, then disrupting these top-down effects in some way may disrupt these online effects of language eliminating the cross-linguistic difference—the pattern observed in the study. It is important to note that for verbal interference to have some effect does not require the involvement of language to be strategic. Rather, the involvement could be one of "the spontaneous but unspoken use of lexical codes," (Gilbert et al., 2006, p. 489). On this account, previous associations of using the word *blue* to describe the color blue cause the word to become reactivated when the color is seen (Lupyan, 2012a,b). The automatic recruitment of color labels may temporarily warp the perceptual space, producing cross-linguistic differences of the sort observed by Winawer et al.

Using similar logic, verbal interference has been used to argue for a role of language in number concepts (Frank and Barner, 2012), spatial memory (Hermer-Vazquez et al., 1999), categorization (Lupyan, 2009), and theory of mind (Newton and de Villiers, 2007). A similar logic underlies behavioral *up*-regulation of linguistic processes through redundant presentation of labels (e.g., Lupyan et al., 2007) or overt self-directed speech (Lupyan and Swingley, 2012). If internally generated labels support some cognitive or perceptual process, then redundant externally-presented labels can be thought to *up-regulate* the linguistic contribution and verbal interference to *down-regulate* it.

Verbal interference has also been used to examine mechanisms of developmental change. For example, in a task requiring participants to locate an object hidden in the corner of a room, children rely on the room's shape, while adults use one of the walls—painted in a distinct color from the other walls—to find the object (Hermer and Spelke, 1994; cf. Twyman and Newcombe, 2010). While developments in spatial language correlate with developments in using the wall as a landmark, this cannot tell us whether language causally influences spatial memory. In order to determine whether language supports development of spatial memory, Hermer-Vazquez et al. (1999) used a verbal interference paradigm. When adults performed the task while shadowing speech, their performance mirrored that of children's. Thus, in addition to improving our understanding of how language affects cognition in real-time, verbal interference has improved our understanding of how influences of language on cognition develop.

However, the use of verbal interference is not without problems. First, despite being used for many years across many domains, there is at present no working theory of verbal interference. Put

bluntly: no one is sure how it works. Its use is vaguely based on the notion that language is a unitary system that can be disrupted, but this assumption is potentially problematic given the degree to which language-related activity is distributed both anatomically (Jung-Beeman, 2005) and functionally, e.g., syntax and semantics, rather than being entirely dissociable systems, tightly interact (MacDonald et al., 1994). Additionally, there is little agreement as to what constitutes verbal interference. Researchers have used numerous tasks, with little theoretical basis for choosing one over another. Tasks falling under the umbrella of verbal interference have included: rehearsing multi-digit numbers for later memory tests (e.g., Gilbert et al., 2006; Lupyan, 2009), repeating the letters "a,b,c" (e.g., Emerson and Miyake, 2003); alternating between naming months and days (e.g., Baddeley et al., 2001), making rhyme judgments (Roberson et al., 2007), answering factual questions such as "What is your name?" (e.g., Hatano et al., 1977), and repeating text, known as speech shadowing (Hermer-Vazquez et al., 1999; Frank and Barner, 2012). These interference tasks differ both in general difficulty and the ease with which performance can be assessed. For example, an interference task with a memory component gives an indication of how well participants rehearsed information—a proxy for how much effort was put into the verbal task—repetition of "a,b,c" does not. Such inconsistencies makes it hard (1) to infer why verbal interference sometimes interferes with primary task performance and sometimes does not and (2) to assess what aspect of language is recruited in the primary task which, when interfered with, disrupts performance.

A final problem with verbal interference is that it requires participants to perform two tasks simultaneously. It is therefore necessary to use a control interference task to determine which changes in primary task performance stem from manipulation to specifically linguistic processes and which stem from having to perform two tasks. Control tasks also vary widely across experiments: from tests of visuospatial memory (e.g., Gilbert et al., 2006; Lupyan, 2009) to foot tapping (Baddeley et al., 2001; Emerson and Miyake, 2003), and rhythm-shadowing (Hermer-Vazquez et al., 1999). Importantly, unless interference tasks are equated in all ways except for their "verbality," little can be said about the role of language in the primary task. For example, it has been argued that verbal, but not non-verbal interference disrupted performance on a false-belief task (Newton and de Villiers, 2007). However, when the interference tasks were better equated for difficulty, both were similarly disruptive (Dungan and Saxe, 2012). Additionally, because verbal interference uses a dual-task paradigm, participants can exert different amounts of effort into the tasks. Such differential effort is difficult both to measure and control.

Despite its appeal, verbal interference paradigms have clear shortcomings. Below, we outline an alternative way of perturbing linguistic processes that solve *some* of the shortcomings, and advocate for systematic cross-method comparisons of linguistic perturbation methods to more fully inform our understanding of how language augments cognition and perception.

TRANSCRANIAL DIRECT CURRENT STIMULATION

One way to avoid some challenges posed by verbal interferences is by manipulating language processing without using secondary tasks through the use of non-invasive brain stimulation. Here, we focus on one such method—tDCS a painless method of regulating cortical excitability through weak electrical current to the scalp—that allows the experimenter to subtly up- and down-regulate neural activity over cortical areas implicated in language processing. For example, using tDCS to up-regulate activity over Wernicke's area (associated with aspects of labeling, particularly comprehension of word meaning; e.g., Price, 2000) is associated with increased ability to map novel words to pictures (Flöel et al., 2008) using it to up-regulate activity over Broca's area (associated with linguistic processes such as speech production; e.g., Gernsbacher and Kaschak, 2003) is associated with increased artificial grammar learning (de Vries et al., 2010).

Similar to using verbal interference, using tDCS to study linguistic influences on cognition assumes that language is a system that can be selectively perturbed. However, tDCS avoids the need to use dual task paradigms. The participant simply performs the main task while undergoing tDCS which, depending on the electrode arrangement, either up- or down-regulates cortical activity in targeted regions. Up-regulating activity is theoretically analogous to behavioral up-regulation through presentation of overt labels (Lupyan, 2008) or behavioral self-directed speech (Lupyan and Swingley, 2012); down-regulating activity is theoretically analogous to verbal interference.

Recently, Lupyan et al. (2012) examined effects of tDCS on non-verbal categorization, showing that down-regulating activity over Broca's area was associated with impairments in the ability to form categories that required selectively representing specific perceptual features to the exclusion of others, e.g., GREEN THINGS, a deficit similar to one shown by individuals performing verbal interference (Lupyan, 2009) or by those with language impairments such as aphasia (Davidoff and Roberson, 2004; Lupyan and Mirman, 2013). In avoiding pitfalls of a dual task design, however, Lupyan and colleagues' tDCS study allows for more definitive conclusions about mechanisms by which language might affect categorization because participants all completed the same task, and between-group differences can be linked to changes in neural activity in a particular cortical region providing at least a foothold for starting to connect behavioral data on the role of language in categorization to particular neural mechanisms.

IMPORTANT FUTURE CONSIDERATIONS

Although tDCS may be well-suited for manipulating linguistic activity, at present there are no direct comparisons of tDCS to verbal interference. It would be useful to know whether domains in which we have seen effects of verbal interference on performance are affected by tDCS over areas associated with language processes and whether domains in which we have *not* previously seen effects are similarly *unaffected*. For example, tDCS can

be used to determine if language-related cortical regions shown to be recruited in non-verbal color-judgment tasks (e.g., Ting Siok et al., 2009) are *causally* implicated by showing perturbations with tDCS have behavioral consequences in color-judgment, thereby informing the neural basis of language-augmented perception. tDCS also has the potential to open up additional domains to investigate effects of language on cognition without the error-fraught hunt for perfect control interference tasks.

In sum, we advocate: (1) a more systematic comparison of linguistic perturbation methods, specifically comparing behavioral linguistic perturbations to perturbations utilizing stimulation techniques such as tDCS; (2) a more rigorous comparison of different verbal interference tasks; and (3) a theory that elucidates the mechanisms by which verbal interference actually works. Such an examination will be critical to clarifying the contributions of language to cognition, helping us answer such questions as whether the mechanisms by which language affects perception of color categories are in some broad sense similar to mechanisms by which language affects spatial cognition. Performing the cross-method comparisons we advocate will highlight possible contradictions leading to further theoretical refinement. For example, what would it mean for the underlying cognitive and neural processes if one verbal interference method affects a primary task but another does not?

The extant empirical literature on effects of language on cognition and perception takes us considerably beyond the question as it is often phrased: "Does language affect thought," (Boroditsky, 2010a). This literature (e.g., Gentner and Goldin-Meadow, 2003; Casasanto, 2008; Boroditsky, 2010b; Lupyan, 2012a,b), while rich in demonstrations, requires a more rigorous investigation of the mechanisms by which learning and using language augment and perhaps fundamentally alter cognition and perception. We believe significant clarity on this important question can be achieved by combining creative uses of linguistic perturbation techniques with theoretical refinement of their mechanisms.

ACKNOWLEDGMENTS

We would like to thank Pierce Edmiston for his comments on an earlier version of this paper.

REFERENCES

Baddeley, A., Chincotta, D., and Adlam, A. (2001). Working memory and the control of action: evidence from task switching. *J. Exp. Psychol. Gen.* 130, 641–657. doi: 10.1037/0096-3445.130.4.641

Boroditsky, L. (2010a). Do the languages we speak shape the way we think? in The Economist. Available online at: http://www.economist.com/debate/overview/190

Boroditsky, L. (2010b). "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), 615–632.

Casasanto, D. (2008). Who's afraid of the big bad whorf? Crosslinguistic differences in temporal language and thought. *Lang. Learn.* 58, 63–79. doi: 10.1111/j.1467-9922.2008.00462.x

Davidoff, J., and Roberson, D. (2004). Preserved thematic and impaired taxonomic categorisation: a case study. *Lang. Cogn. Processes* 19, 137–174. doi: 10.1080/01690960344000125

de Vries, M. H., Barth, A. C. R., Maiworm, S., Knecht, S., Zwitserlood, P., and Flöel, A. (2010). Electrical stimulation of Broca's area enhances implicit learning of an artificial grammar. *J. Cogn. Neurosci.* 22, 2427–2436. doi: 10.1162/jocn.2009.21385

Dungan, J., and Saxe, R. (2012). Matched false-belief performance during verbal and nonverbal interference. *Cogn. Sci.* 36, 1148–1156. doi: 10.1111/j.1551-6709.2012.01248.x

Emerson, M. J., and Miyake, A. (2003). The role of inner speech in task switching: a dual-task investigation. *J. Mem. Lang.* 48, 148–168. doi: 10.1016/S0749-596X(02)00511-9

Flöel, A., Rösser, N., Michka, O., Knecht, S., and Breitenstein, C. (2008). Noninvasive brain stimulation improves language learning. *J. Cogn. Neurosci.* 20, 1415–1422. doi: 10.1162/jocn.2008.20098

Fonteneau, E., and Davidoff, J. (2007). Neural correlates of colour categories. *Neuroreport* 18, 1323–1327. doi: 10.1097/WNR.0b013e3282c48c33

Frank, M. C., and Barner, D. (2012). Representing exact number visually using mental abacus. *J. Exp. Psychol. Gen.* 141, 134–149. doi: 10.1037/a0024427

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

Gernsbacher, M. A., and Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annu. Rev. Psychol.* 54, 91–114. doi: 10.1146/annurev.psych.54.101601.145128

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. doi: 10.1073/pnas.0509868103

Hatano, G., Miyake, Y., and Binks, M. G. (1977). Performance of expert abacus operators. *Cognition* 5, 47–55. doi: 10.1016/0010-0277(77)90016-6

Hermer, L., and Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature* 370, 57–59. doi: 10.1038/370057a0

Hermer-Vazquez, L., Spelke, E. S., and Katsnelson, A. S. (1999). Sources of flexibility in human cognition: dual-task studies of space and language. *Cogn. Psychol.* 39, 3–36. doi: 10.1006/cogp.1998.0713

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends Cogn. Sci.* 9, 512–518. doi: 10.1016/j.tics.2005.09.009

Lupyan, G. (2008). The conceptual grouping effect: categories matter (and named categories matter more). *Cognition* 108, 566–577. doi: 10.1016/j.cognition.2008.03.009

Lupyan, G. (2009). Extracommunicative functions of language: verbal interference causes selective categorization impairments. *Psychon. Bull. Rev.* 16, 711–718. doi: 10.3758/PBR.16.4.711

Lupyan, G. (2012a). "What do words do? Toward a theory of language-augmented thought," in *The Psychology of Learning and Motivation*, Vol. 57, ed B. H. Ross (San Diego, CA: Academic Press), 255–297. doi: 10.1016/B978-0-12-394293-7.00007-8

Lupyan, G. (2012b). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front. Cogn.* 3:54. doi: 10.3389/fpsyg.2012.00054

Lupyan, G., and Mirman, D. (2013). Linking language and categorization: evidence from aphasia. *Cortex* 49, 1187–1194. doi: 10.1016/j.cortex.2012.06.006

Lupyan, G., and Swingley, D. (2012). Self-directed speech affects visual search performance. *Q. J. Exp. Psychol.* 65, 1068–1085. doi: 10.1080/17470218.2011.647039

Lupyan, G., Mirman, D., Hamilton, R., and Thompson-Schill, S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition* 124, 36–49. doi: 10.1016/j.cognition.2012.04.002

Lupyan, G., Rakison, D. H., and McClelland, J. L. (2007). Language is not just for talking: redundant labels facilitate learning of novel categories. *Psychol. Sci.* 18, 1077–1083. doi: 10.1111/j.1467-9280.2007.02028.x

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* 101, 676–703. doi: 10.1037/0033-295X.101.4.676

Newton, A. M., and de Villiers, J. G. (2007). Thinking while talking: adults fail nonverbal false-belief reasoning. *Psychol. Sci.* 18, 574–579. doi: 10.1111/j.1467-9280.2007.01942.x

Özgen, E., and Davies, I. R. (2002). Acquisition of categorical color perception: a perceptual learning approach to the linguistic relativity hypothesis. *J. Exp. Psychol. Gen.* 131, 477–493. doi: 10.1037/0096-3445.131.4.477

Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *J. Anat.* 197, 335–359. doi: 10.1046/j.1469-7580.2000.19730335.x

Roberson, D., Damjanovic, L., and Pilling, M. (2007). Categorical perception of facial expressions:

evidence for a "category adjustment" model. *Mem. Cogn.* 35, 1814–1829. doi: 10.3758/BF03193512

Ting Siok, W., Kay, P., Wang, W. S. Y., Chan, A. H. D., Chen, L., Luke, K.-K., et al. (2009). Language regions of brain are operative in color perception. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8140–8145. doi: 10.1073/pnas.0903627106

Twyman, A. D., and Newcombe, N. S. (2010). Five reasons to doubt the existence of a geometric module. *Cogn. Sci.* 34, 1315–1356. doi: 10.1111/j.1551-6709.2009.01081.x

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. doi: 10.1073/pnas.0701644104

Received: 15 July 2013; accepted: 29 August 2013; published online: 17 September 2013.

Citation: Perry LK and Lupyan G (2013) What the online manipulation of linguistic activity can tell us about language and thought. Front. Behav. Neurosci. 7:122. doi: 10.3389/fnbeh.2013.00122

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

HYPOTHESIS AND THEORY ARTICLE
published: 18 December 2013
doi: 10.3389/fnbeh.2013.00204

Computational principles of syntax in the regions specialized for language: integrating theoretical linguistics and functional neuroimaging

***Shinri Ohta*[1,2], *Naoki Fukui*[3,4] *and Kuniyoshi L. Sakai*[1,4,5]*

[1] *Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan*
[2] *Japan Society for the Promotion of Science, Tokyo, Japan*
[3] *Department of Linguistics, Sophia University, Tokyo, Japan*
[4] *CREST, Japan Science and Technology Agency, Tokyo, Japan*
[5] *Department of Basic Science, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan*

Edited by:
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA

Reviewed by:
Cedric A. Boeckx, Catalan Institute for Research and Advanced Studies, Spain
Noriaki Yusa, Miyagi Gakuin Women's University, Japan

***Correspondence:**
Kuniyoshi L. Sakai, Department of Basic Science, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
e-mail: sakai@mind.c.u-tokyo.ac.jp

The nature of computational principles of syntax remains to be elucidated. One promising approach to this problem would be to construct formal and abstract linguistic models that parametrically predict the activation modulations in the regions specialized for linguistic processes. In this article, we review recent advances in theoretical linguistics and functional neuroimaging in the following respects. First, we introduce the two fundamental linguistic operations: Merge (which combines two words or phrases to form a larger structure) and Search (which searches and establishes a syntactic relation of two words or phrases). We also illustrate certain universal properties of human language, and present hypotheses regarding how sentence structures are processed in the brain. Hypothesis I is that the Degree of Merger (DoM), i.e., the maximum depth of merged subtrees within a given domain, is a key computational concept to properly measure the complexity of tree structures. Hypothesis II is that the basic frame of the syntactic structure of a given linguistic expression is determined essentially by functional elements, which trigger Merge and Search. We then present our recent functional magnetic resonance imaging experiment, demonstrating that the DoM is indeed a key syntactic factor that accounts for syntax-selective activations in the left inferior frontal gyrus and supramarginal gyrus. Hypothesis III is that the DoM domain changes dynamically in accordance with iterative Merge applications, the Search distances, and/or task requirements. We confirm that the DoM accounts for activations in various sentence types. Hypothesis III successfully explains activation differences between object- and subject-relative clauses, as well as activations during explicit syntactic judgment tasks. A future research on the computational principles of syntax will further deepen our understanding of uniquely human mental faculties.

Keywords: syntax, universal grammar, recursive computation, inferior frontal gyrus, supramarginal gyrus, fMRI

INTRODUCTION

Tree structures are one of the most ubiquitous structures in nature, appearing in the branchings of rivers, lightning, snowflakes, trees, blood vessels, nervous systems, etc., and can be simulated in part by fractal geometry (Mandelbrot, 1977). To properly quantify the complexity of such tree structures, various models have been proposed. The number of nodes would be one of the simplest models; this approach consists of simply counting the total number of non-terminal nodes (branching points) and terminal nodes of a tree structure (**Figure 1A**). This model obviously cannot capture hierarchical levels within the tree (sister relations in linguistic terms). To properly measure the hierarchical levels of a tree structure, we have proposed the Degree of Merger (DoM) as a key computational concept (**Figure 1B**) (Ohta et al., 2013). The DoM is defined as the *maximum depth* of merged subtrees (called Mergers) within a given domain. With this model, the same numbers are assigned to the nodes with an identical hierarchical level. The DoM corresponds to the number of iterations for generating fractal figures, when the tree structures are self-similar.

In this article, we first explain certain universal properties of human language discovered in modern linguistics, and we present hypotheses regarding how sentence structures are processed in the brain. We then introduce our recent functional magnetic resonance imaging (fMRI) study, which demonstrated that the DoM is indeed a key syntactic factor that accounts for syntax-selective activations in the regions specialized for language (Ohta et al., 2013). We also show that the top-down connectivity from the left inferior frontal gyrus to the left supramarginal gyrus is critical for the syntactic processing. Next, we clarify that the DoM can account for activation modulations in the frontal region, depending on different sentence structures. Finally, we hypothesize that the DoM domain changes dynamically in accordance with iterative Merge applications, the distance

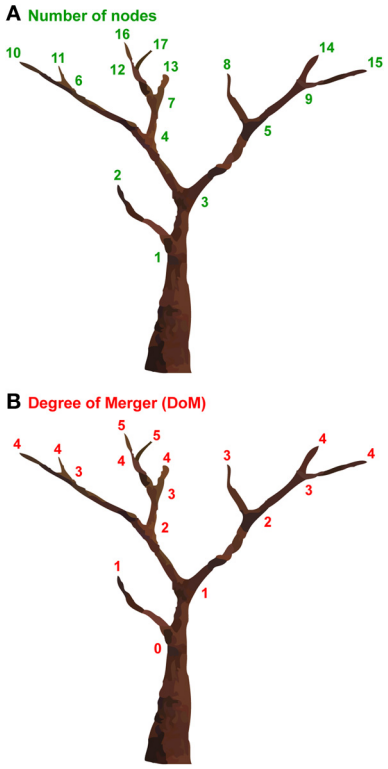


FIGURE 1 | Two models for measuring the complexity of tree structures. (A) "The number of nodes" counts the total number of nonterminal nodes (branching points) and terminal nodes of a tree structure. The number of nodes of the tree structure shown is 17. **(B)** "The Degree of Merger (DoM)" quantifies the maximum depth of merged subtrees, or the degree of branching. We increased the number one by one for each node, starting from the trunk (zero) to terminal nodes. The DoM of the tree structure shown is 5.

required for Search operations (or simply the "Search distance"), and/or task requirements. This hypothesis accounts for activation differences between subject-relative and object-relative clauses, as well as for activations during explicit syntactic judgment tasks.

UNIVERSAL PROPERTIES OF HUMAN LANGUAGE

THEORETICAL BACKGROUND

Modern linguistics has clarified universal properties of human language, which, directly or indirectly, reflect the computational power, or engine, of the human language faculty. A sentence is not a mere string of words, but is made of phrase structure (called constituent structure). Moreover, a single phrase contains the key element (i.e., the "head") that determines the basic properties of the phrase. Furthermore, a sentence can be recursively embedded within other sentences, as in, e.g., "*I think that John believes that Mary assumes that…* ," and there is in principle no upper bound for the length of sentences. These universal properties can be adequately and minimally expressed by hierarchical tree structures with a set of relevant structural relations defined on such structures (Chomsky, 1957, 1965).

To construct hierarchical tree structures, modern linguistics has proposed the fundamental linguistic operation of *Merge* (capitalized in linguistics to indicate a formal operation). Merge is a structure-building operation that combines two syntactic objects (words or phrases) to form a larger structure (Chomsky, 1995). Merge would be theoretically "costless," requiring no driving force for its application (Saito and Fukui, 1998; Chomsky, 2004; Fukui, 2011). Besides Merge, we have proposed *Search* operation of searching syntactic features, which applies to a syntactic object already constructed by Merge, where Search couples and connects two distinct parts of the same structure, thereby assigning relevant features from one to the other part (Fukui and Sakai, 2003). Various other "miscellaneous" operations that have been employed in the linguistics literature, such as Agree, Scope determination, Copy, etc., are in fact different manifestations of one and the same, i.e., more generalized, operation of Search (Fukui and Sakai, 2003). Human language, therefore, should minimally contain two universal operations, Merge and Search. The total number of Merge and Search applications within an entire sentence are here simply denoted as "number of Merge" and "number of Search," respectively. The number of Merge in a sentence becomes always one less than the number of terminal nodes, *irrespective of sentence structures* (see Appendix S2 of Ohta et al., 2013).

SYMBOL SEQUENCES AND FORMAL LANGUAGES

In regard to formal symbol sequences beyond the bounds of finite state languages, three specific types of language have been discussed in the linguistics literature: (i) "counter language," (ii) "mirror-image language," and (iii) "copying language" (cf. Chomsky, 1957, p. 21).

(i) *ab*, *aabb*, *aaabbb*, …, and in general, all sentences consisting of *n* occurrences of *a* followed by *n* occurrences of *b* and only these;
(ii) *aa*, *bb*, *abba*, *baab*, *aaaa*, *bbbb*, *aabbaa*, *abbbba*, …, and in general, all sentences consisting of a string *X* followed by the "mirror image" of *X* (i.e., *X* in reverse), and only these;
(iii) *aa*, *bb*, *abab*, *baba*, *aaaa*, *bbbb*, *aabaab*, *abbabb*, …, and in general, all sentences consisting of a string *X* of *a*'s and *b*'s followed by the identical string *X*, and only these.

The counter language can be handled by a counting mechanism to match the number of each symbol, whereas the mirror-image language contains a mirror-image dependency, requiring more than a mere counter. If the number of symbols is not fixed (i.e., infinite), both of these languages are beyond the bounds of finite-state grammars, and are to be generated by context-free (simple) phrase structure grammars, while the copying language with a cross-serial dependency clearly goes beyond the bounds of even context-free phrase structure grammars, requiring a more powerful device, viz., context-sensitive phrase structure grammars or transformational grammars (Chomsky, 1959; Hopcroft and Ullman, 1979).

It remains a central issue in cognitive sciences whether or not the faculty of language is also shared by animals. Animals have thus been tested with regular symbol sequences such as

A^nB^n ($n \geq 2$; i.e., AABB, AAABBB, ...) and $(AB)^n$ ($n \geq 2$; i.e., ABAB, ABABAB, ...), which differ in *symbol order*. In an animal study, songbirds were trained to discriminate patterns of A^nB^n and $(AB)^n$ in more than ten thousand trials (Gentner et al., 2006). However, this learning can be achieved by tracking symbol repetition or counting strategy alone (Corballis, 2007). There is also a recent report that songbirds seemed to discriminate strings with or without nesting (Abe and Watanabe, 2011), but this learning can be achieved by simply remembering partial strings (Beckers et al., 2012). Along the lines of contrasting A^nB^n and $(AB)^n$, fMRI studies have tested participants with different symbol sequences, such as $A_2A_1B_1B_2$ vs. $A_1B_1A_2B_2$ (each subscript denotes a matching order), which also differ in matching order (Bahlmann et al., 2008). The difference in activation patterns can be simply explained by differences in any factor associated with matching orders and symbol orders, i.e., temporal order-related factors. It is thus necessary to completely control these general factors when extracting any syntactic factor from a number of cognitive factors involved in actual symbol processing.

Since the number of symbols is inevitably fixed (i.e., finite) in any actual experiment, it should be noted that any symbol sequence can be expressed by a regular (finite state) grammar, i.e., the least powerful grammar in the so-called Chomsky hierarchy. Therefore, one cannot, in principle, claim from the experiments that individual grammars (e.g., context-free phrase structure grammars vs. regular grammars) are differentially represented in the brain. Thus, the neural representation of individual grammars was *not* within the scope of Ohta et al. (2013). In addition to the various models examined, other non-structural and non-symbolic models with simple recurrent networks have been proposed to process some examples of even context-free and context-sensitive phrase structure languages, generalizing to some degree to longer strings than the training set (Rodriguez, 2001). However, these models do not account for any parametric modulation of the activations reported in Ohta et al. (2013), except the length of sentences.

In the previous experiment, we introduced letter strings, which had no lexical associations but had both symbol orders (e.g., AABB and ABAB) and matching orders (e.g., $A_2A_1B_1B_2$). There were two basic types of strings: reverse-order strings (Reverse) and same-order strings (Same). In the Reverse strings, the first and second halves of a string were presented in the reverse order, while in the Same strings the halves were presented in the same order (**Figure 2**). Under these conditions, there was actually no path connecting the non-terminal nodes of symbol pairs (e.g., A_1B_1 and A_2B_2), as there was *no* Merge application to connect the multiple pairs. In regard to the symbol orders, both the Reverse and Same strings took the above type (i) of A^nB^n. In regard to the matching orders, the Reverse string took the type (ii) of $A_2A_1B_1B_2$ or $A_3A_2A_1B_1B_2B_3$, while the Same string took the type (iii) of $A_1A_2B_1B_2$ or $A_1A_2A_3B_1B_2B_3$.

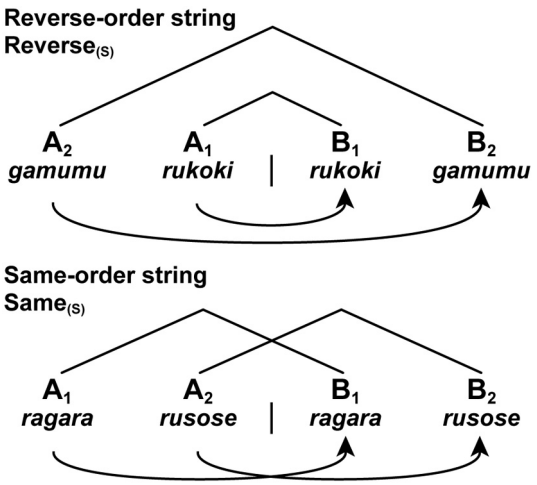


FIGURE 2 | Two basic types of letter strings related to formal languages. We tested two string conditions with short [(S) as a subscript] stimuli: $Reverse_{(S)}$ and $Same_{(S)}$. Each letter string was formed by jumbling letters of either a pseudonoun or pseudoverb (see **Figure 4**). We also tested the long stimuli with six items. Each curved arrow with an arrowhead denotes a Search operation, as in the following figures. Symbols used: A, sample stimulus; B, comparison stimulus.

HYPOTHESIS I

Given a tree structure with a formal property of Merge and *iterativity* (recursiveness) (Fukui, 2011), we propose the following hypothesis (Hypothesis I):

(1) The DoM, which can be defined as the *maximum depth* of merged subtrees within a given domain, is a key computational concept to properly measure the complexity of tree structures.

The DoM can quantify and compare various syntactic phenomena, such as self-embedding, scrambling, *wh*-movement, etc. Furthermore, when Search applies to each syntactic object with its hierarchical structure, the calculation of the DoM plays a critical role. Indeed, from a nested sentence "$[[The\ boy_2\ [we_3\ like_3]_2]_1\ sings_1]_0$" (subscripts denote the DoM for each node), two sentences "$[The\ boy...]_1\ sings_1$" and "$we_3\ like_3$" are obtained, where relevant features (numbers and persons here) are searched and matched between the nodes with the identical DoM. Since such analyses of hierarchical structures would produce specific loads in syntactic computation, we expect that the DoM and associated "number of Search" would affect performances and cortical activations.

Sentences with various constructions have been previously discussed in terms of the acceptability of sentences (cf. Chomsky, 1965, p. 12).

(i) nested constructions
(ii) self-embedded constructions
(iii) multiple-branching constructions
(iv) left-branching constructions
(v) right-branching constructions

The nested constructions are created by *centrally* embedding a phrase within another phrase (with some non-null element to its left and some non-null element to its right), and the self-embedded constructions are the special case of nested constructions when nesting occurs within the *same* type of

phrases (e.g., noun phrases). The multiple-branching constructions are made by conjoining phrases at the same hierarchical level, and the left/right-branching constructions are yielded by merging a phrase in the left-most or right-most phrase. The degrees of nesting and self-embedding have already been proposed to model the understanding of sentences (Miller and Chomsky, 1963). By generalizing this attractive idea in such a way as to include any construction with merged phrases, we introduced the DoM as a key computational concept.

Based on the nested (self-embedded), left/right-branching, and multiple-branching constructions, three basic types of sentences can be distinguished: the nested sentence (Nested), simple sentence (Simple), and conjoined sentence (Conjoined), respectively. The sentences shown in **Figure 3** are some examples in Japanese. Given syntactic structures like the ones shown, the correspondence of each subject-verb pair becomes fixed. Here N and V denote a noun phrase and a verb phrase, respectively. For the sentence shown in **Figure 3A**, an entire sentence is constructed by nesting sentences in the form of $[N_2[N_1V_1]V_2]$, where $[N_iV_i]$ represents a subject-verb pair of a sentence. Since Japanese is a head-last, and hence an SOV (verb-final) language, a main verb is placed after a subordinate clause. Therefore, Japanese sentences naturally yield nested structures without having to employ, as in English, object-relative clauses (e.g., "*The boy who$_i$ we like t_i sings*"), which require "movement" of an object (i.e., with more Merge applications) and thus leave behind a "trace" (t_i, subscripts denote the same entity). For the sentence shown in **Figure 3B**, a simple sentence is constructed by adding the same number of left/right branches to both Ns and Vs. The last noun (i.e., head) in the branches of Ns made a subject-verb pair with the last verb (i.e., head) of a compound verb. Each simple sentence thus takes the form of $[(NN_1)(VV_1)]$. For the sentence shown in **Figure 3C**, an entire sentence is constructed by conjoining sentences in the form of $[N_1V_1][N_2V_2]$. When considering longer sentences like $N_3N_2N_1V_1V_2V_3$, these constructions have distinct values for DoM.

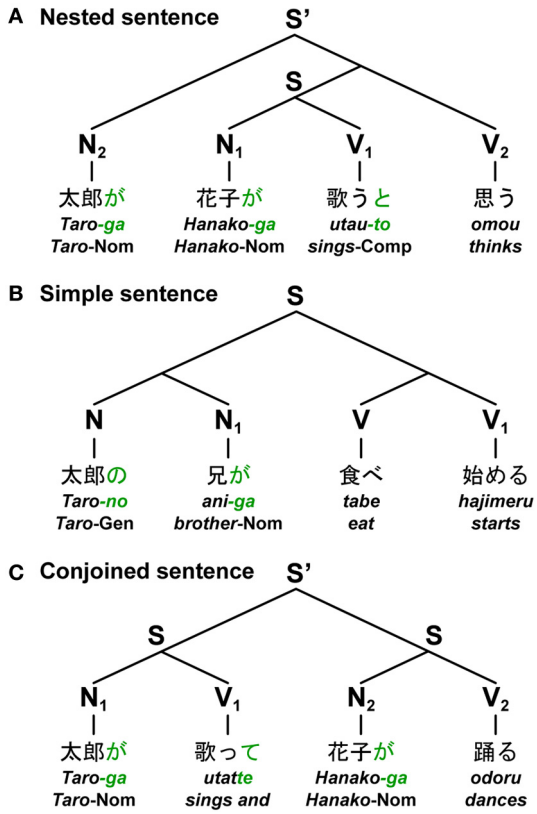


FIGURE 3 | Japanese sentences with three major constructions. The figure shows three basic types of sentences in Japanese: the nested sentence, simple sentence, and conjoined sentence. Based on contemporary linguistics, each diagram represents a unique tree structure of each sentence constructed from nouns and verbs. Below each example, word-by-word translations in English are shown. **(A)** A sentence (S) at the lowest hierarchical level was nested into an entire sentence (S') ("*Taro-ga Hanako-ga utau-to omou*," "*Taro thinks that Hanako sings*"). **(B)** A simple sentence was constructed by adding the same number of left/right branches to both nouns and verbs ("*Taro-no ani-ga tabe hajimeru*," "*Taro's brother starts eating*"). **(C)** An entire sentence (S') was constructed by conjoining two sentences ("*Taro-ga utatte Hanako-ga odoru*," "*Taro sings, and Hanako dances*"). Symbols used: S and S', sentence; N, noun phrase; V, verb phrase; *-ga*, nominative case marker; *-no*, genitive case marker; *-to*, complementizer; *-te*, gerundive form; Nom, nominative case; Gen, genitive case; Comp, complementizer.

HYPOTHESIS II

In any sentence, functional elements, such as inflections, auxiliary verbs, and grammatical particles, serve an essentially grammatical function without descriptive content. In regard to the fundamental role of these functional elements, we propose the following hypothesis (Hypothesis II):

(2) The basic frame of the syntactic structure of a given linguistic expression (e.g., sentence) is determined essentially by functional elements, which trigger Merge and Search operations.

In the non-sense poem "Jabberwocky" by Lewis Carroll, e.g., "'*Twas* (*'It was'*) *brillig, and the slithy toves did* ... ," the basic frames of syntactic structures are indeed determined by the functional elements of "*'Twas*," "*and*," "*the*," "*-s*," and "*did*." In the Japanese language, grammatical particles and morphosyntactic inflections are functional elements. The sentences shown in **Figure 3** actually contain only three kinds of grammatical particles, which represent *canonical* (i.e., in a prototypical use) case markings and syntactic information in Japanese: *-ga*, a nominative case marker; *-no*, a genitive case marker; and *-to*, a complementizer. It should be noted that both the nested and simple sentences have the same symbol order of N^nV^n, but they have different grammatical particles and syntactic structures. In contrast, both the simple and conjoined sentences have the same tree structures as a result, but they have different symbol orders of N^nV^n or $(NV)^n$ $(n \geq 2)$. It is the grammatical particles and morphosyntactic inflections, but not symbol orders or matching orders themselves, that determine the basic frame of syntactic structures of a sentence.

Following morphosyntactic and phonological features of Japanese verbs (Tsujimura, 2007), Vs take a non-past-tense form (*-ru*), past-tense form (*-ta*), or gerundive form (*-te*); Vs ending with *-to* and *-te* introduce *that*-clauses and *and*-conjunctives, respectively. The gerundive form can be used not only in *and*-conjunctives, but in compound verbs (e.g., "*tabete-simau*," "*finish*

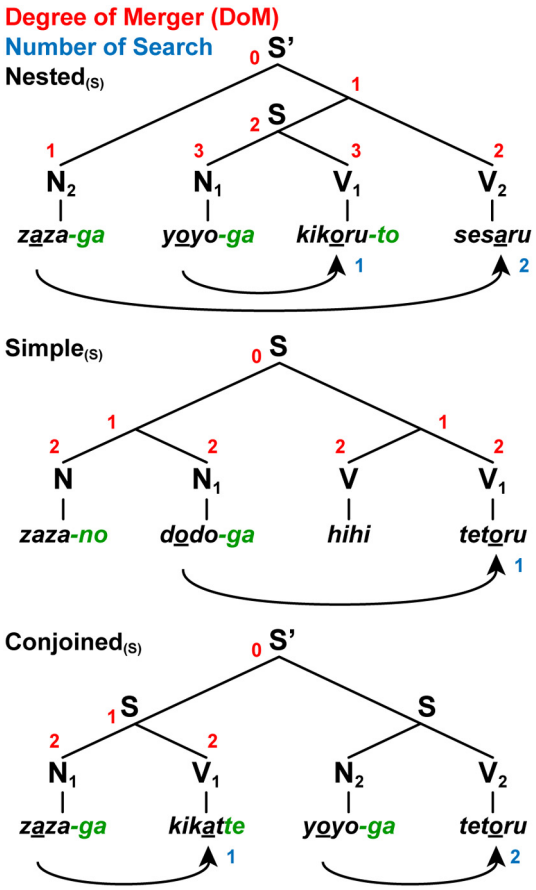


FIGURE 4 | A paradigm for testing various sentence structures. We tested three sentence conditions with short [(S) as a subscript] jabberwocky sentences: Nested$_{(S)}$, Simple$_{(S)}$, and Conjoined$_{(S)}$. Note the syntactic structures of these jabberwocky sentences are same as those of real sentences in **Figure 3**. The digits shown in red and blue denote the DoM for each node and "number of Search," respectively. We also tested the long stimuli with six words.

eating"; actual Japanese words will be translated hereafter), much as gerunds can in English. The *-ga*, *-no*, *-to*, and *-te* endings (*green* letters in **Figures 3**, **4**), together with the first verb of a compound verb in an adverbial form (e.g., "*tabe*"), are associated with Merge applications to connect multiple nouns/verbs or sentences, amounting to "number of Merge." The Japanese language lacks the "agreement features" (i.e., number, person, gender, etc.), but it is nevertheless equipped with the general Search procedure that is employed in agreement phenomena in other languages. This Search mechanism is in fact attested for various other phenomena in Japanese (see Fukui and Sakai, 2003 for further discussion). For example, the Japanese language exhibits a phenomenon called "honorification," where a noun phrase denoting an honored person and the form of honorifics in verbs are to be matched (Gunji, 1987; Ivana and Sakai, 2007).

In this section, we provided some theoretical discussions based on modern linguistics, focusing on the two fundamental linguistic operations of Merge and Search. We hypothesized that the DoM is a key computational concept to properly quantify the complexity of tree structures, and that the basic frame of the syntactic structure of a given linguistic expression is determined essentially by grammatical particles and morphosyntactic inflections, which trigger Merge and Search operations.

THE DoM AS A KEY SYNTACTIC FACTOR ELUCIDATED BY AN fMRI EXPERIMENT

One possible way to elucidate the neural basis of computational properties of natural language is to examine how the brain responds to the modulation of specified syntactic factors. We should not be content with such a general cognitive factor as so-called "syntactic complexity" or "syntactic working memory," which could involve both linguistic and non-linguistic factors. We should instead identify minimal factors that *sufficiently* explain any activation change obtained. In our recent study, we focused on different sentence constructions, and found that the DoM and "number of Search" were the *minimal* syntactic factors associated with phrase structures, which parametrically modulate cortical responses measured with event-related fMRI (Ohta et al., 2013). In this section, we will present the basic paradigm and results of this work.

A PARADIGM TO TEST HYPOTHESES I AND II

We used jabberwocky sentences, which consist of pseudonoun phrases (Ns) and pseudoverb phrases (Vs) that lack lexical associations, but have grammatical particles and morphosyntactic inflections (**Figure 4**). According to Hypothesis II stated above, these jabberwocky sentences had the same syntactic structures as normal sentences. Under the sentence conditions of Nested, Simple, and Conjoined with the same structures shown in **Figure 3**, the jabberwocky sentences were visually presented in a phrase-by-phrase manner to the participants. We made six pseudonouns by repeating the same syllables with voiced consonants and any one of /a/, /u/, or /o/: *rara*, *zaza*, *mumu*, *gugu*, *yoyo*, and *dodo*. We also made four pseudoverb roots by repeating the same syllables with voiceless consonants and either /i/ or /e/: *kiki*, *hihi*, *sese*, and *tete*. Here, vowel harmony was adopted to change the last, i.e., the second, vowel of the verb root, so that this vowel harmonized with the vowel (i.e., /a/, /u/, or /o/) of the corresponding subject (e.g., "*rara-ga tetaru*" from "*teteru*," underlined vowels within pseudowords). These features of vowels were only *experimentally* introduced, and these pseudoverbs lacked grammatical features, as in the Japanese verbs. In all jabberwocky sentences, the distinction between Ns and Vs was clear without memorizing pseudowords, because Ns, but not Vs, ended with either *-ga* or *-no*, i.e., case markers in Japanese such as *-ga* and *-no* can be generally attached only to nominal phrases.

To test whether participants actually paid attention to the correspondence of each subject-verb pair, we used a matching task, such that the vowel of a subject (N_i as a sample stimulus) was matched with the last vowel of the corresponding verb root (V_i as a comparison stimulus), probing the goal with the same vowel as explained above. It follows that the same syntactic structures were constructed from matching stimuli and non-matching stimuli (e.g., "*rara-ga teturu*"), which were both well-formed, i.e., *grammatical*, in Japanese. A matching strategy (counting, for example,

the first and the fourth stimuli for matching) was useful in solving the task, but performing the task was *not* prerequisite for constructing syntactic structures. Our matching task is different from classification tasks for symbol orders (e.g., AABB vs. ABAB, where A and B are symbols representing certain sets of stimuli), which can be solved by counting the maximum number of consecutively repeated symbols. The order of the Nested, Simple, and Conjoined was pseudo-randomized without repetition. We further examined whether cortical activations were modulated by the length of sentences: short (S as a subscript, e.g., Conjoined$_{(S)}$; four-word) and long (L as a subscript, e.g., Conjoined$_{(L)}$; six-word) sentences, where the DoM domain spanned four and six relevant words, respectively.

We also used the same matching task under the string conditions of Reverse and Same (**Figure 2**), such that the first half of a string (A_i as a sample stimulus) was matched with the corresponding second half (B_i as a comparison stimulus) in the reverse or same order. These string conditions also controlled any involvement of the matching strategy stated above. Between the Nested ($N_2N_1V_1V_2$) and Reverse ($A_2A_1B_1B_2$) conditions, the curved arrows shown in **Figures 2**, **4** represent the *same* matching order of sequentially presented stimuli. The symbol order was also identical among the Nested, Simple, Reverse, and Same conditions in the form of N^nV^n or A^nB^n. Combining these multiple conditions, we were able to properly examine whether different structures were actually constructed between sentences and strings. The spatial and temporal resolution of fMRI, as well as its sensitivity, has been proven to be high enough to confirm various hypotheses about human cognitive functions like ours.

SYNTAX-SELECTIVE ACTIVATIONS MODULATED BY THE DoM AND THE NUMBER OF SEARCH

To control both matching orders and symbol orders, we directly compared the Nested with the Reverse condition, using the Simple and Same conditions as respective references, i.e., (Nested − Simple) > (Reverse − Same), where we combined the short and long stimuli. This contrast further controlled various linguistic and non-linguistic factors, such as the number of Merge, number of case markers, number of nodes, memory span, and counting. This point is particularly important, because temporal order-related or memory-related factors have often been confused with differences in structure or grammar type. Significant activation was elicited by this contrast in the pars opercularis and pars triangularis of the left inferior frontal gyrus (L. F3op/F3t) [local maximum: $(x, y, z) = (-51, 24, 24)$, $Z = 5.8$], and the left supramarginal gyrus (L. SMG) [$(-39, -45, 42)$, $Z = 5.7$] (**Figure 5A**). Our results are best explained by the linguistic factors associated with the Nested condition, supporting our second hypothesis that basic syntactic structures are constructed when well-formed sentences are given even without lexical meanings.

For these two critical regions, we examined the percent signal changes under the Nested and Simple conditions by subtracting those under the Conjoined condition, which had the simplest tree structures (**Figure 4** and **Table 1**), separately for

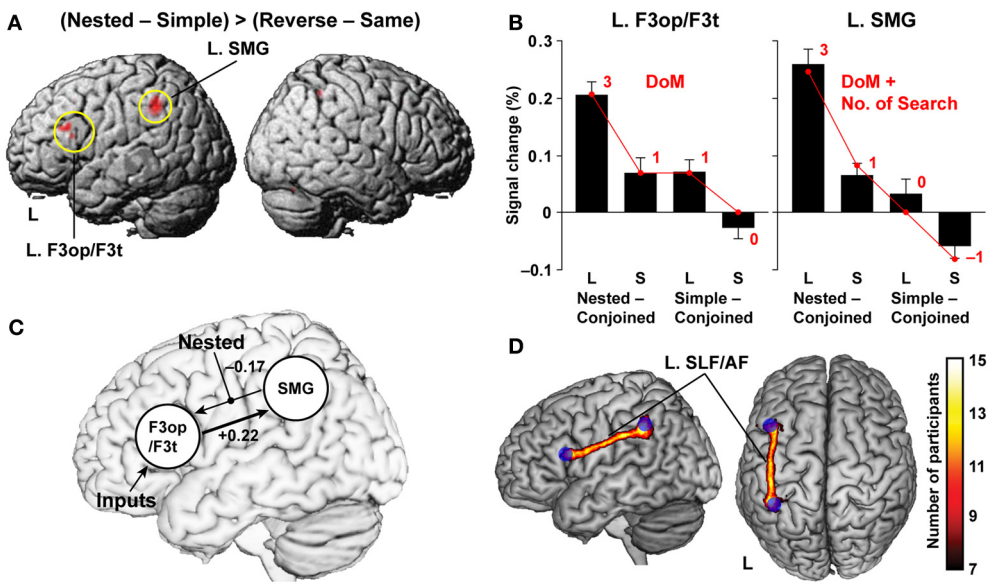


FIGURE 5 | Functional and anatomical evidence of a syntax-related network. (A) Regions identified by the (Nested − Simple) > (Reverse − Same) contrast (see **Figure 4**). Activations were projected onto the left (L) and right lateral surfaces of a standard brain. **(B)** Percent signal changes for Nested − Conjoined and Simple − Conjoined in the L. F3op/F3t and L. SMG. Overlaid red dots and lines denote the values fitted with the estimates (digits in red) for the best models: DoM for the L. F3op/F3t and "DoM + number of Search" for the L. SMG. **(C)** The results of DCM, testing effective connectivity between the L. F3op/F3t and L. SMG. The best model included a significant top-down connection from the L. F3op/F3t to L. SMG (a thick line). **(D)** Anatomical connectivity between the L. F3op/F3t and L. SMG revealed by DTI. The population probability map is shown on the left lateral and dorsal surfaces of a standard brain with maximum intensity projection. Blue spheres represent seed regions of the L. F3op/F3t and L. SMG. Symbols used: L, long sentences; S, short sentences.

Table 1 | Estimates of various factors to account for activations in Ohta et al. (2013).

Factor	Nested$_{(L)}$	Nested$_{(S)}$	Simple$_{(L)}$	Simple$_{(S)}$	Conjoined$_{(L)}$	Conjoined$_{(S)}$
Degree of Merger (DoM)	5	3	3	2	2	2
No. of Search	3	2	2	1	3	2
No. of nodes	11	7	11	7	10	7
	Nested$_{(L)}$ − Conjoined$_{(L)}$	**Nested$_{(S)}$ − Conjoined$_{(S)}$**	**Simple$_{(L)}$ − Conjoined$_{(L)}$**	**Simple$_{(S)}$ − Conjoined$_{(S)}$**		
DoM	3	1	1	0		
DoM + No. of Search	3	1	0	−1		
No. of Search	0	0	−1	−1		
No. of nodes	1	0	1	0		

Estimates under the Conjoined condition were subtracted from those under the other Nested and Simple conditions [e.g., DoM for Nested$_{(L)}$ − Conjoined$_{(L)}$, 5 − 2 = 3], separately for long and short sentences. We regarded "DoM + number of Search" (i.e., adding the estimates of two factors) as an additional factor.

long and short sentences. Since we used the Conjoined$_{(L)}$ and Conjoined$_{(S)}$ as appropriate references, we examined whether likewise *subtracted* estimates of each factor (e.g., DoM for Nested$_{(L)}$ − Conjoined$_{(L)}$; see **Table 1**) directly explained the parametric modulation of activations in the four contrasts of Nested$_{(L)}$ − Conjoined$_{(L)}$, Nested$_{(S)}$ − Conjoined$_{(S)}$, Simple$_{(L)}$ − Conjoined$_{(L)}$, and Simple$_{(S)}$ − Conjoined$_{(S)}$. The percent signal changes in the L. F3op/F3t and L. SMG, averaged across significant voxels, indeed correlated exactly in a step-wise manner with the parametric models of the DoM [3, 1, 1, 0] and "DoM + number of Search" [3, 1, 0, −1], respectively (**Figure 5B**). By generalizing the role of Search, we assumed that Search applied to a subject-verb pair, where the relevant features (vowels here) are experimentally "inserted" (Ohta et al., 2013).

We further examined 19 models proposed in theoretical linguistics, psycholinguistics, and natural language processing to verify that the models of the DoM and "DoM + number of Search" best explained the cortical activations (Ohta et al., 2013). All contrasts of Nested$_{(L)}$ − Conjoined$_{(L)}$, etc. predicted that the activations should be exactly zero when a factor produced no effect or load relative to the Conjoined. We thus adopted a no-intercept model, in which percent signal changes of each region were fitted with a single (thus minimal) scale parameter to a model of each factor using its subtracted estimates. For the four contrasts, a least-squares method was used to minimize the residual sum of squares (RSS) for the four fitted values (i.e., four estimates multiplied by a fitting scale) against the corresponding signal changes averaged across participants (**Table 2**).

The model of the DoM for the L. F3op/F3t, as well as that of "DoM + number of Search" for the L. SMG, indeed produced by far the least RSS ($\leq$0.0020) and largest coefficient of determination (r^2) ($\geq$ 0.97). Goodness of fit was further evaluated for each model by using a one-sample *t*-test (significance level at $\alpha = 0.0125$, Bonferroni corrected) between the fitted value for each contrast and individual activations. The model of the DoM for the L. F3op/F3t, as well as that of "DoM + number of Search" for the L. SMG, produced no significant deviation for the four contrasts ($P \geq 0.17$). To further take account of interindividual variability, we fitted "linear mixed-effects models" with individual activations, and found that the models of the DoM and "DoM + number of Search" were by far more likely for the L. F3op/F3t and L. SMG, respectively. Even if we took the Simple condition as a reference for subtracted estimates, we obtained the same results of best models. These results directly support Hypotheses I and II, such that the basic frame of syntactic structures are determined essentially by functional elements, whereas the DoM, together with the number of Search, is a key factor to properly quantify the complexity of the syntactic structures.

THE SIGNIFICANCE OF THE CONNECTIVITY BETWEEN THE L. F3op /F3t AND L. SMG

It has been reported that the L. F3op/F3t is specialized for syntactic processing (Stromswold et al., 1996; Dapretto and Bookheimer, 1999; Embick et al., 2000; Hashimoto and Sakai, 2002; Friederici et al., 2003; Musso et al., 2003; Suzuki and Sakai, 2003; Kinno et al., 2008), suggesting that this region subserves a grammar center (Sakai, 2005). On the other hand, the left angular gyrus and SMG (L. AG/SMG) have been suggested to be important for vocabulary knowledge or lexical processing (Lee et al., 2007; Pattamadilok et al., 2010). To elucidate the relationships between the L. F3op/F3t and L. SMG, we modeled the effective connectivity between these two regions by using dynamic causal modeling (DCM). Our interest was to identify the direction of the connectivity modulated by the Nested condition, which has the largest DoM of all conditions. First, we assumed intrinsic, i.e., task-independent, bi-directional connections, and the models were grouped into three "modulatory families": families with modulation for the bottom-up connection from the L. SMG to L. F3op/F3t, for the top-down connection from the L. F3op/F3t to L. SMG, and for both connections. Each family was composed of three "input models" as regards the regions receiving driving inputs. We found that the model with the modulation for the bottom-up connection, in which the L. F3op/F3t received driving inputs, was the best and most probable model (**Figure 5C**). We further confirmed that the intrinsic top-down connectivity was significantly positive (+0.22; $P < 0.0002$), while the bottom-up connectivity was negatively modulated.

A recent DCM study with a picture-sentence matching task has suggested that the L. F3op/F3t received driving inputs (den

Table 2 | Fittings and likelihood of various models tested in Ohta et al. (2013).

Factor	RSS	r^2	*P*-values for four contrasts	Log-likelihood	Likelihood ratio
L. F3op/F3t					
*DoM	0.0007	0.99	0.17, 0.92, 0.97, 0.99	65.0	1.0
DoM + No. of Search	0.0065	0.88	0.0035, 0.064, 0.63, 0.88	59.2	3.1×10^{-3}
No. of Search	0.052	<0.1	<0.0001, 0.018, 0.019, 0.031	33.4	2.0×10^{-14}
No. of nodes	0.015	0.72	0.0050, 0.0082, 0.018, 0.17	53.7	1.2×10^{-5}
L. SMG					
DoM	0.0063	0.92	0.013, 0.083, 0.44, 0.49	58.8	0.079
*DoM + No. of Search	0.0020	0.97	0.22, 0.30, 0.42, 0.62	61.4	1.0
No. of Search	0.075	<0.1	<0.0001, 0.0061, 0.045, 0.090	23.6	3.8×10^{-17}
No. of nodes	0.033	0.56	0.0004, 0.0005, 0.0061, 0.013	40.1	6.0×10^{-10}

*Percent signal changes in the L. F3op/F3t and L. SMG were fitted with a single scale parameter to a model of each factor using its subtracted estimates (**Table 1**) for the four contrasts of $Nested_{(L)} - Conjoined_{(L)}$, $Nested_{(S)} - Conjoined_{(S)}$, $Simple_{(L)} - Conjoined_{(L)}$, and $Simple_{(S)} - Conjoined_{(S)}$. The P-values for the t-tests are shown in ascending order. The models with an asterisk resulted in the best fit of 19 models tested (four models are shown here) for explaining activations in the L. F3op/F3t or L. SMG, i.e., with the least residual sum of squares (RSS), largest coefficient of determination (r^2), and larger P-values. The likelihood ratio was taken as the ratio of each model's likelihood to the best model's likelihood. The best models were by far more likely than the other models.*

Ouden et al., 2012), which was consistent with our DCM results. Moreover, our previous studies revealed that the functional connectivity between the L. F3t/F3O (pars orbitalis) and L. AG/SMG was selectively enhanced during sentence processing (Homae et al., 2003), and that the L. AG/SMG was also activated during the identification of correct past-tense forms of verbs, probably reflecting an integration of syntactic and vocabulary knowledge (Tatsuno and Sakai, 2005). Considering the role of the L. AG/SMG in lexical processing, the Search operation based on the DoM would be essential in assigning relevant features to the syntactic objects derived from lexical items.

To further confirm the anatomical plausibility of the network between the L. F3op/F3t and L. SMG revealed by DCM, we used diffusion tensor imaging (DTI) with a probabilistic tractography. We observed that a single continuous cluster of the left superior longitudinal and arcuate fasciculi (SLF/AF) connected these regions (cluster size, 3189 mm^3), together with much smaller clusters or islands (**Figure 5D**). Moreover, the left SLF/AF was consistently observed in all participants.

The findings of recent DTI studies have been controversial regarding the functional roles of two different pathways in language processes: the dorsal tracts of the SLF/AF, and the ventral tracts of the middle longitudinal fasciculus (MdLF) and extreme capsule (EmC). Both pathways connect the inferior frontal and superior/middle temporal areas (Saur et al., 2008; Wilson et al., 2011; Wong et al., 2011; Griffiths et al., 2013). Our DCM and DTI results indicate that the L. SMG activations reflecting the DoM mirrored a top-down influence from the L. F3op/F3t through the left dorsal pathway of the SLF/AF, revealing the most crucial network and pathway for syntactic computation.

FURTHER CONFIRMATION OF HYPOTHESES I AND II

A PICTURE-SENTENCE MATCHING PARADIGM

We further examined whether our hypotheses hold for various cases discussed in previous studies. In our fMRI study (Kinno et al., 2008), we used a picture-sentence matching task with three sentence types in Japanese: active, passive, and scrambled sentences (**Figure 6A**). In the picture-sentence matching task, the participants read a sentence covertly and judged whether or not the action depicted in a picture matched the meaning of the sentence. Each sentence had two noun phrases called *arguments*, each of which assumes a different grammatical relation ("subject, direct object, or indirect object" in linguistic terms) and a semantic role ("agent, experiencer, or patient" in linguistic terms, i.e., an agent who performs the action, and an experiencer/patient who is affected by it); these three conditions were thus called Two-argument conditions. More specifically, the active, passive, and scrambled sentences corresponded to "agent and patient" (subject and direct object), "experiencer and agent" (subject and indirect object), and "patient and agent" (direct object and subject) types, respectively. Pictures consisted of two stick figures, each of which was distinguished by a "head" symbol: a circle (○), square (□), or triangle (△). These sentences excluded the involvement of pragmatic information about word use (e.g., "*An officer chases a thief*" is more acceptable than "*A thief chases an officer*"). To minimize the effect of general memory demands, a whole sentence of a minimal length was visually presented for a longer time than was needed to respond.

In Japanese syntax, the grammatical relations are first marked by grammatical particles (nominative, dative, or accusative), which in turn allow the assignment of semantic roles. In the active sentences we used, a noun phrase with the nominative case marker *-ga* (*green* letters in **Figure 6B**) is associated with an agent, and the one with the accusative case marker *-o* is associated with a patient. For the passive sentences we used, however, a noun phrase with the nominative case marker *-ga* is associated with an experiencer (a person experiencing a situation), whereas a passive bound verb "*-(r)areru*" marks passiveness, making a subject-verb pair with the experiencer. In contrast, a noun phrase with the dative marker *-ni* is associated with an agent, whereas an action verb (e.g., "*hik(u)*," "*pull*") makes a subject-verb pair with the agent, forming a subordinate clause within the main clause "○*-ga*... *-(r)areru.*" Note that there exist similar causative

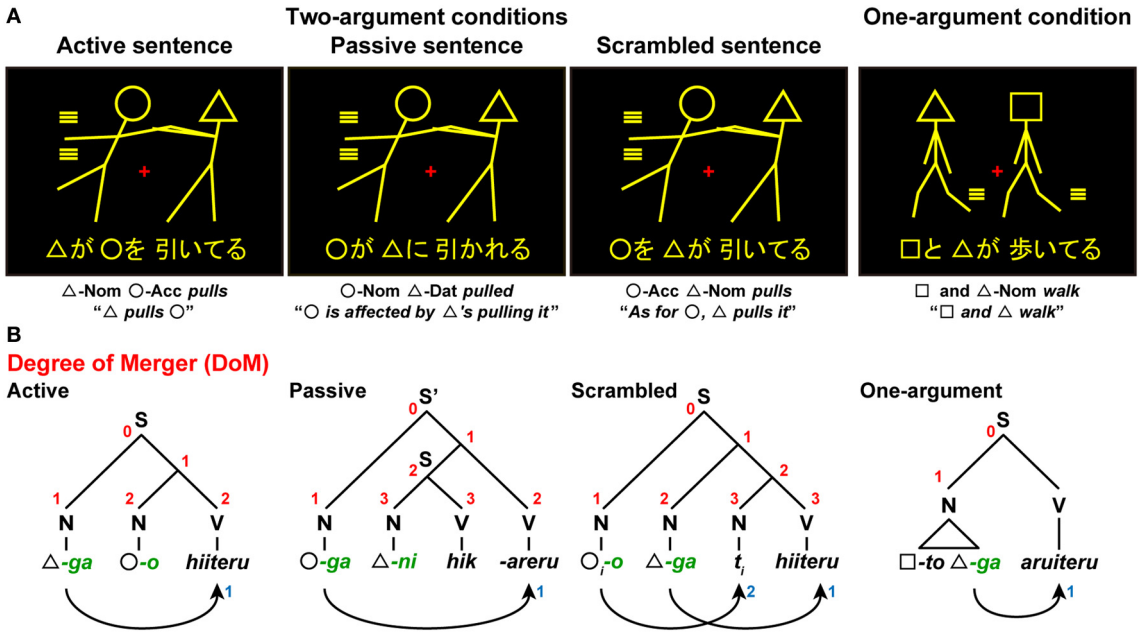


FIGURE 6 | A picture-sentence matching paradigm in Kinno et al. (2008). (A) A picture-sentence matching task under either Two-argument conditions or a One-argument condition. Each stimulus consisted of one picture (top) and one sentence (bottom). Below each example, word-by-word and full translations in English are shown. An identical picture set was used under the Two-argument conditions, where we tested three sentence types: active sentences ("△-*ga* ○-*o hiiteru*"), passive sentences ("○-*ga* △-*ni hik-areru*"), and scrambled sentences ("○-*o* △-*ga hiiteru*"). Under the One-argument condition, we presented syntactically simpler active sentences ("□-*to* △-*ga aruiteru*"). **(B)** The syntactic structures of three sentence types. The digits shown in red and blue denote the DoM for each node and "number of Search," respectively. Symbols used: S and S′, sentence; N, noun phrase; V, verb phrase; Nom, nominative case; Acc, accusative case; Dat, dative case; *-ga*, nominative case marker; *-o*, accusative case marker; *-ni*, dative case marker; *-to*, coordinator; t_i, trace (subscripts denote the same entity).

structures in both Japanese and English: "*Hanako-ga kare-ni hik-aseta*," "*Hanako made him pull*." Actually, there are two types of passivization in Japanese: *ni* passive (e.g., "*Hanako-ga Taro-ni hik-areru*," "*Hanako is affected by Taro's pulling her*") and *ni yotte* passive (e.g., "*Hanako-ga Taro-ni yotte hik-areru*," "*Hanako is pulled by Taro*"). According to Kuroda (1992), the *ni* passive involves no noun-phrase movement, while the *ni yotte* passive involves a movement similar to the case in English. For the scrambled sentences, an object moves from its canonical position to higher nodes by undergoing another Merge operation. This type of constructions is perfectly normal, not only in Japanese but in German, Finnish, and other languages. We also tested the One-argument condition, under which each sentence was presented with an intransitive verb and double agents. This condition did not involve two-argument relationships, and was thus syntactically simpler than any of the Two-argument conditions.

HYPOTHESIS III

Here we present the following hypothesis (Hypothesis III):

(3) The DoM domain changes dynamically in accordance with iterative Merge applications, the Search distances, and/or task requirements.

Since Merge combines two syntactic objects to form a larger structure, Merge always produces a one-level higher node. When Merge applies iteratively to an existing phrase or sentence, the DoM domain becomes thus larger in accordance with the number of Merge applications. The Search distance is the structural distance between two distinct parts to which the Search operation applies, regardless of the nodes that are irrelevant to the Search operation. As observed from **Figure 4**, the DoM domain changes in accordance with the Search distance. On the other hand, for every sentence stimulus in the study of Ohta et al. (2013), the construction of syntactic structures was ensured by task requirements, in which three sentence types had to be distinguished while they were completely mixed. Task requirements include not only certain constraints required by experimental tasks, but detailed parsing naturally required to understand a part of phrases or sentences (e.g., subject-verb relationships and noun-pronoun (coreference) relationships).

In the above mentioned paradigm (Kinno et al., 2008), the four task conditions (three sentence types under the Two-argument conditions, as well as one type under the One-argument condition) were completely mixed (see **Figure 6A**). With such task requirements, the DoM domain spanned three relevant words for all sentence types under the Two-argument conditions. Under the One-argument condition, the action of two stick figures was always identical, and thus a subject (a triangle just below N in **Figure 6B**) is regarded as a unit. Under these four task conditions, participants were required to check at least one of the argument-verb relationships, demanding Search at least once. For

the scrambled sentences alone, an additional Search operation should match the identical indices of the moved object and its trace. For the active, passive, and scrambled sentences, the estimates of DoM were 2, 3, and 3, respectively, while those of the DoM was 1 under the One-argument condition.

APPLYING THE DoM TO VARIOUS SENTENCE TYPES

In the study of Kinno et al. (2008), we directly contrasted passive and active sentence conditions to identify a cortical region that is activated by purely syntactic processes. This stringent contrast resulted in significant activation in the left dorsal F3t (L. dF3t) alone [(−48, 24, 21), $Z = 3.8$] (**Figure 7A**), which was very close to the L. F3op/F3t activation in the study of Ohta et al. (2013). The L. dF3t activation was significantly enhanced under both the passive and scrambled sentence conditions compared to that under the active sentence condition ($P \leq 0.033$) (**Figure 7B**), whereas there was no significant difference between the passive and scrambled sentence conditions ($P = 0.15$). Taking the One-argument condition as a reference for subtracted estimates, the signal changes in the L. dF3t were precisely correlated in a step-wise manner with the parametric model of the DoM [1, 2, 2], producing the RSS of 0.0001 and r^2 of 0.99, without significant deviation for the three contrasts ($P \geq 0.87$). The model of the DoM thus *sufficiently* explains the L. dF3t activations. It should be noted that the parametric model of "the number of nodes" [2, 4, 4] also yielded the same fitting results in this case. The design of experimental paradigms limits the separation of multiple factors.

In a recent fMRI study, only right-branching constructions were examined, and activations in the L. F3t were modulated by the size of constituents (i.e., number of terminal nodes) (Pallier et al., 2011). Since the estimates of the DoM were identical to those of "the number of Merge" or "the number of non-terminal nodes" in this case, it was not possible to separate these factors. Taking their simplest condition (lists of unrelated words) as an appropriate reference, the model of the DoM actually showed a comparable or better goodness of fit for activations in the L. F3t, when compared with their log-fitting models.

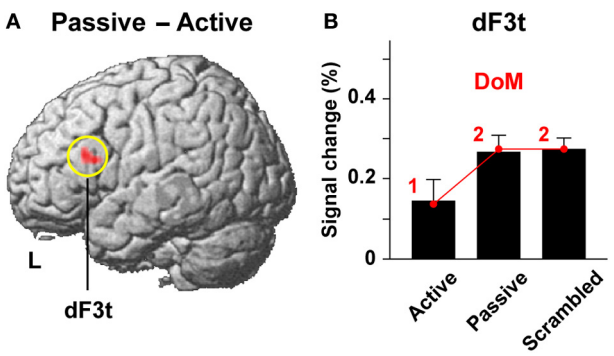


FIGURE 7 | Activations in the L. dF3t modulated by the DoM. **(A)** A region identified by the Passive – Active contrast (see **Figure 6**). Activations were projected onto the left (L) lateral surface of a standard brain. **(B)** Percent signal changes for the active, passive, and scrambled sentence conditions in the L. dF3t, taking the One-argument condition as a reference. Overlaid red dots and lines denote the values fitted with the estimates (digits in red) for the model of the DoM.

FURTHER CONFIRMATION OF HYPOTHESIS III

THE EFFECT OF THE SEARCH DISTANCES ON THE DoM

Neuroimaging and psycholinguistic studies have reported that English sentences with object-relative clauses have higher processing loads than those with subject-relative clauses (Just et al., 1996; Stromswold et al., 1996; Gibson, 2000). To properly parse the relative clauses, the relative pronoun and its antecedent are coindexed; "*who*$_i$" and "*the boy*$_i$," respectively, in the example shown in **Figure 8**. In a subject-relative clause, a relative pronoun "*who*$_i$" was displaced from the *subject* position denoted by a trace t_i (originally, "*the boy*$_i$ *likes the girl*"), while in an object-relative clause, a relative pronoun was displaced from the *object* position (originally, "*the girl likes the boy*$_i$"). Following the proposal by Hawkins (1999), we assume that the relative pronoun searches the corresponding trace within tree structures of a sentence (see curved arrows in **Figure 8**). In a subject-relative clause, Search ends at the initiation of the verb phrase, while in an object-relative clause, Search ends *after* a verb appears within a subordinate clause. In accordance with the Search distances for these examples, the DoM would become one unit larger for the object-relative clause than the subject-relative one. Higher processing

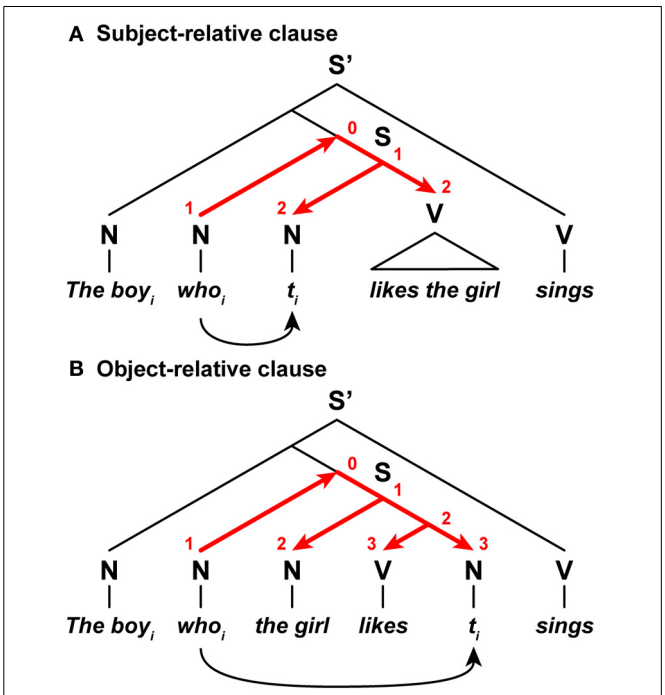


FIGURE 8 | The DoM domains varied with the Search distances. **(A)** A sentence with a subject-relative clause. **(B)** A sentence with an object-relative clause. In these relative clauses, a relative pronoun *who*$_i$ is displaced from its subject or object position denoted by a trace t_i. A set of red straight arrows corresponds to the DoM domain. The digits shown in red denote the DoM for each node within the domain. Symbols used: S and S′, sentence; N, noun phrase; V, verb phrase; t_i, trace (subscripts denote the same entity).

loads observed with object-relative clauses are consistent with this inference about the DoM domain.

THE EFFECT OF TASK REQUIREMENTS ON THE DoM

If Hypothesis III is correct, then the L.F3op/F3t activations can be different in accordance with task requirements, even when the same sentences are presented. In our previous fMRI study, we compared three explicit linguistic tasks with the same set of normal two-word sentences: syntactic decision, semantic decision, and phonological decision tasks (Suzuki and Sakai, 2003). In the syntactic decision task, the participants judged whether or not the presented sentence was syntactically correct, and this judgment required syntactic knowledge about the distinction between transitive and intransitive verbs (e.g., normal sentence, "*yuki-ga tumoru*," "*snow lies* (*on the ground*)"; anomalous sentence, "*yuki-o tumoru*," "(*something*) *lies snow*"). In the semantic decision task, lexico-semantic knowledge about selectional restrictions was indispensable. In the phonological decision task, phonological knowledge about accent patterns was required. Neither the semantic decision task nor the phonological decision task, both with *implicit* syntactic processing, elicited significant activations in the L. F3op/F3t (−57, 9, 6), which was significantly activated during *explicit* syntactic processing, even by a direct comparison between the syntactic decision task and the other tasks. These results suggest the presence of the DoM domain in accordance with the task requirements of explicit syntactic processing.

THE MIXED EFFECTS OF THE SEARCH DISTANCES AND TASK REQUIREMENTS ON THE DoM

In another fMRI study, we directly compared syntactic decision and short-term memory tasks (Hashimoto and Sakai, 2002). In this unique paradigm, we visually presented nested sentences that included two proper nouns, two verbs, and one pronoun, in which either verb or pronoun was underlined. After presenting one complete sentence in a phrase-by-phrase manner, paired phrases including an underlined phrase were shown. In one syntactic decision task (SYN-1), participants were required to judge whether the subject of an underlined verb corresponded to the person in paired phrases (**Figure 9A**). In this case, the Search distance was the structural distance between the subject and verb of the same clause. In the other syntactic decision task (SYN-2), the participants were required to judge whether an underlined pronoun was able to refer to the person in paired phrases (**Figure 9B**). In this case, the Search distance was the structural distance between the coindexed noun and pronoun. In these syntactic decision tasks, the Search distance, and consequently the DoM domain, changed dynamically in accordance with the different task requirements, even when the same sentences were presented. The estimate of the resultant DoM was 2 for both cases. In a short-term memory task with a sentence, the participants memorized the linear order of the phrases, and judged whether the left-hand phrase preceded the right-hand one in the original sequence (**Figure 9C**). With such a task requirement, the factor of DoM would become less effective. Indeed, we found that activations in the L. F3op/F3t were equally enhanced in both syntactic decision tasks when compared with the short-term memory task.

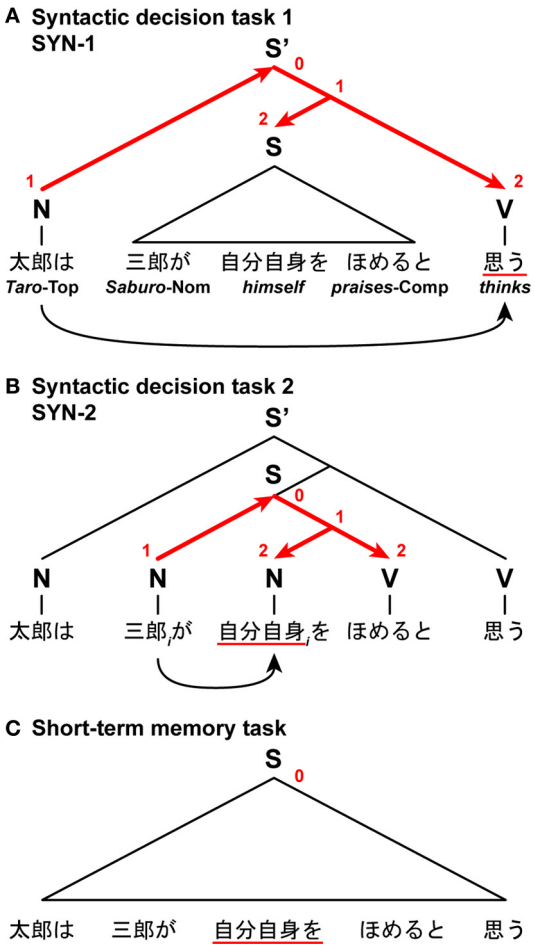


FIGURE 9 | The DoM domains varied with the Search distances and task requirements. In this task, participants read Japanese nested sentences ("*Taro-wa Saburo-ga jibunjishin-o homeru-to omou*," "*Taro thinks that Saburo praises himself*"), in which phrases were sequentially presented. **(A)** A syntactic decision task 1, in which participants judged subject-verb relationships. A set of red straight arrows corresponds to the DoM domain. The digits shown in red denote the DoM for each node within the domain. **(B)** A syntactic decision task 2, in which participants judged noun-pronoun (coreference) relationships (subscripts denote the same entity). **(C)** A short-term memory task with a sentence, in which participants judged the temporal order of the phrases. Symbols used: S and S', sentence; N, noun phrase; V, verb phrase; Top, topic; Nom, nominative case; Comp, complementizer.

CONCLUSIONS

In this article, we reviewed recent advances in theoretical linguistics and functional neuroimaging in the following respects. First, we provided theoretical discussions about the hierarchical tree structures of sentences, and introduced the two fundamental linguistic operations of Merge and Search. We also presented our hypotheses that the DoM is a key computational concept to properly measure the complexity of tree structures (Hypothesis I), and that the basic frame of the syntactic structure of a given linguistic expression is determined essentially by functional elements, which trigger Merge and Search operations (Hypothesis

II). Second, we presented our recent fMRI studies, which have demonstrated that the DoM, together with the number of Search, is indeed a key syntactic factor that accounts for syntax-selective activations in the L. F3op/F3t and L. SMG (Ohta et al., 2013). Moreover, based on the DCM and DTI results, we revealed the significance of the top-down connection from the L. F3op/F3t to L. SMG, suggesting that information about the DoM is transmitted through this specific dorsal pathway. Third, we further hypothesized that the DoM domain changes dynamically in accordance with iterative Merge applications, the Search distances, and/or task requirements (Hypothesis III). We showed that the DoM sufficiently explains activation modulations due to different structures reported in previous fMRI studies (Kinno et al., 2008; Pallier et al., 2011). Finally, we confirmed that Hypothesis III accounts for higher processing loads observed with object-relative clauses, as well as activations in the L. F3op/F3t during explicit syntactic decision tasks, reported in the previous neuroimaging and psycholinguistic studies (Just et al., 1996; Stromswold et al., 1996; Gibson, 2000; Hashimoto and Sakai, 2002; Suzuki and Sakai, 2003). It is likely that the DoM serves as a key computational principle for other human-specific cognitive capacities, such as mathematics and music, both of which can be expressed by hierarchical tree structures. A future investigation into the computational principles of syntax will further deepen our understanding of uniquely human mental faculties.

ACKNOWLEDGMENTS

We would like to thank R. Kinno, H. Miyashita, K. Iijima, and T. Inubushi for their helpful discussions, N. Komoro for her technical assistance, and H. Matsuda for her administrative assistance. This research was supported by a Core Research for Evolutional Science and Technology (CREST) grant from the Japan Science and Technology Agency (JST), by Grants-in-Aid for Scientific Research (S) (Nos. 20220005) from the Ministry of Education, Culture, Sports, Science and Technology, and by a Grant-in-Aid for Japan Society for the Promotion of Science (JSPS) Fellows (No. 24·8931).

REFERENCES

Abe, K., and Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* 14, 1067–1074. doi: 10.1038/nn.2869

Bahlmann, J., Schubotz, R. I., and Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *Neuroimage* 42, 525–534. doi: 10.1016/j.neuroimage.2008.04.249

Beckers, G. J. L., Bolhuis, J. J., Okanoya, K., and Berwick, R. C. (2012). Birdsong neurolinguistics: Songbird context-free grammar claim is premature. *NeuroReport* 23, 139–145. doi: 10.1097/WNR.0b013e32834f1765

Chomsky, N. (1957). *Syntactic Structures*. Hague: Mouton Publishers.

Chomsky, N. (1959). On certain formal properties of grammars. *Inform. Control* 2, 137–167. doi: 10.1016/S0019-9958(59)90362-6

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: The MIT Press.

Chomsky, N. (2004). "Beyond explanatory adequacy," in *Structures and Beyond: The Cartography of Syntactic Structures*, Vol. 3, ed A. Belletti (Oxford: Oxford University Press), 104–131.

Corballis, M. C. (2007). Recursion, language, and starlings. *Cogn. Sci.* 31, 697–704. doi: 10.1080/15326900701399947

Dapretto, M., and Bookheimer, S. Y. (1999). Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron* 24, 427–432. doi: 10.1016/S0896-6273(00)80855-7

den Ouden, D.-B., Saur, D., Mader, W., Schelter, B., Lukic, S., Wali, E., et al. (2012). Network modulation during complex syntactic processing. *Neuroimage* 59, 815–823. doi: 10.1016/j.neuroimage.2011.07.057

Embick, D., Marantz, A., Miyashita, Y., O'Neil, W., and Sakai, K. L. (2000). A syntactic specialization for Broca's area. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6150–6154. doi: 10.1073/pnas.100098897

Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., and Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb. Cortex* 13, 170–177. doi: 10.1093/cercor/13.2.170

Fukui, N. (2011). "Merge and Bare Phrase Structure," in *The Oxford Handbook of Linguistic Minimalism*, ed C. Boeckx (Oxford: Oxford University Press), 73–95.

Fukui, N., and Sakai, H. (2003). The visibility guideline for functional categories: Verb raising in Japanese and related issues. *Lingua* 113, 321–375. doi: 10.1016/S0024-3841(02)00080-3

Gentner, T. Q., Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207. doi: 10.1038/nature04675

Gibson, E. (2000). "The dependency locality theory: a distance-based theory of linguistic complexity," in "*Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, eds A. Marantz, Y. Miyashita, and W. O'Neil, (Cambridge, MA: The MIT Press), 95–126.

Griffiths, J. D., Marslen-Wilson, W. D., Stamatakis, E. A., and Tyler, L. K. (2013). Functional organization of the neural language system: dorsal and ventral pathways are critical for syntax. *Cereb. Cortex* 23, 139–147. doi: 10.1093/cercor/bhr386

Gunji, T. (1987). *Japanese Phrase Structure Grammar: A Unification-Based Approach*. Dordrecht: D. Reidel Publishing Company. doi: 10.1007/978-94-015-7766-3

Hashimoto, R., and Sakai, K. L. (2002). Specialization in the left prefrontal cortex for sentence comprehension. *Neuron* 35, 589–597. doi: 10.1016/S0896-6273(02)00788-2

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language* 75, 244–285. doi: 10.2307/417261

Homae, F., Yahata, N., and Sakai, K. L. (2003). Selective enhancement of functional connectivity in the left prefrontal cortex during sentence processing. *Neuroimage* 20, 578–586. doi: 10.1016/S1053-8119(03)00272-6

Hopcroft, J. E., and Ullman, J. D. (1979). *Introduction to Automata Theory, Languages, and Computation*. Reading, MA: Addison-Wesley.

Ivana, A., and Sakai, H. (2007). Honorification and light verbs in Japanese. *J. East Asian Ling.* 16, 171–191. doi: 10.1007/s10831-007-9011-7

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., and Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science* 274, 114–116. doi: 10.1126/science.274.5284.114

Kinno, R., Kawamura, M., Shioda, S., and Sakai, K. L. (2008). Neural correlates of noncanonical syntactic processing revealed by a picture-sentence matching task. *Hum. Brain Mapp.* 29, 1015–1027. doi: 10.1002/hbm.20441

Kuroda, S.-Y. (1992). "On Japanese passives," in *Japanese Syntax and Semantics: Collected Papers*, ed S.-Y. Kuroda (Dordrecht: Kluwer Academic), 183–221. doi: 10.1007/978-94-011-2789-9_6

Lee, H., Devlin, J. T., Shakeshaft, C., Stewart, L. H., Brennan, A., Glensman, J., et al. (2007). Anatomical traces of vocabulary acquisition in the adolescent brain. *J. Neurosci.* 27, 1184–1189. doi: 10.1523/JNEUROSCI.4442-06.2007

Mandelbrot, B. B. (1977). *The Fractal Geometry of Nature*. New York, NY: W. H. Freeman and Company.

Miller, G. A., and Chomsky, N. (1963). "Finitary Models of Language Users," in *Handbook of Mathematical Psychology,* Vol. 2, eds R. D. Luce, R. R. Bush, and E. Galanter, (New York, NY: John Wiley and Sons), 419–491.

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., et al. (2003). Broca's area and the language instinct. *Nat. Neurosci.* 6, 774–781. doi: 10.1038/nn1077

Ohta, S., Fukui, N., and Sakai, K. L. (2013). Syntactic computation in the human brain: the degree of merger as a key factor. *PLoS ONE* 8:e56230, 1–16. doi: 10.1371/journal.pone.0056230

Pallier, C., Devauchelle, A.-D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2522–2527. doi: 10.1073/pnas.1018711108

Pattamadilok, C., Knierim, I. N., Duncan, K. J. K., and Devlin, J. T. (2010). How does learning to read affect speech perception? *J. Neurosci.* 30, 8435–8444. doi: 10.1523/JNEUROSCI.5791-09.2010

Rodriguez, P. (2001). Simple recurrent networks learn context-free and context-sensitive languages by counting. *Neural Comput.* 13, 2093–2118. doi: 10.1162/089976601750399326

Saito, M., and Fukui, N. (1998). Order in phrase structure and movement. *Ling. Inq.* 29, 439–474. doi: 10.1162/002438998553815

Sakai, K. L. (2005). Language acquisition and brain development. *Science* 310, 815–819. doi: 10.1126/science.1113530

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., et al. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18035–18040. doi: 10.1073/pnas.0805234105

Stromswold, K., Caplan, D., Alpert, N., and Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain Lang.* 52, 452–473. doi: 10.1006/brln.1996.0024

Suzuki, K., and Sakai, K. L. (2003). An event-related fMRI study of explicit syntactic processing of normal/anomalous sentences in contrast to implicit syntactic processing. *Cereb. Cortex* 13, 517–526. doi: 10.1093/cercor/13.5.517

Tatsuno, Y., and Sakai, K. L. (2005). Language-related activations in the left prefrontal regions are differentially modulated by age, proficiency, and task demands. *J. Neurosci.* 25, 1637–1644. doi: 10.1523/JNEUROSCI.3978-04.2005

Tsujimura, N. (2007). *An Introduction to Japanese Linguistics,* 2nd Edn. Malden, MA: Blackwell Publishing.

Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., et al. (2011). Syntactic processing depends on dorsal language tracts. *Neuron* 72, 397–403. doi: 10.1016/j.neuron.2011.09.014

Wong, F. C. K., Chandrasekaran, B., Garibaldi, K., and Wong, P. C. M. (2011). White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. *J. Neurosci.* 31, 8780–8785. doi: 10.1523/JNEUROSCI.0999-11.2011

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 September 2013; paper pending published: 17 October 2013; accepted: 01 December 2013; published online: 18 December 2013.

*Citation: Ohta S, Fukui N and Sakai KL (2013) Computational principles of syntax in the regions specialized for language: integrating theoretical linguistics and functional neuroimaging. Front. Behav. Neurosci. **7**:204. doi: 10.3389/fnbeh.2013.00204*

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 19 September 2013
doi: 10.3389/fnbeh.2013.00121

Neural substrates of figurative language during natural speech perception: an fMRI study

Arne Nagels[1]*, Christina Kauschke[2], Judith Schrauf[2], Carin Whitney[3], Benjamin Straube[1] and Tilo Kircher[1]

[1] *Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Marburg, Germany*
[2] *Department of Germanic Linguistics, Philipps-University Marburg, Marburg, Germany*
[3] *Department of Psychology and York Neuroimaging Centre, University of York, York, UK*

Edited by:
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA

Reviewed by:
Seth D. Norrholm, Emory University School of Medicine, USA
Yueqiang Xue, The University of Tennessee Health Science Center, USA

***Correspondence:**
Arne Nagels, Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Rudolf-Bultmann-Str. 8, 35039 Marburg, Germany
e-mail: nagels@med.uni-marburg.de

Many figurative expressions are fully conventionalized in everyday speech. Regarding the neural basis of figurative language processing, research has predominantly focused on metaphoric expressions in minimal semantic context. It remains unclear in how far metaphoric expressions during continuous text comprehension activate similar neural networks as isolated metaphors. We therefore investigated the processing of similes (figurative language, e.g., "He smokes like a chimney!") occurring in a short story. Sixteen healthy, male, native German speakers listened to similes that came about naturally in a short story, while blood-oxygenation-level-dependent (BOLD) responses were measured with functional magnetic resonance imaging (fMRI). For the event-related analysis, similes were contrasted with non-figurative control sentences (CS). The stimuli differed with respect to figurativeness, while they were matched for frequency of words, number of syllables, plausibility, and comprehensibility. Similes contrasted with CS resulted in enhanced BOLD responses in the left inferior (IFG) and adjacent middle frontal gyrus. Concrete CS as compared to similes activated the bilateral middle temporal gyri as well as the right precuneus and the left middle frontal gyrus (LMFG). Activation of the left IFG for similes in a short story is consistent with results on single sentence metaphor processing. The findings strengthen the importance of the left inferior frontal region in the processing of abstract figurative speech during continuous, ecologically-valid speech comprehension; the processing of concrete semantic contents goes along with a down-regulation of bilateral temporal regions.

Keywords: figurative speech, simile, abstractness, inferior frontal gyrus, fMRI

INTRODUCTION

Figurative expressions are an established part of everyday speech and are often fully conventionalized. For example, parts of the human body are used in a multitude of figurative expression (head of department, eye of needle, arm of tree, etc.). Thus, figurative speech use goes far beyond the concept of a mere stylistic device and can be seen as an integral part of day-to-day communication. So far, it remains unclear in how far the comprehension of figurative speech draws on additional resources when presented in a naturally-evolving continuous and coherent story. Increases in activation of the left inferior frontal gyrus (IFG) have previously been reported for isolated sentences that contained figurative as opposed to non-figurative elements (Rapp et al., 2004, 2007; Kircher et al., 2007). However, prior context and repeated exposure to figurative speech—as it appears in more natural environments—can have an impact on how we process figures of speech such as similes. Context and conventionalization may facilitate comprehension (for idiom comprehension see Gibbs, 1992) and therefore neural activation of contextualized similes might not differ from non-simile sentences.

The current study focuses on the processing of figurative speech in form of similes such as "The sun is like the eye of heaven" in a natural context without constraining instructions or predetermined cognitive tasks, e.g., decision tasks; asking the participants to press a button, when either abstract or concrete content was presented.

A simile, such as "The sun is like the eye of heaven," can be divided into three different components: (1) the explained element ("the sun"), (2) the explaining element ("the eye of heaven"), (3) and the term of comparison (TOC; "is like"), which connects the two elements in the simile (Leech, 1969). For successful comprehension, the listener needs to refer back to the explained element and identify similarities or common features with the corresponding explaining element (e.g., "He smokes like a chimney!" stresses that someone smokes heavily). The aspect that the two elements have in common is the so-called tertium comparationis, i.e., the third domain involved in a comparison. According to prevailing theories, similes are strongly linked to metaphors, which can be regarded as similes without the TOC (i.e., elliptical similes; cf. Aristotle).

When metaphors are presented, the listener is confronted with a semantic conflict between the explained and explaining element, which needs to be resolved. The initiator of a figurative utterance selects, emphasizes, suppresses, and organizes

features of the explained element by applying characteristics of the explaining elements. Thus, a "mental linkage" between the explained element and the corresponding explaining element is required which goes beyond the usual semantic or word-by-word analysis (Rapp et al., 2004). The TOC ("like") in similes may facilitate this linking process, since it explicitly points to the comparative nature of the utterance.

Usually, similes or metaphors do not stand alone but are integrated into a speech or text. This context can alter the ease by which meaning is integrated, including processing of figures of speech (Gibbs, 1992). Prior context facilitates the comprehension of idioms when it is consistent with the specific entailments of the idiom (i.e., their conceptual representation): though there might be different ways of expressing anger in an idiomatic way (e.g., "bite your head-off" as opposed to "blow your stack"), it is easier to comprehend idioms whose specific conceptual representation has been primed by previous information e.g., describing anger in a way that refers to anger as "animal behavior" (Nayak and Gibbs, 1990). A coherent and evolving story can provide such information, and participants should therefore understand figurative expressions that are embedded in a story easier than isolated sentences or sentence pairs with little detail (e.g., Rapp et al., 2007; Schmidt and Seger, 2009). Behavioral studies of other complex semantic operations, such as ambiguity processing, have also shown that cognitive resources can be saved when the target item is embedded in semantically coherent context as opposed to isolated or neutral environments [for a review see Simpson (1994)]. Also, the amount of prior, coherent information can facilitate inference making and improve comprehension of non-figurative material (Zwaan and Radvansky, 1998).

A number of functional Magnetic Resonance Imaging (fMRI) studies have investigated the neural correlates of figurative speech mostly in the form of metaphoric sentences (Rapp et al., 2004, 2007; Eviatar and Just, 2006; Kircher et al., 2007; Mashal et al., 2007, 2009; Shibata et al., 2007; Stringaris et al., 2007; Schmidt and Seger, 2009; Desai et al., 2011; Diaz and Hogstrom, 2011; Diaz et al., 2011), and for similes (Shibata et al., 2012). The brain response to similes in a story context, however, remains unexplored, thus far. The results of these previous studies support the involvement of the left lateral prefrontal cortex [for a critical review on the neural basis of metaphor processing see Schmidt et al. (2010)], as a correlate for increased cognitive demand during the comprehension of figurative language. In particular, Rapp et al. (2004) for the first time reported enhanced cortical activation in the left IFG [Brodmann area (BA) 45/47] for metaphor reading as compared to literal sentences. The left IFG is an integral component of the semantic processing network and has been related to executive aspects of meaning retrieval, such as semantic search, retrieval, selection, and integration (e.g., Thompson-Schill et al., 1997; Wagner et al., 2001; Noppeney et al., 2004; Badre et al., 2005; Badre and Wagner, 2007; Bedny et al., 2008; Binder et al., 2009). Moreover, difficult metaphors in comparison to easy metaphors such as "Political success is a house of cards" vs. "Books are treasure chests of information" (Schmidt and Seger, 2009) as well as anomalous metaphors ["Their (financial) capital has a lot of rhythm' (Ahrens et al., 2007)] were found to selectively activate the left IFG. Similarly, conventional metaphors as compared to novel metaphors were found to engage the left IFG, whereas novel metaphors activated the left middle frontal gyrus (LMFG) during a reading paradigm (Mashal et al., 2007). The graded response in prefrontal cortex, particularly the left IFG, suggests that activation correlates with the level of cognitive-semantic resources required for successful performance (e.g., integration effort). This is in accordance with previous semantic retrieval studies of graded difficulty, which showed enhanced left IFG response during tasks of high vs. low executive-semantic demands (Thompson-Schill et al., 1997; Roskies et al., 2001; Wagner et al., 2001; Badre et al., 2005; Snyder et al., 2007; Zempleni et al., 2007; Kuperberg et al., 2008; Nagel et al., 2008; Ruff et al., 2008; Snijders et al., 2009; Whitney et al., 2009a,b). A recent study found activations in the head of the caudate presenting metaphors in a context (Uchiyama et al., 2012). However, the majority of past studies on figurative language comprehension utilized highly controlled sentence reading paradigms with minimal prior information (Rapp et al., 2004, 2007; Eviatar and Just, 2006; Stringaris et al., 2006, 2007; Mashal et al., 2007; Yang et al., 2009). The observed left IFG activations for metaphors might partly reflect executive-semantic processes that are related to the lack of sufficient contextual priming, as it would occur in naturally evolving texts.

The aim of the current study was therefore to analyze the neural responses to simile processing and determine the involvement of the left IFG when similes were presented within a natural, unconstrained short story context. Naturalistic stimuli were successfully investigated in the context of different experimental fMRI paradigms (Hasson et al., 2004; Skipper et al., 2007; Domahs et al., 2012). However, the neural correlates of understanding similes in a short story have not been investigated, so far. We hypothesized left IFG activation during processing of sentences containing similes (e.g., he jumps like a gazelle) vs. literal sentences of comparable frequency, plausibility, comprehensibility, and length. In addition, we expected significant correlations between unfamiliar as well as highly abstract similes and enhanced blood-oxygenation-level-dependent (BOLD) responses in the left frontal region.

MATERIALS AND METHODS

PARTICIPANTS

Initially, 19 male subjects took part in the fMRI study. Due to head movement, data of three subjects had to be discarded from further analysis. For the remaining 16 participants movement was minimal as the maximum change in translation and rotation for each participant was less than one voxel size (i.e., 3.5 mm) and less than 1°, respectively. All 16 participants (mean age = 27.00 years, $SD = 6.65$; mean years of education = 14.50 years, $SD = 1.67$) were native speakers of German, right-handed according to the Edinburgh Inventory of Handedness (Oldfield, 1971) and showed average or above average verbal IQ as assessed by the German MWT-B multiple choice vocabulary test (Lehrl et al., 1995; mean estimated verbal IQ = 120.06, $SD = 17.16$). Subjects with recent substance use or general MRI incompatibility (e.g., metal implants) were excluded. All subjects gave informed consent and were paid 20 Euros for participation in the study.

The local ethics committee at RWTH Aachen University approved the study.

STIMULI

A slightly modified version of the short story "Der Kuli Kimgun" by Dauthendey (1930) was chosen for this study as in Whitney et al. (2009b). Low-frequent or foreign words were substituted by more familiar or high-frequent words. The final version included a total of 3581 words.

In general, short stories are well-structured narratives restricted to a few protagonists and basic narrative events which, together, are ascribed to a single, central conflict. Our story was written from an omniscient perspective, leaving out elaborate descriptions of the character's emotional or mental states. The sequence of events occurs chronologically, which allows the listener to build up a temporally continuous mental representation of the story. The short story chosen for the current fMRI study contained figurative descriptions of events and situations.

For auditory presentation during fMRI, the story was professionally recorded and spoken in a natural way by a trained, male speech therapist. The duration of the story was 23:32 min.

PROCEDURE

The story was presented via MRI compatible headphones in two successive runs lasting 14:32 and 9:00 min, respectively. Participants were instructed to close their eyes and listen to the story carefully. To make sure that subjects attended to the content, they were informed at the beginning of the experiment about a short interview after the MRI session about the content of the story. Hereby, 10 questions regarding critical episodes of the short story had to be answered.

FIGURATIVE AND CONTROL SENTENCES

First, 32 similes as well as 50 randomly chosen control sentences (CS) were extracted from the short story independently by two linguists (Christina Kauschke and Judith Schrauf). In a behavioral test, the isolated similes and CS were rated by 20 volunteers, who did not take part in the fMRI study, according to the dimensions "plausibility," "comprehensibility," and "figurativeness." For this purpose, an analogue scale from 1 to 7 was used. Regarding the plausibility rating the instruction was as follows: "Please rate the subsequent sentences according to their plausibility. Very plausible sentences describe ordinary events being easy to follow." Comprehensibility was rated according to the instruction: "Please rate the subsequent sentences according to their comprehensibility. Very comprehensible sentences are those where the meaning can be understood easily and within a very short period of time." Regarding the dimension figurativeness, the rating instruction was formulated: "Sentences can be distinguished with regard to their properties to evoke inner pictures from the content being conveyed. The figurative content can be perceived faster and be grasped more easily in some sentences. Please rate the degree of figurativeness in the following sentences."

For the imaging analysis, 30 similes and 30 CS were chosen so that no significant differences were found between the similes and the CS according to the dimensions of plausibility [$F_{(1.58)} = 2.68$, $p = 0.108$], comprehensibility [$F_{(1.58)} = 493$, $p = 0.49$], and figurativeness [$F_{(1.58)} = 1.76$, $p = 0.19$]. The similes were found to be rather unfamiliar [mean = 2.91 ($SD = 2.14$)] and abstract [mean = 4.86 ($SD = 2.29$)]. All similes and CS were matched according to word frequency as well as to the number of syllables. Similes in comparison to CS revealed no significant differences with regard to their individual length in the story.

fMRI DATA ACQUISITION

All scanning was performed on a 1.5 T scanner (Gyroscan Intera, Philips Medical, Eindhoven, The Netherlands) using standard gradients and a circular polarized phase array head coil. For each subject, we acquired two series of functional volumes of T2*-weighted axial EPI-scans parallel to the AC/PC line with the following parameters: number of slices (NS), 22; slice thickness (ST), 5.0 mm; interslice gap (IG), 0.55 mm; matrix size (MS), 64 × 64; field of view (FOV), 240 × 240 mm; echo time (TE), 50 ms; repetition time (TR), 2.0 s. Four hundred and thirty-six functional volumes were acquired for the first part of the story and 270 functional volumes for the second part, adding up to 706 volumes in total.

fMRI DATA ANALYSIS

MR images were analyzed using Statistical Parametric Mapping software (SPM5; www.fil.ion.ucl.ac.uk) implemented in MATLAB (v. R2006b, Mathworks Inc., Sherborn, MA). After discarding the first three volumes, all images were realigned to the first image to correct for head movement. Unwarping was used to correct for the interaction of susceptibility artifacts and head movement. After realignment and unwarping, the signal measured in each slice was shifted relative to the acquisition time of the middle slice using a sinc interpolation in time to correct for their different acquisition times. Volumes were then normalized into standard stereotaxic anatomical MNI-space by using the transformation matrix calculated from the first EPI-scan of each subject and the EPI-template. Afterwards, the normalized data with a resliced voxel size of 4 × 4 × 4 mm were smoothed with a 10 mm full width at half maximum (FWHM) isotropic Gaussian kernel to accommodate intersubject variation in brain anatomy. The time series data were filtered with a high-pass cut-off of 1/128 Hz. The autocorrelation of the data was estimated and corrected for.

Onsets for the simile phrases were set at the beginning of the phrase that referred to the explained element. The duration was measured individually for each simile and included the explained element, TOC, and explaining element. CS were modeled in a similar way, with the onset at the beginning of the phrase and the duration of the event being equal to the duration of the complete phrase. A random-effects group analysis was performed entering the contrast images for similes and CS from the first-level analysis into a full-factorial design matrix.

In a separate analysis, rating values for figurativeness, familiarity, and abstractness were entered individually as parametric variates for each simile into the first-level analysis in order to analyze correlations between brain responses during figurative speech processing and the aforementioned dimensions.

A further *post-hoc* analysis was performed with respect to the particular role of the TOC in the processing of abstract figurative speech. Therefore, the particular onset of the TOC (engl. "as"

or "like"; e.g., in "He smokes like a chimney") was individually calculated using an event-related design.

The results were corrected on a voxel-wise threshold of $p < 0.001$. Hereby, a Monte Carlo simulation of the brain volume of the current study was conducted to establish an appropriate voxel contiguity threshold (Slotnick et al., 2003). The procedure is based on the fact that the probability of observed clusters of activity due to voxel-wise Type I error (i.e., noise) decreases systematically as cluster size increases. Assuming an individual voxel type I error of $p < 0.001$ in our study, a cluster extent of 13 contiguous resampled voxels was indicated as necessary to correct for multiple voxel comparisons.

Each of the reported activations was determined with the Anatomy Toolbox for SPM5 (v. 1.7b, http://www.fz-juelich.de/inm/inm-1/spm_anatomy_toolbox). The imaging figures were made with the MRIcron software package (http://www.cabiatl.com/mricro/mricron/).

RESULTS

BEHAVIORAL RESULTS

During the post-scan interview about critical story episodes, all participants were able to recall all the desired details in response to each of the 10 questions (Whitney et al., 2009b). Answers to all questions were provided quickly and effortlessly.

fMRI RESULTS

Simile > CS

In the whole brain analysis, the simile sentences as contrasted with the CS selectively activated the left IFG including the pars triangularis and adjacent middle frontal gyrus (LMFG; $p < 0.05$, Monte Carlo corr.; **Figure 1**). Local maxima were found in both regions, though BOLD responses for the left IFG were stronger (**Table 1**). Extracted beta values for the activated region revealed activations for CS as well, however, activations were significantly stronger for the simile condition.

CS > simile

A network of activations encompassing the precuneus, bilateral middle temporal gyri as well as the LMFG was found for the whole brain *CS > simile* contrast. A large cluster of activation extended from the left (BA 23) and right precuneus (BA 7) to left middle (BA 31) and posterior cingulate cortex (BA 7). Activations in the right middle temporal gyrus (BA 19) extended into the region of the angular gyrus (BA 39) as well as the middle occipital gyrus. Contralaterally, left middle temporal as well as the left angular gyrus (BA 39) were found to be more activated during CS processing. Enhanced BOLD responses were also found in the LMFG as well as in the bilateral middle and inferior temporal gyri (BA 20; **Table 1**).

CORRELATION ANALYSES

Familiarity

The correlation analysis between the degree of familiarity and BOLD signal changes revealed a pattern of activation in the left parahippocampal region (**Table 1**, **Figure 1**).

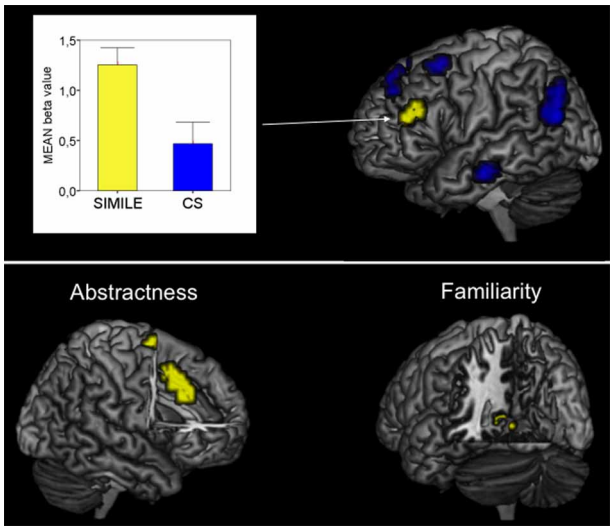


FIGURE 1 | Top: Imaging results for the contrast Simile > CS. The bar graph illustrates the contrast estimates (beta values; yellow bar = Simile, blue bar = CS). **Bottom**: Imaging results for the interaction of brain responses with both familiar and abstract similes.

Abstractness

Highly abstract similes resulted in BOLD enhancements in the anterior cingulate cortex as well as in activations in the right superior frontal gyrus (**Table 1**, **Figure 1**).

Figurativeness

No significant relation was found between BOLD enhancements and figurativeness.

Post-hoc ANALYSES RESULTS

TOC > CS

Enhanced BOLD responses for the TOC as opposed to CS were found in the left-hemispheric IFG (p. triangularis) and the superior parietal region (Supplementary Material).

CS > TOC

The opposed contrast revealed pronounced activations in the right precuneus, middle temporal gyrus, and angular gyrus. In the left hemisphere, CS > TOC elicited neural responses in the middle temporal region (Supplementary Material).

([TOC > CS] > [SIMILE > CS])

The contrast for TOC > CS as opposed to SIMILE > CS again resulted in left lateralized activations encompassing the IFG and the superior area of the parietal cortex (Supplementary Material).

([SIMILE > CS] > [TOC > CS])

The inverse contrast resulted in right hemispheric activations in the precuneus and the middle temporal region (Supplementary Material).

SIMILE > CS ∩ TOC > CS

The conjunction analyses for both similes and TOC as contrasted with CS activated the left pars triangularis (Supplementary Material).

Table 1 | Peak activation within clusters for the contrasts Simile > CS, CS > Simile as well as for the correlation analysis with familiarity and abstractness (whole-brain analysis, Monte Carlo corr. $p < 0.001$).

		BA	Coordinates			*t*-value	No. voxels
			x	*y*	*z*		
SIMILE >CS							
L	Inferior frontal gyrus (p. Triangularis)	45	−48	32	24	4.93	17
			−48	44	20	4.56	
CS > SIMILE							
R	Precuneus	7	4	−60	32	6.20	184
			0	−60	16	5.80	
			0	−44	36	4.85	
R	Middle temporal gyrus	20	52	−4	−24	6.12	44
			60	−12	−20	4.42	
L	Middle temporal gyrus	39	−56	−64	20	5.97	77
			−44	−60	24	5.54	
			−56	−68	32	5.12	
L	Middle frontal gyrus	6	−32	20	56	5.18	17
L	Middle temporal gyrus	20	−60	−12	−24	4.72	28
			−48	−16	−20	4.48	
R	Middle temporal gyrus	39	60	−64	12	4.69	102
			60	−60	24	4.64	
			52	−68	28	4.64	
FAMILIARITY							
L	Parahippocampal gyrus	30	−8	−40	0	4.50	25
ABSTRACTNESS							
L	Anterior cingulate	24	−4	24	28	4.87	42
			−8	16	36	4.47	
			0	8	44	3.46	
R	Superior frontal gyrus	6	12	4	72	4.01	14

Coordinates refer to MNI space.

CS > SIMILE ∩ CS > TOC

CS as opposed to similes and TOC activated a neural pattern, encompassing the right precuneus, middle temporal gyrus as well as the angular gyrus. In the left hemisphere, enhanced BOLD responses were for the middle temporal region (Supplementary Material).

DISCUSSION

Figurative expressions, such as metaphors and similes are fundamental to language and thought. They represent a conventionalized part of everyday communication (Lakoff and Johnson, 2003). The processing of such figurative expressions requires a mental linkage between the explained and the explaining element. In case of a particular kind of metaphors, i.e., similes, this linkage is made explicit by the use of a TOC ("as" or "like," German: "wie"), also referred to as a hedge word (Shibata et al., 2012). The aim of the present study was to investigate the neural basis of simile processing using a highly naturalistic, continuous speech perception paradigm. This allowed us to examine whether brain activation reported in previous investigations of figurative language in the left IFG also holds true in a natural setting of short story comprehension. In line with findings on metaphor processing (Rapp et al., 2004) we observed an involvement of the left IFG for the simile condition, suggesting that more neural resources are required to interpret the figurative meaning of the simile under naturalistic conditions. We moreover found significant correlations between enhanced BOLD responses in the anterior cingulate region and the superior frontal gyrus in the context of highly abstract figurative expressions. Correlation analysis for familiar similes resulted in activations in the left hippocampal region.

The neural processing of continuous, naturalistic stimuli has thus far only rarely been performed. Recent studies have either explored natural speech production (Kircher et al., 2000, 2004; Buchheim et al., 2006), narrative comprehension (Wilson et al., 2008; Whitney et al., 2009b; Brennan et al., 2012; Domahs et al., 2012), or naturalistic audio-visual processing mechanisms (Bartels and Zeki, 2004; Hasson et al., 2004, 2010).

SIMILE PROCESSING IN THE LEFT IFG

The results demonstrate that similes elicited enhanced BOLD responses in the dorsal part of the pars triangularis (left IFG) and the ventral portion of the LMFG when contrasted with non-figurative CS. The increased activation in these brain regions might reflect the enhanced demand on deep semantic processing integrating figurative expressions into the surrounding context. Based on the initial semantic conflict between explained and explaining element, the listener compares the figurative expression by means of a parallel, which is drawn to a different entity ("tertium comparationis"). Thus, common characteristics as well as distinct features of both elements are to be selected, emphasized, inhibited, and organized which goes beyond the usual level of contextual semantic word processing (Rapp et al., 2004).

Enhanced neural responses in the left IFG have previously been found in studies on metaphors compared to literal phrases using single sentences (Rapp et al., 2004). In a recent fMRI study, Schmidt and Seger (2009) compared sentences with easy and difficult metaphors. While easy metaphors were found to selectively activate the left MFG, difficult metaphors elicited enhanced BOLD responses in the left IFG. Similarly, the processing of metaphorical sentences taken from poetry resulted in activations of the left dorsolateral prefrontal cortex (Mashal et al., 2009). Bambini and colleagues investigated the neural correlates of implicit metaphor processing as compared to non-metaphorical passages, while being explicitly involved in an adjective matching task to be performed after reading the target passages (Bambini et al., 2011). The authors found a widespread neural network encompassing the left and right inferior frontal gyri, the right superior temporal gyrus, the left angular gyrus, and the anterior cingulate region. Imaging results were interpreted in terms of integrating linguistic material and world knowledge into the context. The left IFG and in particular the pars triangularis may hence represent a key region for figurative speech processing, including similes in a story context.

ACTIVATION FOR CONTROL SENTENCES

A widespread bilateral cortical network was found to be activated for control vs. simile sentences. Thus, enhanced BOLD responses were observed for CS in contrast to similes in the middle temporal gyrus bilaterally, the precuneus and the LMFG. Since the CS were matched with regard to plausibility and comprehensibility, enhanced activation may reflect general semantic analysis of concrete information continuously presented as previously found by Whitney et al. (2009a,b). The activations, in particular in the bilateral temporal gyri, were consistently reported for auditory language processing [for review see: Ferstl et al. (2008)]. Continuous listening to the concrete sentences includes inferences for bridging successive utterances, the use of background knowledge about concrete entities of the world and discourse context as well as lexical retrieval (Ferstl et al., 2008), all processes that have been attributed to the neural network found in our study.

The fact that the processing of similes resulted in a reduced involvement of this bilateral "concrete sentence" network indicates that either concrete representations were inhibited in favor of the relevant abstract interpretation, or the double representation (e.g., in the sentence "He smokes like a chimney!" the representation of smoke will be activated by both smoke and chimney) of the respective concept led to a facilitation of related processing mechanisms. Nevertheless, this finding in general supports the theory that abstract figurative meaning is mainly represented in the left hemisphere (Perlovsky and Ilin, 2010).

ACTIVATIONS FOR FAMILIARITY

All of our similes were non-conventional, but more or less familiar to the listeners. We found a positive association between familiarity of similes and activation in the left parahippocampal region. No negative correlations with BOLD enhancements were found. These data suggest that familiar, lexicalized and therefore well-known similes, e.g., "serve like a slave," are associated with enhanced semantic memory processes (Hoenig and Scheef, 2005). Thus, the enhanced semantic memory retrieval from the long-term storage as well as the integrative associative-mnemonic processes (Hoenig and Scheef, 2005) suggest a contribution of the left parahippocampal region to the processing of familiar figurative speech. With regard to the neural substrates of metaphor processing easy and familiar metaphors as contrasted with literal sentences have previously been found to activate the left parahippocampal gyrus (Schmidt and Seger, 2009). Similarly, Yang and colleagues (Yang et al., 2009) revealed activations in the bilateral hippocampal gyri for conventional metaphors, e.g., "She is a peach," as opposed to a redundant condition such as "She is a female." In the current investigation significant correlations for familiar similes were solely restricted to BOLD enhancements in the left parahippocampal region. A number of reasons may account for the selective recruitment of the parahippocampal gyrus. First, the linguistic structure of similes as compared to other metaphoric expressions differs with regard to the presence of a TOC (usually "like"). The presence of a mental linkage presumably facilitates the understanding of the figurative speech part, which might be explained by evoking wider associations of memory. Second, the auditory task design asking the participants to listen carefully to the narrative instead of reading or judging the isolated figurative or literal expressions, respectively, differs from recent experimental designs (Shibata et al., 2007; Schmidt and Seger, 2009; Yang et al., 2009).

ACTIVATIONS FOR ABSTRACTNESS

In general, the anterior cingulate is involved in many processes, such as verbal working memory as well as in selective attention, online-monitoring processes, and abstract auditory sequencing (Carter et al., 1998; Macdonald et al., 2000; Lee et al., 2011). With regard to figurative speech processing, enhanced neural responses in the anterior cingulate were previously reported (Rapp et al., 2004; Shibata et al., 2007; Bambini et al., 2011; Diaz and Hogstrom, 2011). In the current study, however, BOLD enhancements in this region and in the right superior frontal gyrus were found for more abstract similes, such as "he moved forward like a swimmer against the tide." This pattern of activation suggests an involvement of enhanced semantic selection and monitoring processes, since relevant aspects and appropriate literal meanings of the abstract explaining element must be filtered and interpreted. These enhanced cognitive control mechanisms together with the comparatively stronger abstract semantic integration demands may have resulted in the recruitment of the anterior cingulate.

ACTIVATIONS FOR TERM OF COMPARISON

Post-hoc analyses for the TOC resulted in a neural pattern of activations encompassing left hemispheric pars triangularis as well as superior parts of the parietal region. The conjunction analyses with similes (including the whole figurative speech phrase) again resulted in BOLD enhancements in the pars triangularis. It might be hypothesized that the TOC early predicts the upcoming figurative information; the TOC represents a bridging element linking the concrete—the explained element—to the subsequent abstract mental image. The TOC ("like") may support this linking process, since it explicitly points to the comparative nature of the utterance. Moreover, it can be assumed that the competition between the abstract and the literal meaning resulting in the additional recruitment of the left IFG (Chen et al., 2008).

LIMITATIONS

Analyzing continuous and authentic speech perception in a natural context goes also along with a number of methodological problems. CS —though carefully matched—still represent an arbitrary selection that could differ in a specific aspect, which cannot be systematically controlled for. Finally, ratings (e.g., plausibility evaluations) have been performed on isolated sentences and consequently do not consider the specific narrative context. However, we could demonstrate a highly comparable result pattern for the processing of similes in contrast to CS previously found for highly controlled experiments on figurative speech processing (e.g., Rapp et al., 2004, 2007; Kircher et al., 2007). Correlation analyses moreover revealed plausible result patterns indicating that sentence evaluations are associated with corresponding cognitive mechanisms.

CONCLUSIONS

The present study suggests that the left IFG plays a crucial role in the processing of figurative comparisons embedded into highly naturalistic continuous speech processing within a short story. These findings add novel plausibility to previous, highly restrained experiments and show the applicability of this approach. In general, future investigations may consider employing experimental paradigms, using ecologically valid and naturally evolving stimulus material, e.g., including contextual information or multi-modal processing (audio-visual perception), with a high resemblance to the real world.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/BehavioralNeuroscience/10.3389/fnbeh.2013.00121/abstract

REFERENCES

Ahrens, K., Liu, H. L., Lee, C. Y., Gong, S. P., Fang, S. Y., and Hsu, Y. Y. (2007). Functional MRI of conventional and anomalous metaphors in Mandarin Chinese. *Brain Lang.* 100, 163–171. doi: 10.1016/j.bandl.2005.10.004

Badre, D., Poldrack, R. A., Pare-Blagoev, E. J., Insler, R. Z., and Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47, 907–918. doi: 10.1016/j.neuron.2005.07.023

Badre, D., and Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* 45, 2883–2901. doi: 10.1016/j.neuropsychologia.2007.06.015

Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M., and Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Res. Bull.* 86, 203–216. doi: 10.1016/j.brainresbull.2011.07.015

Bartels, A., and Zeki, S. (2004). Functional brain mapping during free viewing of natural scenes. *Hum. Brain Mapp.* 21, 75–85. doi: 10.1002/hbm.10153

Bedny, M., McGill, M., and Thompson-Schill, S. L. (2008). Semantic adaptation and competition during word comprehension. *Cereb. Cortex* 18, 2574–2585. doi: 10.1093/cercor/bhn018

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., and Pylkkanen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain Lang.* 120, 163–173. doi: 10.1016/j.bandl.2010.04.002

Buchheim, A., Erk, S., George, C., Kachele, H., Ruchsow, M., Spitzer, M., et al. (2006). Measuring attachment representation in an FMRI environment: a pilot study. *Psychopathology* 39, 144–152. doi: 10.1159/000091800

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., and Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280, 747–749. doi: 10.1126/science.280.5364.747

Chen, E., Widick, P., and Chatterjee, A. (2008). Functional-anatomical organization of predicate metaphor processing. *Brain Lang.* 107, 194–202. doi: 10.1016/j.bandl.2008.06.007

Dauthendey, M. (1930). *Das Rauschen der großen Muschel*. Stuttgart: DVA.

Desai, R. H., Binder, J. R., Conant, L. L., Mano, Q. R., and Seidenberg, M. S. (2011). The neural career of sensory-motor metaphors. *J. Cogn. Neurosci.* 23, 2376–2386. doi: 10.1162/jocn.2010.21596

Diaz, M. T., Barrett, K. T., and Hogstrom, L. J. (2011). The influence of sentence novelty and figurativeness on brain activity. *Neuropsychologia* 49, 320–330. doi: 10.1016/j.neuropsychologia.2010.12.004

Diaz, M. T., and Hogstrom, L. J. (2011). The influence of context on hemispheric recruitment during metaphor processing. *J. Cogn. Neurosci.* 23, 3586–3597. doi: 10.1162/jocn_a_00053

Domahs, F., Nagels, A., Domahs, U., Whitney, C., Wiese, R., and Kircher, T. (2012). Where the mass counts: common cortical activation for different kinds of nonsingularity. *J. Cogn. Neurosci.* 24, 915–932. doi: 10.1162/jocn_a_00191

Eviatar, Z., and Just, M. A. (2006). Brain correlates of discourse processing: an fMRI investigation of irony and conventional metaphor comprehension. *Neuropsychologia* 44, 2348–2359. doi: 10.1016/j.neuropsychologia.2006.05.007

Ferstl, E. C., Neumann, J., Bogler, C., and Von Cramon, D. Y. (2008). The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Hum. Brain Mapp.* 29, 581–593. doi: 10.1002/hbm.20422

Gibbs, W. (1992). What do idioms really mean? *J. Mem. Lang.* 31, 485–506. doi: 10.1016/0749-596X(92)90025-S

Hasson, U., Malach, R., and Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* 14, 40–48. doi: 10.1016/j.tics.2009.10.011

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506

Hoenig, K., and Scheef, L. (2005). Mediotemporal contributions to semantic processing: fMRI evidence from ambiguity processing during semantic context verification. *Hippocampus* 15, 597–609. doi: 10.1002/hipo.20080

Kircher, T. T., Brammer, M. J., Levelt, W., Bartels, M., and McGuire, P. K. (2004). Pausing for thought: engagement of left temporal cortex during pauses in speech. *Neuroimage* 21, 84–90. doi: 10.1016/j.neuroimage.2003.09.041

Kircher, T. T., Brammer, M. J., Williams, S. C., and McGuire, P. K. (2000). Lexical retrieval during fluent speech production: an fMRI study. *Neuroreport* 11, 4093–4096. doi: 10.1097/00001756-200012180-00036

Kircher, T. T., Leube, D. T., Erb, M., Grodd, W., and Rapp, A. M. (2007). Neural correlates of metaphor processing in schizophrenia. *Neuroimage* 34, 281–289. doi: 10.1016/j.neuroimage.2006.08.044

Kuperberg, G. R., Sitnikova, T., and Lakshmanan, B. M. (2008). Neuroanatomical distinctions within the semantic system during sentence comprehension: evidence from functional magnetic resonance imaging. *Neuroimage* 40, 367–388. doi: 10.1016/j.neuroimage.2007.10.009

Lakoff, G., and Johnson, M. (2003). *Metaphors We Live By*. Chicago; London: The University of Chicago Press. doi: 10.7208/chicago/9780226470993.001.0001

Lee, Y. S., Janata, P., Frost, C., Hanke, M., and Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *Neuroimage* 57, 293–300. doi: 10.1016/j.neuroimage.2011.02.006

Leech, G. (1969). *A Linguistic Guide to English Poetry*. London: Longman.

Lehrl, S., Triebig, G., and Fischer, B. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurol. Scand.* 91, 335–345. doi: 10.1111/j.1600-0404.1995.tb07018.x

Macdonald, A. W. 3rd., Cohen, J. D., Stenger, V. A., and Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288, 1835–1838. doi: 10.1126/science.288.5472.1835

Mashal, N., Faust, M., Hendler, T., and Jung-Beeman, M. (2007). An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain Lang.* 100, 115–126. doi: 10.1016/j.bandl.2005.10.005

Mashal, N., Faust, M., Hendler, T., and Jung-Beeman, M. (2009). An fMRI study of processing novel metaphoric sentences. *Laterality* 14, 30–54. doi: 10.1080/13576500802049433

Nagel, I. E., Schumacher, E. H., Goebel, R., and D'Esposito, M. (2008). Functional MRI investigation of verbal selection mechanisms in lateral prefrontal cortex. *Neuroimage* 43, 801–807. doi: 10.1016/j.neuroimage.2008.07.017

Nayak, N. P., and Gibbs, R. W. Jr. (1990). Conceptual knowledge in the interpretation of idioms. *J. Exp. Psychol. Gen.* 119, 315–330. doi: 10.1037/0096-3445.119.3.315

Noppeney, U., Phillips, J., and Price, C. (2004). The neural areas that control the retrieval and selection of semantics. *Neuropsychologia* 42, 1269–1280. doi: 10.1016/j.neuropsychologia.2003.12.014

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)9006

Perlovsky, L. I., and Ilin, R. (2010). Neurally and mathematically motivated architecture for language and thought. *Open Neuroimag. J.* 4, 70–80. doi: 10.2174/1874440001004010070

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., and Kircher, T. T. (2004). Neural correlates of metaphor processing. *Brain Res. Cogn. Brain Res.* 20, 395–402. doi: 10.1016/j.cogbrainres.2004.03.017

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., and Kircher, T. T. (2007). Laterality in metaphor processing: lack of evidence from functional magnetic resonance imaging for the right hemisphere theory. *Brain Lang.* 100, 142–149. doi: 10.1016/j.bandl.2006.04.004

Roskies, A. L., Fiez, J. A., Balota, D. A., Raichle, M. E., and Petersen, S. E. (2001). Task-dependent modulation of regions in the left inferior frontal cortex during semantic processing. *J. Cogn. Neurosci.* 13, 829–843. doi: 10.1162/08989290152541485

Ruff, I., Blumstein, S. E., Myers, E. B., and Hutchison, E. (2008). Recruitment of anterior and posterior structures in lexical-semantic processing: an fMRI study comparing implicit and explicit tasks. *Brain Lang.* 105, 41–49. doi: 10.1016/j.bandl.2008.01.003

Schmidt, G. L., Kranjec, A., Cardillo, E. R., and Chatterjee, A. (2010). Beyond laterality: a critical assessment of research on the neural basis of metaphor. *J. Int. Neuropsychol. Soc.* 16, 1–5. doi: 10.1017/S1355617709990543

Schmidt, G. L., and Seger, C. A. (2009). Neural correlates of metaphor processing: the roles of figurativeness, familiarity and difficulty. *Brain Cogn.* 71, 375–386. doi: 10.1016/j.bandc.2009.06.001

Shibata, M., Abe, J., Terao, A., and Miyamoto, T. (2007). Neural mechanisms involved in the comprehension of metaphoric and literal sentences: an fMRI study. *Brain Res.* 1166, 92–102. doi: 10.1016/j.brainres.2007.06.040

Shibata, M., Toyomura, A., Motoyama, H., Itoh, H., Kawabata, Y., and Abe, J. (2012). Does simile comprehension differ from metaphor comprehension? A functional MRI study. *Brain Lang.* 121, 254–260. doi: 10.1016/j.bandl.2012.03.006

Simpson, G. B. (1994). "Context and the processing of ambiguous words," in *Handbook of Psycholinguistics*, ed M. A. Gernsbacher (San Diego, CA: Academic Press), 359–374.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., and Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang.* 101, 260–277. doi: 10.1016/j.bandl.2007.02.008

Slotnick, S. D., Moo, L. R., Segal, J. B., and Hart, J. Jr. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Res. Cogn. Brain Res.* 17, 75–82. doi: 10.1016/S0926-6410(03)00082-X

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J., Petersson, K. M., and Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: an FMRI study using word-category ambiguity. *Cereb. Cortex* 19, 1493–1503. doi: 10.1093/cercor/bhn187

Snyder, H. R., Feigenson, K., and Thompson-Schill, S. L. (2007). Prefrontal cortical response to conflict during semantic and phonological tasks. *J. Cogn. Neurosci.* 19, 761–775. doi: 10.1162/jocn.2007.19.5.761

Stringaris, A. K., Medford, N., Giora, R., Giampietro, V. C., Brammer, M. J., and David, A. S. (2006). How metaphors influence semantic relatedness judgments: the role of the right frontal cortex. *Neuroimage* 33, 784–793. doi: 10.1016/j.neuroimage.2006.06.057

Stringaris, A. K., Medford, N. C., Giampietro, V., Brammer, M. J., and David, A. S. (2007). Deriving meaning: distinct neural mechanisms for metaphoric, literal, and non-meaningful sentences. *Brain Lang.* 100, 150–162. doi: 10.1016/j.bandl.2005.08.001

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., and Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 14792–14797. doi: 10.1073/pnas.94.26.14792

Uchiyama, H. T., Saito, D. N., Tanabe, H. C., Harada, T., Seki, A., Ohno, K., et al. (2012). Distinction between the literal and intended meanings of sentences: a functional magnetic resonance imaging study of metaphor and sarcasm. *Cortex* 48, 563–583. doi: 10.1016/j.cortex.2011.01.004

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., and Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31, 329–338. doi: 10.1016/S0896-6273(01)00359-2

Whitney, C., Grossman, M., and Kircher, T. T. (2009a). The influence of multiple primes on bottom-up and top-down regulation during meaning retrieval: evidence for 2 distinct neural networks. *Cereb. Cortex* 19, 2548–2560. doi: 10.1093/cercor/bhp007

Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., and Kircher, T. (2009b). Neural correlates of narrative shifts during auditory story comprehension. *Neuroimage* 47, 360–366. doi: 10.1016/j.neuroimage.2009.04.037

Wilson, S. M., Molnar-Szakacs, I., and Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cereb. Cortex* 18, 230–242. doi: 10.1093/cercor/bhm049

Yang, F. G., Edens, J., Simpson, C., and Krawczyk, D. C. (2009). Differences in task demands influence the hemispheric lateralization and neural correlates of metaphor. *Brain Lang.* 111, 114–124. doi: 10.1016/j.bandl.2009.08.006

Zempleni, M. Z., Renken, R., Hoeks, J. C., Hoogduin, J. M., and Stowe, L. A. (2007). Semantic ambiguity processing in sentence context: evidence from event-related fMRI. *Neuroimage* 34, 1270–1279. doi: 10.1016/j.neuroimage.2006.09.048

Zwaan, R. A., and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162–185. doi: 10.1037/0033-2909.123.2.162

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 July 2013; paper pending published: 13 August 2013; accepted: 26 August 2013; published online: 19 September 2013.

Citation: Nagels A, Kauschke C, Schrauf J, Whitney C, Straube B and Kircher T (2013) Neural substrates of figurative language during natural speech perception: an fMRI study. Front. Behav. Neurosci. 7:121. doi: 10.3389/fnbeh.2013.00121

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 03 February 2014
doi: 10.3389/fnbeh.2014.00013

Making fingers and words count in a cognitive robot

Vivian M. De La Cruz[1], Alessandro Di Nuovo[2,3]*, Santo Di Nuovo[4,5] and Angelo Cangelosi[2]

[1] *Dipartimento di Scienze Cognitive, della Formazione e degli Studi Culturali, Università degli Studi di Messina, Messina, Italy*
[2] *Centre for Robotics and Neural Systems, School of Computing and Mathematics, Plymouth University, Plymouth, UK*
[3] *Facoltà di Ingegneria e Architettura, Università degli Studi di Enna "Kore", Enna, Italy*
[4] *Dipartimento dei Scienze della Formazione, Università degli Studi di Catania, Catania, Italy*
[5] *Unità operativa di Psicologia, IRCCS Oasi Maria SS di Troina, Enna, Italy*

Edited by:
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA

Reviewed by:
Anna M. Borghi, University of Bologna and Institute of Cognitive Sciences and Technologies, Italy
Giovanni Acampora, Nottingham Trent University, UK

****Correspondence:***
Alessandro Di Nuovo, Centre for Robotics and Neural Systems, School of Computing and Mathematics, Plymouth University, B109 PSQ, Drake Circus, PL4 8AA, Plymouth, UK
e-mail: alessandro.dinuovo@plymouth.ac.uk

Evidence from developmental as well as neuroscientific studies suggest that finger counting activity plays an important role in the acquisition of numerical skills in children. It has been claimed that this skill helps in building motor-based representations of number that continue to influence number processing well into adulthood, facilitating the emergence of number concepts from sensorimotor experience through a bottom-up process. The act of counting also involves the acquisition and use of a verbal number system of which number words are the basic building blocks. Using a Cognitive Developmental Robotics paradigm we present results of a modeling experiment on whether finger counting and the association of number words (or tags) to fingers, could serve to bootstrap the representation of number in a cognitive robot, enabling it to perform basic numerical operations such as addition. The cognitive architecture of the robot is based on artificial neural networks, which enable the robot to learn both sensorimotor skills (finger counting) and linguistic skills (using number words). The results obtained in our experiments show that learning the number words in sequence along with finger configurations helps the fast building of the initial representation of number in the robot. Number knowledge, is instead, not as efficiently developed when number words are learned out of sequence without finger counting. Furthermore, the internal representations of the finger configurations themselves, developed by the robot as a result of the experiments, sustain the execution of basic arithmetic operations, something consistent with evidence coming from developmental research with children. The model and experiments demonstrate the importance of sensorimotor skill learning in robots for the acquisition of abstract knowledge such as numbers.

Keywords: embodied cognition, developmental robotics, finger counting, number words, number cognition

INTRODUCTION

Whether finger counting is an essential stage in the development of the cognition of number is still highly debated, though strong evidence exists on the positive contribution of sensorimotor skills and representation in numerical cognition. A growing number of researchers, consider finger counting an important tool children (as well as adults) use across a variety of cultures in the development of numerical cognition (e.g., Andres et al., 2008; Di Luca and Pesenti, 2011). Consideration of the links between counting and the emergence of number concepts is not new. Piaget (1952) for example, considered the linking of numbers to objects as being an important characteristic of the sensorimotor stage of cognitive development, possibly being one of a series of prerequisites for the child's construction of the concept of number. Quite recently, however, the topic of finger based number knowledge has seen a surge of new interest, especially from embodied cognition perspectives (for a recent special issue on the topic see Fischer et al., 2012). Finger counting has generally been assumed to be important to the acquisition of a mature counting system (e.g., Gelman and Gallistel, 1978; Fuson et al., 1982; Butterworth, 2005) as well as instrumental to the development of children's arithmetic abilities (e.g., Fuson and Kwon, 1992). It has been hypothesized that this capability helps children to acquire a variety of principles proposed as being fundamental to the development of a counting system (Gelman and Gallistel, 1978) such as: the acquisition of the one-to-one correspondence principle (i.e., when counting only one word is assigned to each object) through the tagging or assignment of one number word to each item, the assimilation of the stable order principle (i.e., when counting, number words are always assigned in the same order) and the cardinality principle (i.e., the last number word uttered when counting, is the total number of objects in a set). More recent studies have reported an association between finger gnosis (the ability to mentally represent one's fingers) and mathematical abilities (Noël, 2005; Costa et al., 2011), and found finger training helpful in improving the performance of children with weak numerical skills (Gracia-Bafalluy and Noël, 2008). Evidence such as this supports the view that finger representations play a special role in number cognition, and might serve as a basic building block in the child's unfolding capacity to mentally manipulate abstract numerical information.

That a close link might exist between finger counting strategies and patterns, and that they may influence the mental representation and processing of number, has also been suggested by evidence coming from neuroimaging studies. For example, studies using fMRI on adult subjects to investigate aspects of embodied theories of cognition have found intrinsic functional links between finger counting and number processing. Cortical motor activity is evoked by Arabic digits and number words, which reflects particular individual finger counting habits (i.e., whether when counting small digits subjects started with their right or left hand) (Tschentscher et al., 2012). These results have been interpreted in several different ways by the authors of this study, one interpretation invoking a shared neural network for number processing and planning of finger movements, which would include parietal cortical areas, the precentral gyrus and the primary motor cortex, in which number perception might very well elicit the sub-threshold tendency to move associated fingers. Another interpretation used by these authors, to explain how the association between numbers, number words and individual finger counting movements might have come about in their subjects, during their individual development of numerical skills in childhood, would be predicted by a Hebbian learning approach to semantic circuits (Pulvermüller, 1999). The prediction is, that due to the fact that children often use their fingers when counting and solving simple counting problems, a correlation between the neuronal activation for the processing of numbers and the movement of fingers is established. A number of neuroimaging studies done in the last decade, using both PET and fMRI, had already found activation of part of the left precentral gyrus (where hand movements are represented) when subjects were asked to engage in numerical tasks such as addition (e.g., Pesenti et al., 2000), subtraction (Rueckert et al., 1996) and multiplication (Dehaene et al., 1996), leading some authors to suggest that the activation of the left precentral gyrus, along with the inferior parietal cortex, might be evidence of a finger moving network that might, in turn, be reflecting a trace of a finger counting strategy (Sato and Lalain, 2008).

The act of counting often involves the acquisition and use of a verbal number system, of which, number words are the basic building blocks. Number words are highly frequent in child directed speech, but their meanings are acquired slowly, with effort and in stages. Wynn (1992) has argued, in fact, that the developing knowledge of the meanings of counting words is a central part of the process of understanding the counting system. Though children as young as 6 months can discriminate between set sizes (Xu et al., 2005), and children 1 and 2 years of age show to be good at reciting the count sequence (Fuson, 1988), as well as capable of recognizing number words as designators of quantity (Bloom and Wynn, 1997), their difficulty seems to lie in understanding *how* specific words match to specific quantities. One proposal is that the syntax of number words as well as the contexts, in which they appear, might be serving as cues that help children bootstrap this process early (e.g., Gleitman, 1990; Wynn, 1992; Bloom and Wynn, 1997), but finger counting may also very well be serving as an early entry point to this understanding.

Other evidence coming from developmental as well as neurocognitive studies, in keeping with what has been found in neuroimaging studies, suggest that finger counting activity, helps build motor-based representations of number that continue to influence number processing well into adulthood, suggesting that abstract cognition may be rooted in bodily experience (Domahs et al., 2010). In fact, these motor-based representations have been argued to facilitate the emergence of number concepts from sensorimotor experience through a bottom-up process (Andres et al., 2008). In our view, finger counting, can also be seen as a means by which direct sensory experience with the body can serve the purpose of *grounding* number as well as number words initially as low level labels, that later serve as the basis for the acquisition of new higher level symbols from the combination of already grounded ones, something known as grounding transfer (e.g., Harnad, 1990; Cangelosi and Riga, 2006). The grounding approach has also been useful for the modeling of the acquisition of words for objects (Morse et al., 2010; Tikhanoff et al., 2011) and for actions (Marocco et al., 2010; Stramandoli et al., 2012) as well as for numbers (Ruciński et al., 2011, 2012).

In sum, while finger counting may not be strictly necessary for children to get on their way to the cognition of number, there is evidence that it does seem to help the learning process, serving as a bridge between possibly innate abilities to perceive and respond to numerosity (e.g., Butterworth, 2005) and the development of the capacity to mentally represent and process number as well as linguistic number related concepts (Lafay et al., 2013). Not much work using robotics has attempted to build on this.

A number of connectionist models have simulated different aspects of number learning. Ma and Hirai (1989) for example, studied how children learn to count using an associative memory network model, which mimicked three phenomena proposed by Fuson et al. (1982), to be present in the acquisition of counting by children [i.e., number word sequence produced by children dividable into three distinct portions: conventional, stable nonconventional, and unstable; irregular number words (e.g., "fifteen") omitted more often than regular ones ("fourteen," "sixteen"; initially number word sequence is in recitation form)].

Other models yet, have focused on the identification of the number of objects in a visual scene as a result of learning. Dehaene and Changeux (1993), for example, using a system consisting of three modules (an input retina, an intermediate topological map of object locations, and a map of detectors) created a numerosity detector. The system was able to simulate the distance effect in counting, by which performance increases with increasing numerical distance between two discriminated quantities. More recently, Ahmad et al. (2002), explored quantification abilities and how they might arise in development, using a multi neural net approach, that combined supervised and un-supervised nets and learning techniques in order to simulate subitization (phenomenon by which subjects appear to produce immediate quantification judgments, usually involving up to 4 objects, without the need to count them) and counting. They used a combined and modular approach, providing a simulation of different cognitive abilities that might be involved in the cognition of number, (each of which would have their own evolutionary history in the brain), and is in keeping with Dehaene's triple code model (2000). Rajapakse et al. (2005), targeted aspects of language related to number such as linguistic quantifiers.

Using a hybrid artificial vision connectionist architecture, they ground linguistic quantifiers such as *few*, *several*, *many,* in perception, taking into consideration contextual factors. Their model, after being trained and tested with experimental data using a dual-route neural network, is able to count objects (fish) in visual scenes and select the quantifier that best describes the scene. Even more recently, Ruciński et al. (2011) using a cognitive robotics paradigm, have explored embodied aspects of mathematical cognition such as the interactions between numbers and space, reproducing three psychological phenomena connected with number processing, namely size and distance effects, the SNARC effect and the Posner-SNARC effect. The same group in another work using the same paradigm (Ruciński et al., 2012), instead focused on counting, and in particular, on the contribution of counting gestures such as pointing. These models, however, did not consider the role of finger counting in numerical abilities.

In this paper, using a Cognitive Developmental Robotics paradigm (Asada et al., 2009; Cangelosi and Schlesinger, 2014) we present results of an exploration on whether finger counting and the association of number words (or tags) to the fingers, could serve to bootstrap the representation of number in a cognitive robot enabling it to perform basic numerical operations, such as addition.

MATERIALS AND METHODS

The robotic model used for the experiments is a computer simulation model of the iCub humanoid robot (Tikhanoff et al., 2008, 2011). The iCub is an open-source humanoid robot platform designed to facilitate cognitive developmental robotics research as detailed in (Metta et al., 2010). At the current state the iCub platform is a child-like humanoid robot 1,05m tall, with 53 degrees of freedom (DoF) distributed in the head, arms, hands and legs. The simulated iCub has been designed to reproduce, as accurately as possible, the physics and the dynamics of the physical iCub. The simulator allows the creation of realistic physical scenarios in which the robot can interact with a virtual environment. Physical constraints and interactions that occur between the environment and the robot are simulated using a software library that provides an accurate simulation of rigid body dynamics and collisions. One of the most advanced parts of the iCub is the hand, that comprises 9 DoF, for a total of 18 DoF, and it is the result of a design that optimized the level of integration of the hand in the overall robot to meet the child-like project specifications in terms of dimensions, dexterity and sensorization. Details on the iCub hand can be found in (Schmitz et al., 2010).

In this work we focus on the fingers, that means we use 7 DoF for each hand, distributed as follows: 2 DoF for thumb, index and middle fingers, but only one for controlling the ring and pinky fingers, that are "glued" together. Because of the limitation with the last two fingers the finger representation of numbers with the right hand is as in **Figure 1**. Numbers from six to ten are represented by adding left hand fingers with all the right hand fingers open (e.g., six is five right hand plus one left hand). In this work, we suppose that the robot is right handed.

The iCub is not provided with ears, so the auditory input (i.e., the number words) was recorded from a child's voice using a standard microphone and stored as a WAVEform audio file format at 22 KHz with lossless compression. From the waveform, we extracted the mel-frequency cepstral coefficients (MFCC) to represent each number word from one to ten, using the Slaney's auditory toolbox 2.0 for MATLAB (1998). MFCC technique combines an auditory filter-bank with a cosine transform to give a rate representation roughly similar to the auditory system (Davis and Mermelstein, 1980).

Figure 2 presents the architecture of the robot's cognitive system, in which the different units and their connections are presented in a schematic form. The lower part of the implemented neural system is directly connected with the robotic platform, and can be summarized in: (i) the motor controller/memory (Motor System and Right/Left Layers), that is able to plan finger movements by setting the finger joints' angles and to memorize the finger number sequence; (ii) an auditory memory (auditory system and auditory layer), that is able to memorize the number words sequence. Upper part of **Figure 2** presents the inner units that are responsible for abstract functions (i.e., not directly connected with the robot), they are the switch/associative layer, that allows the two lower systems to cooperate in order to perform other functions, and the competitive layer classifier we implemented to test the quality of the number learning. After supervised training, it is able to represent the correspondence between numbers from 1 to 10 and the internal representations (i.e., hidden layer activations and/or cepstral coefficients). The role of the competitive layer classifier is to simulate the final processing of the numbers, after a number is correctly classified into its class, the appropriate action can be started, e.g., the production of the corresponding word, of a symbol, the manipulation of an object and so on.

The motor controller/memory was designed using two different RNNs in order to model lateralization when processing numbers, as shown by (Tschentscher et al., 2012). In this way,

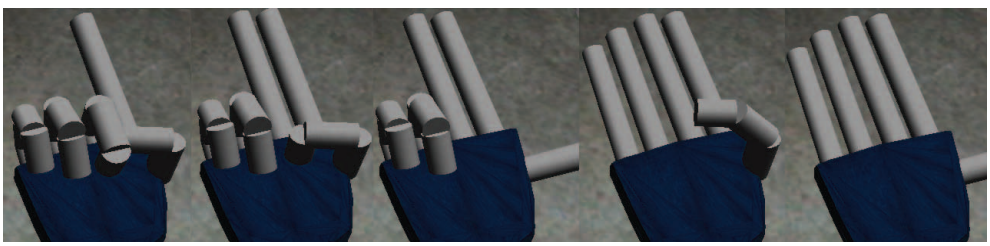


FIGURE 1 | Number representation with the right hand fingers of the iCub. From left to right: one, two, three, four and five.

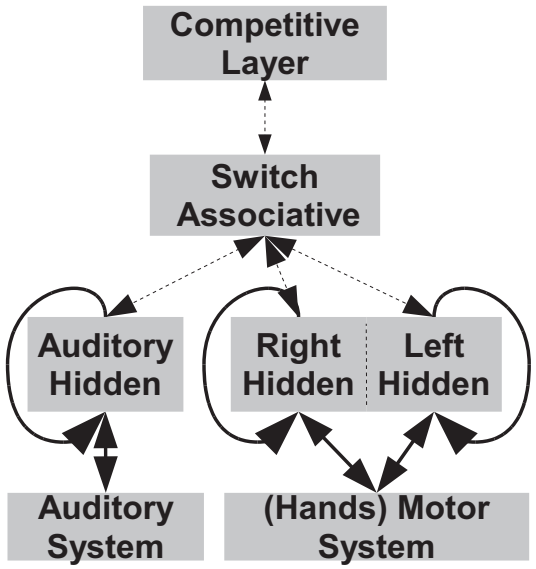


FIGURE 2 | Schematic of the Robot's Cognitive System. In the lower part are the units of the motor controller/memory and of the auditory sub-systems, they are directly connected with the robotic platform. In the upper part there are the units with abstract functions that are the switch/associative network and the competitive layer classifier. Bold links indicate a full (one-to-all) connection between each layer, while dotted links are direct (one-to-one) connections. Note that the system's external inputs coincide with the outputs, indeed proprioceptive information from motor and auditory systems is an input for the system during the training phase, while it is the control output when the system is operating.

the network that controls the left hand will be switched off when low numbers (1–5) are processed. The two RNNs that compose the motor controller/memory were trained separately, i.e., with different random weight initialization. Note that the motor controller is implemented by two different RNNs, trained separately, but that we refer to as a single unit. The use of RNNs to learn to count was investigated by Rodriguez et al. (1999), they explored the capabilities of recurrent networks in the task of learning to predict the next character in a simple deterministic context-free language, in order to provide a more detailed understanding of how dynamics can be harnessed to solve language problems.

The artificial neural networks were implemented using the Matlab Neural Network Toolbox 8.0, the supervised training algorithm for all networks was Levenberg-Marquardt algorithm (LMA), one of the fastest and widely used optimization algorithms that can be applied to artificial neural networks (Hagan and Menhaj, 1994). The LMA interpolates between the Gauss–Newton algorithm (GNA) and the method of gradient descent. The LMA is more robust than the GNA, which means that in many cases it finds a solution even if it starts far from the final minimum. Like the quasi-Newton methods, the LMA was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (as is typical in training feed-forward networks), then the Hessian matrix can be approximated as $\boldsymbol{H} = \boldsymbol{J}^T\boldsymbol{J}$ and the gradient can be computed as $\boldsymbol{g} = \boldsymbol{J}^T\boldsymbol{e}$ where $\boldsymbol{J}$ is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and $\boldsymbol{e}$ is a vector of network errors. In our implementation, the error $\boldsymbol{e}$ is calculated as the average of the squared errors of outputs. The Jacobian matrix can be computed through a standard backpropagation technique (see Hagan and Menhaj, 1994), that is much less complex than computing the Hessian matrix. The LMA uses this approximation to the Hessian matrix in the following Newton-like update:

$$\Delta x = \left[J(x)^T J(x) + \mu I\right]^{-1} J(x)^T e(x)$$

when the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift toward Newton's method as quickly as possible. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function is always reduced at each iteration of the algorithm. In our experiments the initial value of μ was 0.001, increase factor was 10 while decrease factor was 0.1, maximum μ was 10^{10}. The number of iterations (or epochs) of the algorithm was variable because we adopted as stop criterion a minimum performance gradient of 10^{-7} or a maximum of 1000 epochs.

The derivative function of the RNN networks was the back-propagation through time (Rumelhart et al., 1986), that is a gradient based technique that begins by unfolding the recurrent neural network through time into feed-forward neural networks, so that the training then proceeds in a manner similar to training a feed-forward neural network with classic back-propagation, except that each epoch must run through the observations in sequential order.

The competitive layer classifier is implemented using the *softmax* transfer function that gives as output the probability/likelihood of each classification. Naturally, it ensures all of the output values are between 0 and 1, and that their sum is 1. The *softmax* function used is a follows:

$$softmax(q, i) = \frac{e^{q_i}}{\sum_{j=1}^{n} e^{q_j}}$$

where the vector $\boldsymbol{q}$ is the net input to a *softmax* node, and n is the number of nodes in the *softmax* layer.

The architecture of the hidden layers of RNNs was chosen after a performance test, in which after 100 runs with varying number of hidden neurons, the best trade-off solutions were selected in terms of minimization of the error and number of iterations needed to converge. We found that 10 neurons was not surprisingly the ideal solution, this because 10 is also the number of different states to represent. Furthermore, in our preliminary experiments we also found that the pure linear transfer functions for the hidden layers were more effective than the usual sigmoid. We chose not to use a bias or set them to zero for the RNN. Due to these choices, when the networks are not active, i.e., all activations are zero, they can be activated by incepting the activation values to the respective neurons in order to start counting from a specific number.

In addition to the main blocks, an associative network is included in the system to initiate the computation of the system

and to implement the number manipulation. Indeed, after the RNNs have learned the number sequence, the switch is needed to stop the counting and to redirect the signals to the competitive classifier for the processing of the result.

Figure 3 shows the details of the switch/associative layer that, once the two systems have learned to count, allows them to operate and communicate with each other. In particular, the unit is responsible for starting the counting by initializing all the hidden units to 1, and redirecting the hidden unit activation to the competitive classifier when the counting is finished. Furthermore, this unit is crucial in the development for the acquisition of the ability to add numbers, because it can reset one of the two networks to make it count the new operand, and it lets the other continue as a buffer memory. Finally, thanks to the associative connections between the two layers (with weights $\mathbf{w}_1$ and $\mathbf{w}_2$ in **Figure 3**) there are other two states that allow inputting a specific number representation starting from another: from fingers to words and vice versa. These states will be studied in more detail in the number manipulation experiments. All states are reported in the table on the left of **Figure 3**.

As can be seen from the switch/state table in **Figure 3** we set to 1 the initial state of all the hidden layers' neurons in order to start the sequence. Vice versa if the initial state is set to 0, there is no activation because RNNs do not have bias in the hidden layer.

Table 1 presents the actual finger joint positions for the ten number representations plus the rest (zero) position. A high value of the joint position represents the finger when it is closed, while low values indicate the finger is open. Note that because of element collision and tendon limitations the actual values are not the ideal ones (i.e., 90, 180, 220 when finger is closed, 0 when open).

Table 2 reports the MFCCs for the number words extracted from a child voice.

In our experiments, all values in the input/output datasets used in training were pre-processed by dividing them by the maximum absolute value of the series, in order to have them in the range $[-1, 1]$. This is beneficial for the learning of weights and biases of the artificial neural networks.

EXPERIMENTS AND RESULTS

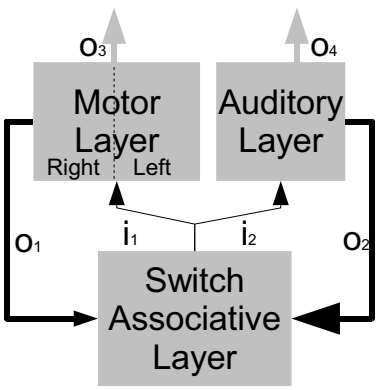
Using the material and methods presented above, first we studied the part of the cognitive system that learns to count. The results of the training are presented in the subsection "Numbers learning." As second step, we build on this by developing the capacity of the associative network to control basic operations like the addition of two operands and to derive the number representation of one of the networks from the other (i.e., from fingers to words and vice versa).

NUMBERS LEARNING

For this first experiment, the main goal was to test the ability of the proposed cognitive system to learn numbers by comparing the performance of different ways of training the number knowledge of the robot with: (1) the internal representation (hidden units activation) of a given finger sequence, (2) the MFCC coefficients of number words out of sequence, (3) the internal representation of the number words sequence, (4) the internal representation of finger sequences plus the MFCC of number words out of sequence (i.e., learning words while counting); (5) internal representations of the sequences of both fingers and number words together (i.e., learning to count with fingers and words).

To this end, we setup the experiment with the following steps: (i) the motor controller learns the opening of the fingers in a given sequence, in order to later establish a finger counting routine, and creates an internal representation for each step in the sequence by means of the hidden units activations; (ii) MFCCs are extracted from number words; (iii) the auditory memory learns the verbal number words in order from 1 to 10 and creates an internal representation for each word in the sequence. From each learning step, relevant data are collected and stored as datasets for the experimentation, these sequences can be summarized as follows:

(1) Internal representations of the finger sequence: 10 values corresponding to the activation values of the hidden units of motor controller/memory network.



s=	$i_1(t)$	$i_2(t)$	$o_1(t+1)$	$o_2(t+1)$	$o_3(t+1)$	$o_4(t+1)$
idle	**0**	**0**	**0**	**0**	*NaN*	*NaN*
start	**1**	**1**	$\Sigma \mathbf{i}_1(t)\times\mathbf{h}_1$	$\Sigma \mathbf{i}_2(t)\times\mathbf{h}_2$	*NaN*	*NaN*
count	$o_1(t)$	$o_2(t)$	$\Sigma \mathbf{i}_1(t)\times\mathbf{h}_1$	$\Sigma \mathbf{i}_2(t)\times\mathbf{h}_2$	*NaN*	*NaN*
done	**0**	**0**	**0**	**0**	$o_1(t)$	$o_2(t)$
sum	$o_1(t)$	**1**	$\Sigma \mathbf{i}_1(t)\times\mathbf{h}_1$	$\Sigma \mathbf{i}_2(t)\times\mathbf{h}_2$	*NaN*	*NaN*
f2w	*NaN*	$o_1(t)\times\mathbf{w}_1$	*NaN*	$\Sigma \mathbf{i}_2(t)\times\mathbf{h}_2$	*NaN*	*NaN*
w2f	$o_2(t)\times\mathbf{w}_2$	*NaN*	$\Sigma \mathbf{i}_1(t)\times\mathbf{h}_1$	*NaN*	*NaN*	*NaN*

× = scalar product; f2w = fingers to words; w2f=words to fingers; NaN preserves the activation at time t (or doesn't activate the units); **w** and **h** are weight vectors, of associative layer and motor/auditory layer respectively.

FIGURE 3 | Details of the Switch/Associative Layer. The table on the right summarizes the outputs according to the different states. In practice the layer operates as a recursive feedback with the possibility to start and reset the motor/auditory layers and to derive the activations of one layer from the ones of the other. Bold lines indicate a full weighted connection, while normal lines are single connections. For simplicity hidden units of the two RNNs of the motor system are represented with one block.

Table 1 | Actual finger joint positions according to number representation.

Fingers	Rest/zero	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten
Thumb right	90.0	90.0	90.0	10.9	88.9	11.1	11.1	11.1	11.1	11.1	11.1
	180.0	180.0	180.0	2.1	88.8	1.2	1.2	1.2	1.2	1.2	1.2
Index right	90.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	180.0	2.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Middle right	90.0	90.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	180.0	180.0	2.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ring and pinky right	220.0	220.0	220.0	220.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0
Thumb left	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	11.0	89.0	11.0
	180.0	180.0	180.0	180.0	180.0	180.0	180.0	180.0	2.2	88.9	1.2
Index left	90.0	90.0	90.0	90.0	90.0	90.0	1.0	0.0	0.0	0.0	0.0
	180.0	180.0	180.0	180.0	180.0	180.0	2.1	0.0	0.0	0.0	0.0
Middle left	90.0	90.0	90.0	90.0	90.0	90.0	90.0	1.0	0.0	0.0	0.0
	180.0	180.0	180.0	180.0	180.0	180.0	180.0	2.1	0.0	0.0	0.0
Ring and pinky left	220.0	220.0	220.0	220.0	220.0	220.0	220.0	220.0	220.0	2.7	0.0

Table 2 | Mel Frequency Cepstral Coefficients for number words.

MFCC	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten
1	−35.5929	−32.9669	−32.4777	−31.2712	−29.7136	−35.4331	−35.442	−32.0295	−31.2157	−31.9479
2	−1.0919	−1.3581	−1.5224	−1.8495	−1.4493	−1.2686	−1.0689	−1.9539	−1.1709	−2.0959
3	0.4216	1.1045	0.6798	0.0099	−0.6858	0.7221	0.6448	0.6668	0.1402	0.5683
4	−0.1042	0.2708	−0.0635	−0.3179	−0.4566	0.184	0.1756	0.2295	0.7539	−0.2115
5	0.3303	0.0268	0.2202	0.0331	0.507	−0.08	−0.2727	−0.0303	−0.4317	−0.2456
6	−0.1156	0.3903	−0.0071	−0.3468	0.3923	0.0328	0.0277	−0.2201	−0.189	0.1962
7	0.0052	0.1658	−0.0939	0.3523	−0.3028	0.1184	0.7074	0.0011	0.9789	0.2468
8	−0.1069	0.0182	0.0204	0.3312	−0.1998	−0.4751	0.036	−0.0101	0.3488	0.4278
9	−0.1343	−0.1744	0.0419	0.1559	−0.1472	0.0975	0.2879	0.1467	−0.2435	0.5773
10	0.1164	−0.1774	0.0226	−0.0661	−0.1542	0.5202	0.1546	0.1802	0.4552	−0.1427
11	−0.5587	0.3471	−0.0303	−0.0531	0.3098	−0.1306	0.155	0.0578	−0.1662	−0.0612
12	−0.1981	−0.0564	0.1463	0.0979	0.2068	−0.0164	−0.2105	0.2783	−0.2708	−0.0456
13	0.223	−0.0566	−0.0506	0.033	−0.0655	0.0037	−0.1311	−0.0695	−0.176	0.0915

(2) MFCCs from number words: 13 values, not as part of a sequence.
(3) Internal representations of the words sequence: 10 values from hidden units' activations of auditory memory network.
(4) Internal motor representations of the finger sequence and MFCCs: a total of 23 values obtained by merging 1 and 2.
(5) Internal motor and auditory representations: a total of 20 inputs obtained by merging 1 and 3.

Datasets 4 and 5 are built to model the learning when both fingers and number words are presented together as training input to the cognitive system.

Figure 4 shows the activation values of hidden layers of RNNs: finger sequences on the left and word sequences on the right. Note that we present together the activations of the two RNNs that compose the motor controller/memory network. Motor activations show a lateralization because the network that controls the left hand (neurons 6–10) is switched off, furthermore the units from 1 to 5 remains fixed from the number five on because we suppose that the right hand is open (we reason as if the robot is right handed). Moreover, in **Figure 5** we present the dendrogram after the optimal leaf order (Bar-Joseph et al., 2001), that shows how the internal finger representation is more similar to the number sequence, indeed, numbers that are close in the actual sequence are linked together. Meanwhile, the grouping of number words (learned in or out of sequence) is more random, and affects the learning as shown in the classification experiment.

All datasets were used to train the competitive layer classifier to be classified in the ten classes that represent the numbers from 1 to 10. Classification results after 10 epochs of training are presented for each class/number in **Table 3**. The low number of epochs is, in this case, is imposed in order to study the robot's number learning in the early stages. However, results show that 10 epochs are enough for the LMA to converge.

Table 3 reports the medians and standard deviations of class calculated after 100 runs for each classification training dataset. If we consider "good" classification only, the cases in which the likelihood is greater than 0.5, we can consider the plain words dataset as not adequate to train the network because it fails for all numbers. However, if we consider as successful classification the cases when the class has the greatest likelihood, the only misclassification observed is for the number three. All the other datasets are

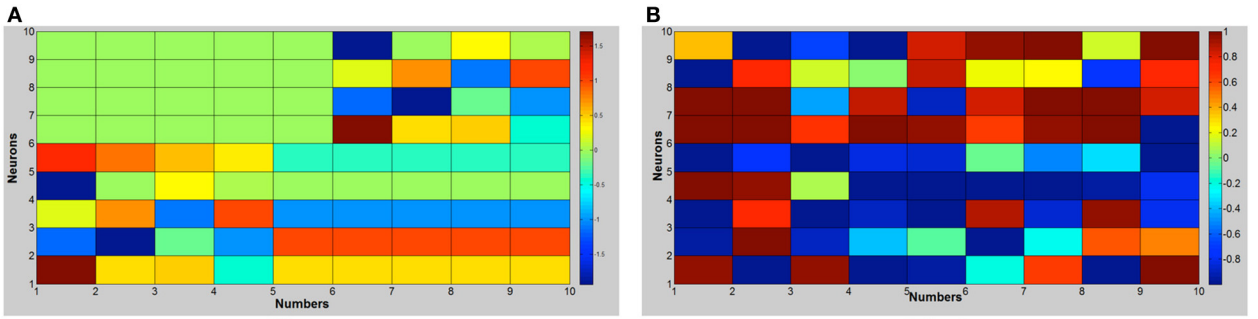


FIGURE 4 | Hidden units' activation with the number sequence from 1 to 10. (A) RNN trained with finger sequence. **(B)** RNN trained with word sequence.

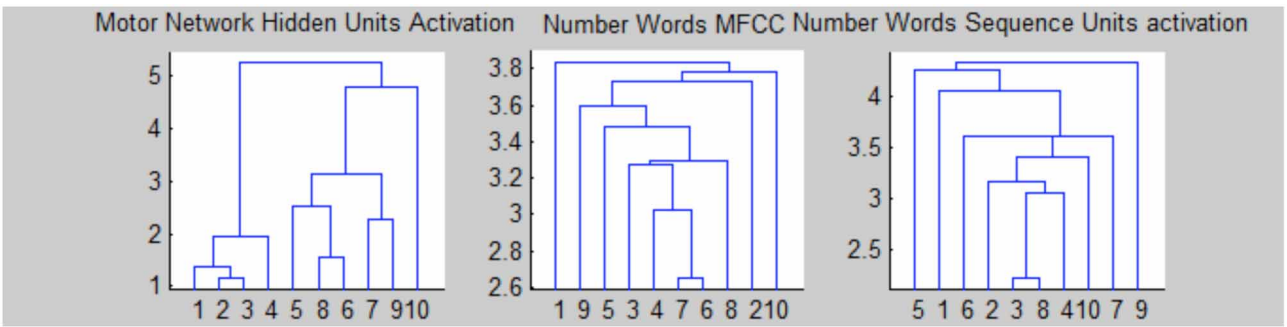


FIGURE 5 | Optimal leaf-order of hidden units' activations.

Table 3 | Likelihood of number classes with different training datasets.

	Fingers sequence only		Number words out of sequence		Words sequence only		Finger sequence and number words		Fingers and words sequences	
	Median	**Std dev**	**Median**	**Std dev**	**Median**	**Std dev**	**Median**	**Std dev**	**Median**	**Std dev**
1	0.779	0.0743	0.3526	0.0218	0.7638	0.0093	0.7355	0.0782	0.9585	0.0108
2	0.662	0.1341	0.2303	0.0083	0.7358	0.0094	0.6584	0.048	0.9621	0.0085
3	0.7015	0.1192	0.0752	0.0043	0.679	0.0145	0.6255	0.0587	0.9459	0.021
4	0.9027	0.1376	0.1777	0.003	0.6414	0.0195	0.7789	0.0677	0.9763	0.1921
5	0.7156	0.0485	0.4448	0.0057	0.6871	0.0116	0.7798	0.0292	0.9148	0.0114
6	0.7574	0.05	0.2411	0.0322	0.6452	0.0161	0.8072	0.0533	0.9202	0.01631
7	0.8226	0.0387	0.2798	0.031	0.6082	0.0308	0.875	0.0229	0.9333	0.01391
8	0.6799	0.0653	0.1499	0.0045	0.6476	0.0198	0.7133	0.0543	0.9125	0.0138
9	0.7854	0.0433	0.3667	0.0058	0.7284	0.0102	0.84	0.0333	0.9443	0.0152
10	0.9069	0.0278	0.2404	0.0058	0.7436	0.0098	0.93	0.0221	0.9708	0.0085
avg	0.7653	0.0738	0.2558	0.0122	0.688	0.0151	0.7743	0.0468	0.9438	0.0311

good and as expected, when finger and word sequences are used together, the cognitive system learns numbers quickly and with a very good likelihood, greater than 90% for all numbers.

We performed a pairwise *t*-test to evaluate the statistical significance of the results reported in **Table 3**. The *t*-test results confirm that all the differences are statistically significant except for the number three, when finger sequences are compared with word sequences, and the two when finger sequences only are compared with finger sequences and number words.

The "finger sequence and number words" (i.e., dataset 4), shows that associating the number words with the fingers sequence helps to drastically improve the classification performance without needing to learn number words in a sequence. However, to learn number words in sequence helps to additionally improve the classification performance to highest likelihood, if internal representations are associated with motor ones.

In order to study in more detail the development of learning, we measured the classification performance over the 10 epochs

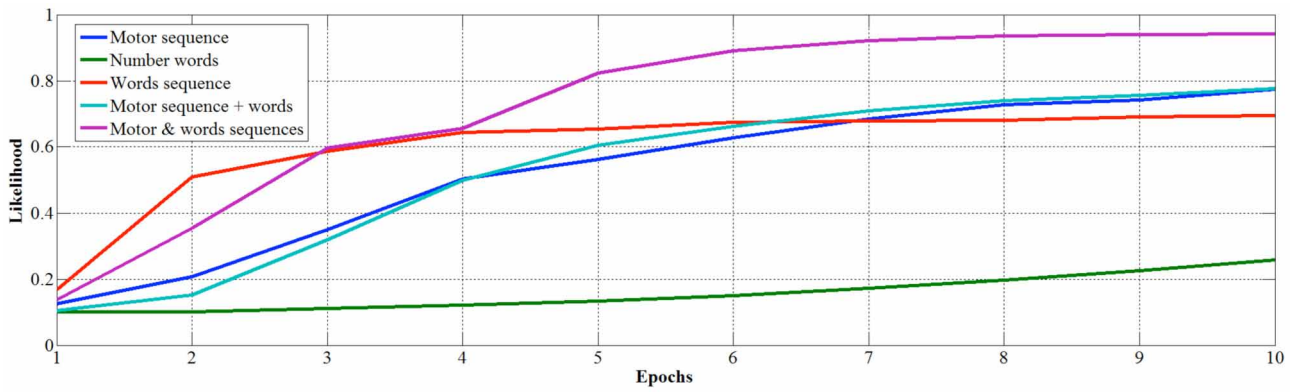


FIGURE 6 | Average likelihood with number classes with varying epochs.

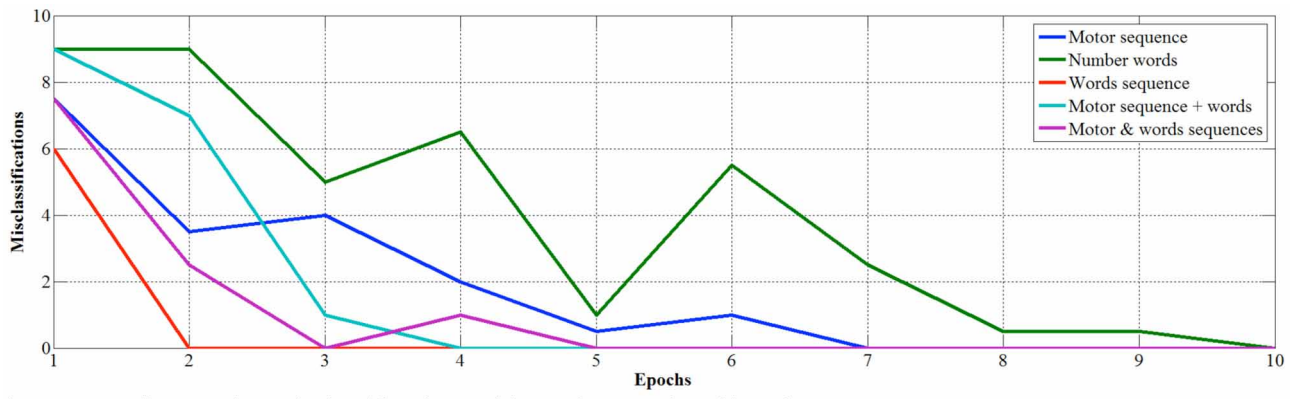


FIGURE 7 | Median number misclassifications with varying epochs of learning.

for the competitive layer trained with the different datasets. In this case, performance is evaluated by means of the average likelihood of classification (**Figure 6**) and median number of misclassifications (**Figure 7**).

Looking at the developmental results, we once again see that number words learned out of a sequence are the less efficient to learn as there are no misclassifications only after 10 epochs, and the average likelihood is still low (0.256) after ten epochs. Conversely, if number words are learned in sequence and internal representations are used as inputs, the learning is faster in terms of precision of classification (i.e., no errors after just 2 epochs) but the maximum average likelihood, that converges at 0.688, is not as strong as when the learning involves also fingers. Indeed, the finger sequence reaches a higher average likelihood (0.765), but best results are obtained when internal representation of words and fingers are used together as input, in fact the average median likelihood is 0.94 just after 8 epochs.

NUMBERS MANIPULATION

Once the number sequences are learned, an interesting feature of the proposed cognitive system is the possibility to easily build up the ability to manipulate numbers with the development of the switch-associative network.

Indeed, this ability can be modeled by extending the capabilities of the associative network from the simple start and stop, to its transferring and mapping to the basic operation of addition.

By transferring, we mean the new mapping of the network's representation derived from the number counted by the other network, when the robot hears the number word "three," to the correlated finger representation. This can be considered, in a sense, an associative mapping between internal representations. This is implemented by activating a weighted connection between the two networks, which can be learned by applying the LMA to the two-layer network that comprises the hidden units of both networks. This training is quite fast and effective both ways, as an average of 4 iterations (over 100 trials) are needed to reach an average estimation error lower than 10^{-15}, which practically does not affect the performance of the classifier, that shows differences in the statistical indicators (mean, median, and standard deviation over 100 trials) lower than 10^{-12} when its inputs are derived by the associative network.

The operation of addition can be seen as a direct development of the concurrent learning of the two recurrent units (motor and auditory). Indeed, if one of the two does the actual counting of the operands, the other can be used as a buffer memory to add the result, when it is done, the final number can be transferred

from the buffer to the other unit and then inputted to the final processor (the classifier in our system).

As an example let us consider 2 + 2, the following steps will be taken:

(1) The first operand is heard by the auditory system and both networks will count until the corresponding activation of number 2 is reached. This step corresponds to the states of the associative network.
(2) The sum operator is recognized so the auditory network is reset, while the first operand remains stored in the motor memory.
(3) The second operand is heard, both networks restart to count as in step 1, until the auditory network reaches the activation corresponding to the number 2. In the meantime, the motor network reaches the activation of the number 4.
(4) After the auditory network stops, the associative network recognizes that the work is done so the total (4) is incepted from the fingers network to the auditory network thanks to the associative connection.
(5) Finally the output of the resulting number (4) is produced for final processing (in our case the classifier).

The steps are depicted in **Figure 8**.

DISCUSSION

The results obtained in our experiments with the iCub child-like robotic platform, show that learning the number words in sequence along with finger configurations helps the fast building of the initial representation of number in the robot. Number knowledge, is instead, not as efficiently developed when number words are learned out of sequence without finger counting. Furthermore, the internal representations of the finger configurations themselves, developed by the robot as a result of the experiments, sustain the execution of basic arithmetic operations,

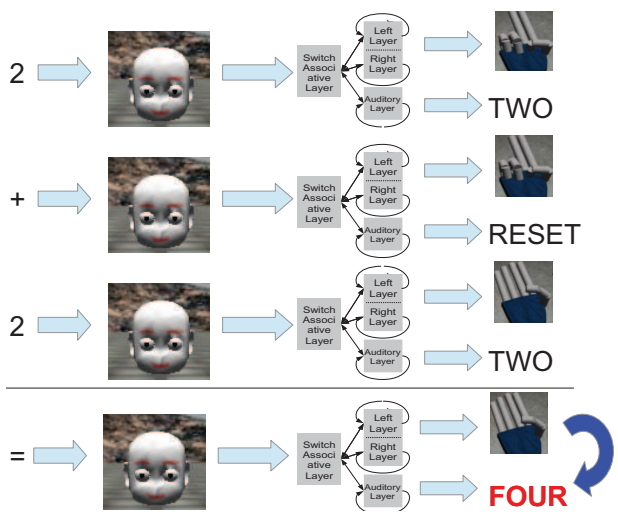


FIGURE 8 | The picture presents the steps done by the Robot's Cognitive System to perform the addition of two numbers, that in this example are both two.

something consistent with evidence coming from developmental research with children.

This does not mean that just learning the counting sequence from one to ten, is enough for children (or our robot), to understand number concepts, but it is the repeated experience using the number word sequence when counting sets of things that is important in the development of numerical understanding (Sarnecka and Carey, 2008; Donlan, 2009). While the use of fingers does not necessarily precede the use of language in the acquisition of a symbolic numerical system (e.g., Nicoladis et al., 2010), what many children seem to be doing initially, in fact, is learning small number word sequences by rote, and later, associations between these small number words and objects in the world (first among which, their readily available fingers). Later on down the developmental path, with the child's early schooling experience, this mapping will also include written representations (or numerals). These written representations, eventually take on the meaning of the spoken number word (Fuson and Kwon, 1992). It is this kind of associative multi-modal learning that we are in a sense reproducing in our model.

Studies focusing on how children acquire abstract words and concepts, have proposed that multiple representational systems involving both sensorimotor as well as linguistic information might be playing a role in conceptual representation (e.g., Louwerse and Jeuniaux, 2010). While the case of the acquisition of number words might be considered as a particular type of abstract word learning, theories such as the LASS theory (Barsalou et al., 2008), according to which both the linguistic system as well as the sensorimotor system (through simulation) are activated in the processing of word meaning to different degrees under different task conditions, and the WAT (Words as Tools) proposal put forth by Borghi and Cimatti (2009) (but also see Borghi et al., 2011, and the special issue of Borghi and Pecher, 2011), have argued and furnished evidence on the synergetic role both language and sensorimotor experience play in the acquisition of abstract concepts, and on how important the modality by which words are learned is. In our model, number words or tags heard repeatedly, when coupled to the experience of moving the fingers, do serve as tools, used in the subsequent manipulation of the quantities they come to represent.

In fact, the internal representations of the finger configurations themselves, found as a result of the experiments, can be considered to be a basis for the building of an embodied number representation in the robot, something in line with embodied and grounded cognition approaches to the study of mathematical cognitive processes. Just as has been found with young children, through the use of finger counting and verbal counting strategies, our model develops finger and word representations that subsequently sustain the robot's learning the basic arithmetic operation of addition. While the experiments done with the model in this work have targeted simple addition, the same model can be easily adapted to implement other operations such as subtraction by training the motor controller/memory with the backward number sequence (from 10 to 1) and then selecting this sequence at the beginning (e.g., by setting all hidden activations to −1), this way the subsequent manipulation will give the result of the subtraction. Future work with the model will implement this.

Another thing we would like to highlight is that, since the hidden layer without any external input does the actual counting, our system is also able to count "mentally" without necessarily producing actions or words. Future work with the model, will explore offline simulation aspects of the motor programs involved in finger based representations of number, or the "mental motor imagery" activated whenever a number and/or number word is encountered, after the robot has learned finger counting and finger calculation early in its training. This direction is stimulated by recent research on the simulation of mental imagery in cognitive systems and robots (see Di Nuovo et al., 2013b), in particular by the successful application of motor imagery models for mental practice in the execution of verbal commands (Di Nuovo et al., 2012) and for performance improvement (Di Nuovo et al., 2013a). The interested reader can find more details about this line of research in a recent special issue (see Di Nuovo et al., 2013b). In recent theoretical accounts of embodied numerosity (e.g., Moeller et al., 2012), something akin to the offline simulation of the motor programs involved in finger based representations of number, has been suggested to take place in children as well as adults. This line of investigation with the model will also compare the representation of embodied numerosity or "manunumeral" representations (Fischer and Brugger, 2011), using different culturally transmitted counting habits or strategies, in order to explore how they might influence the number processing in the robot (e.g., Bender and Beller, 2011; Previtali et al., 2011).

The utility of children's learning finger counting strategies early in their mathematical education continues to be debated in mathematics education research, despite the evidence coming from neurocognitive and psychological studies indicating that it does (for review of debate see Moeller et al., 2011). Our experiments show that in fact, learning to count with the fingers, using verbal tags, can be helpful in the numerical training of a robot as well. While being inspired by the evidence from the studies we have cited in previous sections, our implementation is nonetheless, an abstraction of complex and as of yet not totally understood processes that may underlie the development of numerical cognition. Our results, however, are in line with what has been theoretically claimed in the developmental literature (e.g., Gelman and Gallistel, 1978), that is: that finger counting may be playing a functional role in the acquisition of a variety of principles considered necessary for children to have "under their belts" in order to reach an understanding of number. Examples of these principles and the role of finger counting that are relevant to our present study, and that we think we have simulated at least in part are: finger counting as an aid in the keeping track of the number words while reciting the counting sequence; as it contributing to the induction of the one-to-one correspondence principle by which children are helped by their fingers to coordinate the processes of tagging, or the attribution of a number word to each item; and as facilitating the assimilation of the stable-order principle where numerical labels have to be enumerated in the same order across counting sequences (see also Andres et al., 2008; and Lafay et al., 2013).

The study of mathematical cognitive processes, traditionally considered to be quintessential examples of abstract and symbolic processing, have been assumed to primarily involve the mind rather than the body. Our embodied robot experiments indicate, that aspects of the development of this knowledge can be accounted for not only by way of bodily representations, but also with an artificial network in the place of a mind.

ACKNOWLEDGMENTS

This work was supported in part by the European Commission under Grants n. 288899 (ROBOT-ERA) and n. 288382 (POETICON++) within the Seventh Framework Programme for Research and Technological Development.

REFERENCES

Ahmad, K., Casey, M. C., and Bale, T. (2002). Connectionist simulation of quantification skills. *Connect. Sci.* 14, 1739–1754. doi: 10.1080/09540090208559326

Andres, M., Di Luca, S., and Pesenti, M. (2008). Finger-counting: the missing tool? *Behav. Brain Sci*. 31, 642–643. doi: 10.1017/S0140525X08005578

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., et al. (2009). Cognitive developmental robotics: a survey. *Auton. Ment. Dev.* 1, 12–34. doi: 10.1109/TAMD.2009.2021702

Bar-Joseph, Z., Gifford, D. K., and Jaakkola, T. S. (2001). Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics* 17, S22–S29. doi: 10.1093/bioinformatics/17.suppl_1.S22

Barsalou, L. W., Santos, A., Simmons, W. K., and Wilson, C. D. (2008). "Language and simulation in conceptual processing," in *Symbols, Embodiment, and Meaning*, eds M. De Vega, A. M. Glenberg, and A. C. Graesser (Oxford: Oxford University Press), 245–284. doi: 10.1093/acprof:oso/9780199217274.003.0013

Bender, A., and Beller, S. (2011). Fingers as a tool for counting – naturally fixed or culturally flexible? *Front. Psychol.* 2:256. doi: 10.3389/fpsyg.2011.00256

Bloom, P., and Wynn, K. (1997). Linguistic cues in the acquisition of number words. *J. Child Lang.* 24, 511–533. doi: 10.1017/S0305000997003188

Borghi, A. M., and Cimatti, F. (2009). "Words as tools and the problem of abstract words meanings," in *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, eds N. Taatgen and H. van Rijn (Amsterdam: Cognitive Science Society), 2304–2309.

Borghi, A. M., Flumini, A., Cimatti, F., Marocco, D., and Scorolli, C. (2011). Manipulating objects and telling words: a study on concrete and abstract words acquisition. *Front. Psychol.* 2:15. doi: 10.3389/fpsyg.2011.00015

Borghi, A. M., and Pecher, D. (2011). Introduction to the special topic embodied and grounded cognition. *Front. Psychol.* 2:187. doi: 10.3389/fpsyg.2011.00187

Butterworth, B. (2005). The development of arithmetic abilities. *J. Child Psychol. Psychiatry* 46, 3–18. doi: 10.1111/j.1469-7610.2004.00374.x

Cangelosi, A., and Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots. *Cogn. Sci.* 30, 673–689. doi: 10.1207/s15516709cog0000_72

Cangelosi, A., and Schlesinger, M. (2014). *Developmental Robotics; from Babies to Robots.* Cambridge,MA: MIT Press, Bradford Books.

Costa, A. J., Chagas, P. P., Krinzinger, H., Lonneman, J., Willmes, K., Wood, G., et al. (2011). A hand full of numbers: a role for offloading in arithmetics learning? *Front. Psychol.* 2:368. doi: 10.3389/fpsyg.2011.00368

Davis, S., and Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans. Acoust.* 28, 357–366. doi: 10.1109/TASSP.1980.1163420

Dehaene, S. (2000). "The cognitive neuroscience of numeracy: exploring the cerebral substrate, the development, and the pathologies of number sense," in *Carving Our Destiny: Scientific Research Faces a New Millennium*, eds S. M. Fitzpatrick and J. T. Bruer (Washington, DC: Joseph Henry Press), 41–76.

Dehaene, S., and Changeux, J-P. (1993). Development of elementary numerical abilities: a neuronal model. *J. Cogn. Neurosci.* 5, 390–407. doi: 10.1162/jocn.1993.5.4.390

Dehaene, S., Tzourio, N., Frack, F., Raynaud, L., Cohen, L., Mehler, J., et al. (1996). Cerebral activations during number multiplications and comparisons: a PET study. *Neuropsychologia* 34, 1097–1106. doi: 10.1016/0028-3932(96)00027-9

Di Luca, S., and Pesenti, M. (2011). Finger numeral representations: more than just another symbolic code. *Front. Psychol*. 2:272 doi: 10.3389/fpsyg.2011.00272

Di Nuovo, A. G., Marocco, D., Di Nuovo, S., and Cangelosi, A. (2013a). Autonomous learning in humanoid robotics through mental imagery. *Neural Netw.* 41, 147–155. doi: 10.1016/j.neunet.2012.09.019

Di Nuovo, A., De La Cruz, V. M., and Marocco, D. (2013b). Special issue on artificial mental imagery in cognitive systems and robotics. *Adapt. Behav.* 21, 217–221. doi: 10.1177/1059712313488964

Di Nuovo, A. G., Marocco, D., Cangelosi, A., De La Cruz, V. M., and Di Nuovo, S. (2012). "Mental practice and verbal instructions execution: a cognitive robotics study," in *Proceedings of the 2012 International Joint Conference on Neural Networks* (Brisbane), 1–6. doi: 10.1109/IJCNN.2012.6252751

Domahs, F., Moeller, K., Huber, S., Willmes, K., and Nuerk, H. C. (2010). Embodied numerosity: implicit hand-based representations influence symbolic number processing across cultures. *Cognition* 116, 251–266. doi: 10.1016/j.cognition.2010.05.007

Donlan, C. (2009). "The role of language in mathematical development: typical and atypical pathways," in *Encyclopedia of Language and Literacy Development* (Canadian Language and Literacy Research Network), 1–7. Available online at: http://literacyencyclopedia.ca/pdfs/topic.php?topld=272; retrieved: 06.21.2013

Fischer, M. H., and Brugger, P. (2011). When digits help digits: spatial-numerical associations point to finger counting as prime examples of embodied cognition. *Front. Psychol.* 2:260 doi: 10.3389/fpsyg.2011.00260

Fischer, M. H., Kaufmann, L., and Domahs, F. (2012). Finger counting and numerical cognition. *Front. Psychol.* 3:108 doi: 10.3389/fpsyg.2012.00108

Fuson, K. C. (1988). *Children's Counting and Concepts of Number*. New York, NY: Springer-Verlag.

Fuson, K. C., and Kwon, Y. (1992). "Learning addition and subtraction: effects of number word and other cultural tools," *in Pathways to Number,* eds J. Bideau and C. J. Brainerd (New York, NY: Springer-Verlag), 33–92.

Fuson, K. C., Richards, J., and Briars, D. J. (1982). "The acquisition and elaboration of the number word sequence," *in Children's Logical and Mathematical Cognition,* ed C. G. Brainerd (New York, NY: Springer-Verlag), 33–92.

Gelman, R., and Gallistel, C. R. (1978). *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.

Gleitman, L. (1990). The structural sources of verb meanings. *Lang. Acquis.* 1, 3–55. doi: 10.1207/s15327817la0101_2

Gracia-Bafalluy, M., and Noël, M. P. (2008). Does finger training increase young children's numerical performance? *Cortex* 44, 368–375.

Hagan, M. T., and Menhaj, M. (1994). Training feed-forward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* 5, 989–993. doi: 10.1109/72.329697

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346. doi: 10.1016/0167-2789(90)90087-6

Lafay, A., Thevenot, C., Castel, C., and Fayol, M. (2013). The role of fingers in number processing in young children, *Front. Psychol.* 4:488 doi: 10.3389/fpsyg.2013.00488

Louwerse, M. M., and Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition* 114, 96–104. doi: 10.1016/j.cognition.2009.09.002

Ma, Q., and Hirai, Y. (1989). Modeling the acquisition of counting with an associative network. *Biol. Cybern.* 61, 271–278. doi: 10.1007/BF00203174

Marocco, D., Cangelosi, A., Fischer, K., and Belpaeme, T. (2010). Grounding action words in the sensorimotor interaction with the world: experiments with a simulated iCub humanoid robot. *Front. Neurorobot.* 4:7 doi: 10.3389/fnbot.2010.00007

Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L.. et al. (2010). The iCub humanoid robot: an open-systems platform for research in cognitive development. *Neural Netw.* 23, 1125–1134. doi: 10.1016/j.neunet.2010.08.010

Moeller, K., Fischer, U., Link, T., Wasner, M., Huber, S., and Cress, U. (2012). Learning and development of embodied numerosity. *Cogn. Process.* 13, 271–274. doi: 10.1007/s10339-012-0457-9

Moeller, K., Martignon, L., Wessolowski, S., Engel, J., and Nuerk, H. C. (2011). Effects of finger counting on numerical development – the opposing views of neurocognition and mathematics education, *Front. Psychol.* 2:328 doi: 10.3389/fpsyg.2011.00328

Morse, A. F., Belpaeme, T., Cangelosi, A., and Smith, L. B. (2010). "Thinking with your body: modeling spatial biases in categorization using a real humanoid robot," in *Proceedings of 2010 Annual Meeting of the Cognitive Science Society* (Portland, OR). 1362–1368.

Nguyen, D., and Widrow, B. (1990). "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," in *Proceedings of the International Joint Conference on Neural Networks*, Vol. 3, (San Diego, CA), 21–26.

Nicoladis, E., Pika, S., and Marentette, P. (2010). Are number gestures easier than number words for pre-schoolers? *Cogn. Dev*. 25, 247–261. doi: 10.1016/j.cogdev.2010.04.001

Noël, M. P. (2005). Finger gnosia: a predictor of numerical abilities in children? *Child Neuropsychol.* 11, 413–430. doi: 10.1080/09297040590951550

Pesenti, M., Thioux, M., Seron, X., and de Volder, A. (2000). Neuroanatomical subtrates of arabic number processing, numerical comparison and simple addition: a PET study. *J. Cogn. Neurosci.* 12, 461–479. doi: 10.1162/089892900562273

Piaget, J. (1952). *The Child's Conception Of Number*. London: Routledge and Kegan.

Previtali, P., Rinaldi, L., and Girelli, L. (2011). Nature or nurture in finger counting: a review on the determinants of the direction of number– finger mapping. *Front. Psychol*. 2:363. doi: 10.3389/fpsyg.2011.00363

Pulvermüller, F. (1999). Words in the brain's language. *Behav. Brain Sci*. 22, 253–336. doi: 10.1017/S0140525X9900182X

Rajapakse, R. H., Cangelosi, A., Coventry, K. R., Newstead, S., and Bacon, A. (2005). "Connectionist modeling of linguistic quantifiers," in *Artificial Neural Networks; Formal Models and Their Applications – ICANN 2005, Lecture Notes in Computer Science,* Vol. 3697, eds W. Dutch, J. Kacprzyk, E. Oja, and S. Zadrozny (Springer-Verlag, Berlin Heidelberg), 679–684.

Rodriguez, P., Wiles, J., and Elman, J. L. (1999). A recurrent neural network that learns to count. *Connect. Sci.* 11, 5–40. doi: 10.1080/095400999116340

Ruciński, M., Cangelosi, A., and Belpaeme, T. (2012). "Robotic model of the contribution of gesture to learning to count," in *Proceedings of IEEE ICDL-EpiRob 2012, Joint International Conference on Development and Learning and the International Conference on Epigenetic Robotics* (San Diego, CA).

Ruciński, M., Cangelosi, A., and Belpaeme, T. (2011). "An embodied developmental robotic model of interactions between numbers and space," in *Expanding the Space of Cognitive Science*: *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, eds L. Carlson, C. Hoelscher, and T. F., Shipley (Austin, TX: Cognitive Science Society), 237–242.

Rueckert, L., Lange, N., Partiot, A., Appolonio, I., Litvan, I., Le Bihan, D., et al. (1996). Visualizing cortical activation during mental calculation with functional MRI. *Neuroimage* 3, 97–103. doi: 10.1006/nimg.1996.0011

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation," in *Mit Press Computational Models of Cognition and Perception Series* eds J. L. McClelland and D. E. Rumelhart (Cambridge, MA: MIT Press Cambridge), 318–362.

Sarnecka, B. W., and Carey, S. (2008). How counting represents number: what children must learn and when they learn it. *Cognition* 108, 662–674. doi: 10.1016/j.cognition.2008.05.007

Sato, M., and Lalain, M. (2008). On the relationship between handedness and hand-digit mapping in finger counting. *Cortex* 44, 393–399. doi: 10.1016/j.cortex.2007.08.005

Schmitz, A., Pattacini, U., Nori, F., Natale, L., Metta, G., and Sandini, G. (2010). "Design, realization and sensorization of the dexterous iCub hand," in *Humanoid Robots (Humanoids), 2010 10th IEEE-RAS International Conference on*, (Nashville, Tennesse), 186–191. doi: 10.1109/ICHR.2010.5686825

Slaney, M. (1998). *Auditory toolbox version 2. Interval Research Corporation, (Indiana: Purdue University), 1998–2010.* Technical Report 1998–010.

Stramandoli, F., Cangelosi, A., and Wermter, S. (2012). The grounding of higher order concepts in action and language: a cognitive robotics model. *Neural Netw.* 32, 165–173. doi: 10.1016/j.neunet.2012.02.012

Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L., and Nori, F. (2008). "An open-source simulator for cognitive robotics research: the prototype of the iCub humanoid robot simulator," in *Proceedings of IEEE Workshop on Performance Metrics for Intelligent Systems Workshop (PerMIS08)*, (Gaithersburg, MD).

Tikhanoff, V., Cangelosi, A., and Metta, G. (2011). Language understanding in humanoid robots: iCub simulation experiments. *IEEE Trans. Auton.* 3, 17–29. doi: 10.1109/TAMD.2010.2100390

Tschentscher, N., Hauk, O., Fischer, M. H., and Pulvermüller, F. (2012). You can count on the motor cortex: finger counting habits modulate motor cortex activation evoked by numbers. *Neuroimage* 59, 3139–3148. doi: 10.1016/j.neuroimage.2011.11.037

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cogn. Psychol.* 24, 220–251. doi: 10.1016/0010-0285(92)90008-P

Xu, F., Spelke, E. S., and Goddard, S. (2005). Number sense in human infants. *Dev. Sci.* 8, 88–101. doi: 10.1111/j.1467-7687.2005.00395.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 September 2013; paper pending published: 25 November 2013; accepted: 08 January 2014; published online: 03 February 2014.

Citation: De La Cruz VM, Di Nuovo A, Di Nuovo S and Cangelosi A (2014) Making fingers and words count in a cognitive robot. Front. Behav. Neurosci. ***8****:13. doi: 10.3389/fnbeh.2014.00013*

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 13 September 2013
doi: 10.3389/fnbeh.2013.00120

Supramodal neural processing of abstract information conveyed by speech and gesture

Benjamin Straube[1]*, Yifei He[1,2], Miriam Steines[1], Helge Gebhardt[3], Tilo Kircher[1], Gebhard Sammer[3] and Arne Nagels[1]

[1] Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Marburg, Germany
[2] Department of General Linguistics, Johannes Gutenberg-University Mainz, Mainz, Germany
[3] Cognitive Neuroscience at Centre for Psychiatry, Justus Liebig University Giessen, Giessen, Germany

Edited by:
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA

Reviewed by:
Nashaat Z. Gerges, Medical College of Wisconsin, USA
Yueqiang Xue, The University of Tennessee Health Science Center, USA

***Correspondence:**
Benjamin Straube, Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Rudolf-Bultmann-Str. 8, 35039 Marburg, Germany
e-mail: straubeb@med.uni-marburg.de

Abstractness and modality of interpersonal communication have a considerable impact on comprehension. They are relevant for determining thoughts and constituting internal models of the environment. Whereas concrete object-related information can be represented in mind irrespective of language, abstract concepts require a representation in speech. Consequently, modality-independent processing of abstract information can be expected. Here we investigated the neural correlates of abstractness (abstract vs. concrete) and modality (speech vs. gestures), to identify an abstractness-specific supramodal neural network. During fMRI data acquisition 20 participants were presented with videos of an actor either speaking sentences with an abstract-social [AS] or concrete-object-related content [CS], or performing meaningful abstract-social emblematic [AG] or concrete-object-related tool-use gestures [CG]. Gestures were accompanied by a foreign language to increase the comparability between conditions and to frame the communication context of the gesture videos. Participants performed a content judgment task referring to the person vs. object-relatedness of the utterances. The behavioral data suggest a comparable comprehension of contents communicated by speech or gesture. Furthermore, we found common neural processing for abstract information independent of modality (AS > CS ∩ AG > CG) in a left hemispheric network including the left inferior frontal gyrus (IFG), temporal pole, and medial frontal cortex. Modality specific activations were found in bilateral occipital, parietal, and temporal as well as right inferior frontal brain regions for gesture (G > S) and in left anterior temporal regions and the left angular gyrus for the processing of speech semantics (S > G). These data support the idea that abstract concepts are represented in a supramodal manner. Consequently, gestures referring to abstract concepts are processed in a predominantly left hemispheric language related neural network.

Keywords: gesture, speech, fMRI, abstract semantics, emblematic gestures, tool-use gestures

INTRODUCTION

Human communication is distinctly characterized by the ability to convey abstract concepts such as feeling, evaluations, cultural symbols, or theoretical assumptions. This can be differentiated from references to our physical environment consisting of concrete objects and their relationships to each other. In addition to our language capacity, humans also employ gestures as flexible tool to communicate both concrete and abstract information (Kita et al., 2007; Straube et al., 2011a). The investigation of abstractness and modality of communicated information can deliver important insight into the neural representation of concrete and abstract meaning. However, up to now, evidence about communalities or differences in the neural processing of abstract vs. concrete meaning communicated by speech vs. gesture is missing.

Recently, a hierarchical model of language and thought has been suggested (Perlovsky and Ilin, 2010) which proposes that abstract thinking is impossible without speech (Perlovsky and Ilin, 2013). According to this model, abstract information is processed by a neural language system, regardless of whether speech or gesture is chosen as a tool to convey this information. Following this assumption, concrete object-related information is represented in mind independent of speech and hence in a modality-dependent manner in brain regions sensitive to for example visual or motor information. The latter assumption—at least partly—contradicts existing embodiment theories, which suggest a strong overlap of the sensory-motor and language system in particular with respect to the processing of concrete concepts (Gallese and Lakoff, 2005; Arbib, 2008; Fischer and Zwaan, 2008; D'Ausilio et al., 2009; Pulvermüller and Fadiga, 2010). However, the particular role of the communication modality for the neural representation of abstract as opposed to concrete concepts has not been investigated so far.

The impact of abstractness on speech processing (e.g., Rapp et al., 2004, 2007; Eviatar and Just, 2006; Lee and Dapretto, 2006; Kircher et al., 2007; Mashal et al., 2007, 2009; Shibata et al., 2007;

Schmidt and Seger, 2009; Desai et al., 2011) and on the neural integration of speech and gesture information has been demonstrated in several functional magnetic resonance imaging (fMRI) studies using different experimental approaches (Cornejo et al., 2009; Kircher et al., 2009b; Straube et al., 2009, 2011a, 2013a; Ibáñez et al., 2011). There is converging evidence suggesting that especially the left inferior frontal gyrus (IFG) plays a decisive role in the processing of abstract semantic figurative meaning in speech (Rapp et al., 2004, 2007; Kircher et al., 2007; Shibata et al., 2007). However, results can further differ due to other factors, such as familiarity, imagibility, figurativeness, or processing difficulty (Mashal et al., 2009; Schmidt and Seger, 2009; Cardillo et al., 2010; Schmidt et al., 2010; Diaz et al., 2011).

In contrast to abstract information processing, it has been suggested that concrete information is processed in different brain regions sensitive to the specific information type: e.g., spatial information in the parietal lobe (Ungerleider and Haxby, 1994; Straube et al., 2011c), form or color information in the temporal lobe (Patterson et al., 2007). A similar finding is illustrated by Binder and Desai (2011): by reviewing 38 imaging studies that examined concrete knowledge processing during language comprehension tasks, the authors found that the processing of action-related speech material activates brain regions that are also involved in action execution (see also Hauk et al., 2004; Hauk and Pulvermüller, 2004); similarly, the processing of other concrete speech information such as sound and color all tend to show activations in areas that process these perceptual modalities (Binder and Desai, 2011). In sum, abstract information processing has been shown to recruit a mainly left-lateralized fronto-temporal neural network whereas concrete information comprehension involves rather diverse activation foci, which are primarily related to the corresponding perceptual origin.

In addition to our speech capacity, gesturing is a flexible communicative tool which humans use to communicate both concrete and abstract information via the visual modality. Previous studies on object- or person-related gesture processing have either presented pantomimes of tool or object use, hands grasping for tools or objects (e.g., Decety et al., 1997; Faillenot et al., 1997; Decety and Grèzes, 1999; Grèzes and Decety, 2001; Buxbaum et al., 2005; Filimon et al., 2007; Pierno et al., 2009; Biagi et al., 2010; Davare et al., 2010; Emmorey et al., 2010; Jastorff et al., 2010); or symbolic gestures like "thumbs up" (Nakamura et al., 2004; Molnar-Szakacs et al., 2007; Husain et al., 2009; Xu et al., 2009; Andric et al., 2013). However, few studies directly compared abstract-social (person-related) with concrete-object-related gestures. A previous study demonstrated that the left IFG is involved in the processing of expressive (emotional) in contrast to body referred and isolated (object-related) hand gestures (Lotze et al., 2006). This finding suggests that the left IFG is sensitive for the processing of abstract information irrespective of communication modality (speech or gestures).

In sum, the left IFG represents a sensitive region for abstract information processing in speech or gesture, whereas the brain areas activated by concrete information depend on communication modality and semantic content. However, whether the same neural structures are relevant for the processing of gestures and sentences with an abstract content or gestures and sentences with a concrete content remains unknown.

Common neural networks for the processing of speech and gesture information have been suggested (Willems and Hagoort, 2007), and empirically tested in several recent studies (Xu et al., 2009; Andric and Small, 2012; Straube et al., 2012; Andric et al., 2013). Andric et al. (2013) performed an fMRI study on gesture processing presenting two different kinds of hand actions (emblematic gestures and grasping movements) and speech to their participants. Thus, either emblematic gestures—hand and arm movements conveying social or symbolic meaning (e.g., "thumbs up" for having done a good job)—or grasping movements (e.g., grasping a stapler) not carrying any semantic meaning *per se* were presented. The authors identified two different types of brain responses for the processing of emblematic gestures: the first type was related to the processing of linguistic meaning, the other type corresponded to the processing of hand actions or movements, regardless of the symbolic meaning conveyed. The latter type involved brain responses in parietal and premotor areas in connection with hand movements, whereas meaning bearing information, e.g., emblem and speech, resulted in activations in left lateral temporal and inferior frontal areas. Altogether, different modalities were involved distinguishing the level of mere perceptual recognition and interpretation of socially and culturally relevant emblematic gestures. More importantly, although lacking baseline conditions containing more concrete semantics (either in gesture or speech), the results from this study tentatively imply a common neural network for processing abstract meaning, irrespective of its input modality.

In a similar vein, Xu et al. (2009) investigated the processing of emblems and pantomimes and their corresponding speech utterances via fMRI. Their finding converges with Andric and colleagues imaging results in the sense that both input modalities activated a common, left-lateralized network encompassing inferior frontal and posterior temporal regions. However, although utilizing emblems (abstract) and pantomimes (concrete) as stimuli, the authors did not elaborate on how different levels of semantics (abstract/concrete) are processed via gesture or speech. Moreover, in a recent study from our laboratory, Straube et al. (2012) looked at less conventionalized gesture—iconic gesture, but still found a fronto-temporal network which was responsible for both the processing of gesture and speech semantics. Altogether, the three aforementioned studies unanimously suggest a common fronto-temporal neural network to be responsible for the processing of not only speech but also gesture semantics.

Although tentative proposals regarding a supramodal neural network for speech and gesture semantics have been made (Xu et al., 2009; Straube et al., 2012), it remains unclear how different levels of semantics—either concrete or abstract—are processed differently with respect to the input modalities. To date, no study results on a direct comparison between abstract and concrete semantic information processing with visual (gesture) or auditory (speech) input are available.

As hypothesized above, concrete object-related information might be represented in mind with and/or without speech, whereas abstract information could require/rely on a

representation in speech. Consequently, common processing mechanisms for the processing of speech and gesture semantics can be specifically expected when abstract (in contrast to concrete) information is communicated. Therefore, the current study focused on the neural correlates of abstractness and modality in a communication context. With a factorial manipulation of content (abstract vs. concrete) and communication modality (speech vs. gestures) we wanted to shed light on supramodal neural network properties relevant for the processing of abstract in contrast to concrete information. We tested the following alternative hypotheses: first, if only abstract concepts—activated through speech or gesture in natural communication situations—are processed in a supramodal manner, then we predict consistent neural signatures only for abstract in contrast to concrete contents across different types of communication modality. However, if concrete concepts—activated through speech or gestures—are also represented in a supramodal network, we predict overlapping neural responses for concrete in contrast to abstract contents across modality.

To manipulate abstractness and communication modality we used video clips of an actor either speaking sentences with an abstract-social [AS] or concrete-object-related content [CS], or performing meaningful abstract-social (emblematic) [AG] or concrete-object-related (tool-use) gestures [CG]. Gestures were accompanied by a foreign language (Russian) to increase the comparability between conditions and naturalness of the gesture videos where spoken language frames the communication context. We used emblematic and tool-related gestures to guarantee high comprehensibility of the gestures. During the experiments participants performed a content judgment task referring to the person vs. object-relatedness of the speech and gesture communications to ensure their attention to the semantic information and the adequate comprehension of the corresponding meaning. We hypothesized modality independent activations exclusively for the processing of abstract information (AS > CS ∩ AG > CG) in language-related regions encompassing the left inferior frontal gyrus, the left middle, and superior temporal gyrus (MTG/STG) as well as regions related to social/emotional processing such as the temporal pole, the medial frontal, and anterior cingulate cortex (ACC). In addition, modality specific activations were expected in bilateral occipital, parietal, and temporal brain regions for gesture (G > S) and in left temporal, temporo-parietal, and inferior frontal regions for the processing of speech semantics (S > G).

METHODS

PARTICIPANTS

Twenty healthy subjects (7 females) participated in the study. The mean age of the subjects was 25.4 years (*SD*: 3.42, range: 22.0–35.0). All participants were right handed (Oldfield, 1971), native German speakers and had no knowledge of Russian. All subjects had normal or corrected-to-normal vision, none reported any hearing deficits. Exclusion criteria were a history of relevant medical or psychiatric illness of the participants. All subjects gave written informed consent prior to participation in the study. The study was approved by the local ethics committee.

STIMULUS MATERIAL

Video clips were selected from a large pool of different videos. Some of them have been used in previous fMRI studies, focusing on different aspects of speech and gesture processing (Green et al., 2009; Kircher et al., 2009b; Straube et al., 2009, 2010, 2011a,b, 2012, 2013a,b; Leube et al., 2012; Mainieri et al., 2013). Here, we used emblematic and tool-related gestures and corresponding sentences to guarantee high comprehensibility of the gestures and a strong difference in abstractness between conditions. For the current analysis, 208 (26 videos per condition × 4 conditions × 2 sets) short video clips depicting an actor were used. The actor performed the following conditions: (1) German sentences with an abstract-social content [AS], (2) Russian sentences with abstract-social (emblematic) gestures [AG], (3) German sentences with a concrete-object-related content [CS], and (4) Russian sentences with concrete-object-related (tool-use) gestures [CG] (**Figure 1**). Thus, we presented videos with semantic information only in speech or only in gesture, both of them in either a highly abstract-social or a concrete-object-related version. Additionally, two bimodal meaningful speech-gesture conditions and one meaningless speech-gesture condition have been presented, which are not of interest for the current analysis.

We decided to present gestures accompanied by a foreign language to increase the comparability between conditions and the naturalness of the gesture videos where spoken language frames the communication context. All sentences had a similar grammatical structure (subject—predicate—object) and were translated into Russian for the gesture conditions. Words that sounded similar in each language were avoided. Examples for the German sentences are: "The blacksmith **hammers** on the metal plate" ("Der Schmied hämmert auf die Metallplatte"; CS condition) or "The bishop **exhorts** the believers" ("Der Bischof ermahnt die

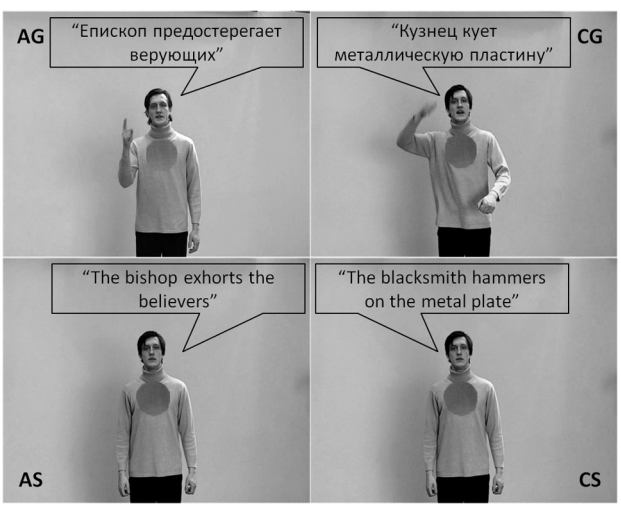


FIGURE 1 | For each of the four conditions (AG, abstract-gesture; CG, concrete-gesture; AS, abstract-speech; CS, concrete-speech) an example of the stimulus material is depicted. Note: For illustrative purposes the spoken German sentences were translated into English and all spoken sentences were written into speech bubbles.

Gläubigen"; AS condition; see **Figure 1**). Thus, the sentences had a similar length of five to eight words and a similar grammatical form, but differed considerably in content. The corresponding gestures (keyword indicated in bold) matched the corresponding speech content, but were presented here only in a foreign language context.

The same male bilingual actor (German and Russian) performed all the utterances and gestures in a natural spontaneous way. Intonation, prosody and movement characteristics in the corresponding variations of one item were closely matched. At the beginning and at the end of each clip the actor stood with arms hanging comfortably. Each clip had a duration of 5 s including 500 ms before and after the experimental manipulation, where the actor neither spoke nor moved. In the present study the semantic aspects of the stimulus material refer to differences in abstractness of the communicated information (abstract vs. concrete content).

For stimulus validation, 20 participants not taking part in the fMRI study rated each video on a scale from 1 to 7 concerning understandability, imageability and naturalness (1 = very low to 7 = very high). In order to assess *understandability* participants were asked: How understandable is the video clip? (original: "Wie VERSTÄNDLICH ist dieser Videoclip?"). The rating scale ranged from 1 = very difficult to understand (sehr schlecht verständlich) to 7 = very easy/good to understand (sehr gut verständlich). For *naturalness* ratings the participants were asked: How natural is the scene? (original: "Wie NATÜRLICH ist diese Szene?"). The rating scale ranged from 1 = very unnatural (sehr unnatürlich) to 7 = very natural (sehr natürlich). Finally, for judgments of *imageability* the participants were asked: How pictorial/imageable is the scene? (original: "Wie BILDHAFT ist dieser Videoclip?"). The rating scale ranged from 1 = very abstract (sehr abstrakt) to 7 = very pictoral/imageable (sehr bildhaft). These scales have been used in previous investigations, too (Green et al., 2009; Kircher et al., 2009b; Straube et al., 2009, 2010, 2011a,b). A set of 338 video clips (52 German sentences with concrete-object-related content, 52 German sentences with abstract-social content and their counterparts in Russian-gesture and German-gesture condition and 26 Russian control condition) were chosen as stimuli for the fMRI experiment on the basis of high naturalness and high understandability for the German and gesture conditions. The stimuli were divided into two sets in order to present each participant with 182 clips during the scanning procedure (26 items per condition), counterbalanced across subjects. A single participant only saw complementary derivatives of one item, i.e., the same sentence or gesture information was only presented once per participant. This was done to avoid speech or gesture repetition or carryover effects. Again, all parameters listed above were used for an equal assignment of the video clips to the two experimental sets, to avoid set-related between-subject differences. As an overview, **Table 1** lists the mean durations of speech and gestures as well as the mean ratings of comprehension, imageability, and naturalness of the items used for the current analyses.

The ratings on understandability for the videos of the four conditions used in this study clearly show a main effect of modality, with the speech varieties scoring higher than the gesture varieties [$F_{(1,\ 113.51)} = 1878.79$, $P < 0.001$, two-factorial

Table 1 | Number of videos and their mean durations of stimulus parameters speech and gesture as well as their mean stimulus ratings of understandability, imageability, and naturalness according to the four conditions abstract-gesture (AG), concrete-gesture (CG), abstract-speech (AS), and concrete-speech (CS) for set 1, set 2 and in total.

Set	Condition	*N*	Stimulus parameter				Rating evaluations					
			Speech duration		Gesture duration		Understandability		Imagebility		Naturalness	
			Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	AG	26	2.163	0.391	2.313	0.440	3.625	0.578	4.498	0.587	4.565	0.379
	CG	26	2.303	0.434	3.033	0.364	3.537	0.808	4.785	0.695	4.340	0.540
	AS	26	2.400	0.308			6.527	0.179	3.481	0.321	4.077	0.258
	CS	26	2.332	0.290			6.650	0.209	2.967	0.308	3.181	0.293
	Total	104	2.299	0.366	2.673	0.540	5.085	1.595	3.933	0.894	4.041	0.649
2	AG	26	2.144	0.296	2.219	0.336	3.392	0.766	4.381	0.698	4.479	0.501
	CG	26	2.160	0.391	2.989	0.415	3.327	0.660	4.598	0.621	4.181	0.444
	AS	26	2.332	0.281			6.490	0.154	3.454	0.372	3.935	0.237
	CS	26	2.274	0.229			6.652	0.155	3.083	0.207	3.181	0.279
	Total	104	2.228	0.311	2.604	0.539	4.965	1.693	3.879	0.810	3.944	0.612
Total	AG	52	2.153	0.343	2.266	0.390	3.509	0.682	4.439	0.641	4.522	0.442
	CG	52	2.231	0.415	3.011	0.387	3.432	0.738	4.691	0.659	4.261	0.496
	AS	52	2.366	0.294			6.509	0.166	3.467	0.344	4.006	0.256
	CS	52	2.303	0.260			6.651	0.182	3.025	0.266	3.181	0.283
	Total	208	2.263	0.340	2.639	0.538	5.025	1.642	3.906	0.851	3.992	0.631

SD, standard deviation.

between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe]. This effect stems from the fact that different languages were used for speech only and gesture with speech conditions. Video clips with German speech scored higher than 6 while Russian speech with gestures videos scored between 3 and 4 (6.58 vs. 3.47, respectively). This difference is in line with the assumption that when presented without the respective sentence context isolated gestures are less meaningful, but even then they still are more or less understandable, which was important for the current study.

Imageability ratings indicated that there were also differences between the conditions concerning their property to evoke mental images. A significant main effect for modality showed that videos consisting of Russian sentences with gesture were evaluated as being better imaginable than videos consisting only of German sentences [4.57 vs. 3.25, respectively; $F_{(1,\ 144.92)} = 349.89$, $P < 0.001$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe]. A significant interaction effect indicated that this difference was even more pronounced for the concrete conditions [$F_{(1,\ 144.92)} = 24.22$, $P < 0.001$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe].

Naturalness ratings showed a main effect for modality as well. Videos including Russian sentences with gestures were evaluated as more natural than videos including German speech [4.39 vs. 3.59, respectively; $F_{(1,\ 160.63)} = 225.65$, $P < 0.001$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe]. There was also a difference in naturalness ratings concerning the abstractness of the included content. Videos depicting concrete content were evaluated as being less natural than videos depicting abstract content [4.26 vs. 3.72, respectively; $F_{(1,\ 160.63)} = 104.48$, $P < 0.001$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe]. Additionally, an interaction effect indicated that videos consisting of German speech with concrete content were evaluated as least natural [$F_{(1,\ 160.63)} = 28.18$, $P < 0.001$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe].

The sentences had an average speech duration of 2263 ms ($SD = 340$ ms), with German sentences being somewhat longer than Russian sentences [2335 vs. 2192 ms, respectively; $F_{(1,\ 180.94)} = 9.51$, $P < 0.05$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe]. The gestures analyzed here had an average gesture duration of 2639 ms ($SD = 538$ ms), with gestures for concrete content being longer than gestures for abstract content [3011 vs. 2266 ms, respectively; $T_{(102)} = 9.78$, $P < 0.001$].

Events for the fMRI statistical analysis were defined in accordance with the bimodal German conditions [compare for example Green et al. (2009); Straube et al. (2012)] as the moment with the highest semantic correspondence between speech and gesture stroke (peak movement): Each sentence contained only one element that could be illustrated, which was intuitively done by the actor. The events occurred on average 2036 ms ($SD = 478$ ms) after the video start and were used for the modulation of events in the event-related fMRI analysis. The use of these predefined integration time points (see Green et al., 2009) for the fMRI data analysis had the advantage that the timing for all conditions of one stimulus was identical since conditions were counterbalanced across subjects. Additionally, speech and gesture duration were used as parameters of no interest on single trial level to control for condition specific differences in these parameters.

EXPERIMENTAL PROCEDURE

During fMRI data acquisition participants were presented with videos of an actor either speaking sentences (S) or performing meaningful gestures (G) with an abstract-social (A) or concrete-object-related (C) content. Gestures were accompanied by an unknown foreign language (Russian). Participants performed a content judgment task referring to the person vs. object-relatedness of the utterances.

fMRI DATA ACQUISITION

All MRI data were acquired on a 3T scanner (Siemens MRT Trio series). Functional images were acquired using a T2-weighted echo planar image sequence ($TR = 2$ s, $TE = 30$ ms, flip angle 90°, slice thickness 4 mm with a 0.36 mm interslice gap, 64×64 matrix, FoV 230 mm, in-plane resolution 3.59×3.59 mm, 30 axial slices orientated parallel to the AC-PC line covering the whole brain). Two runs of 425 volumes were acquired during the experiment. The onset of each trial was synchronized to a scanner pulse.

EXPERIMENTAL DESIGN AND PROCEDURE

An experimental session comprised 182 trials (26 for each condition) and consisted of two 14-min blocks. Each block contained 91 trials with a matched number of items from each condition (13). The stimuli were presented in an event-related design in pseudo-randomized order and counterbalanced across subjects. As described above (stimulus material) across subjects each item was presented in corresponding conditions, but a single participant only saw complementary derivatives of one item, i.e., the same sentence or gesture information was only seen once per participant. Each clip was followed by a gray background with a variable duration of 2154–5846 ms (jitter average: 4000 ms).

Before scanning, each participant received at least six practice trials outside the scanner to ensure comprehensive understanding of the experimental task. Prior to the start of the experiment, the volume of the videos was individually adjusted so that the clips were clearly audible. During scanning, participants were instructed to watch the videos and to indicate via left hand key presses whether the content of the sentence or the gesture referred to objects index finger or interpersonal social information (e.g., feelings, requests, etc.) middle finger. This task enabled us to focus participants' attention to the semantic content of speech and gesture and to investigate comprehension in a rather implicit manner. Performance rates and reaction times were recorded.

MRI DATA ANALYSIS

MR images were analyzed using Statistical Parametric Mapping (SPM8) standard routines and templates (www.fil.ion.ucl.ac.uk). After discarding the first five volumes to minimize T1-saturation effects, all images were spatially and temporally realigned, normalized (resulting voxel size $2 \times 2 \times 2\,\text{mm}^3$), smoothed (8 mm isotropic Gaussian filter) and high-pass filtered (cut-off period 128 s).

Statistical whole-brain analysis was performed in a two-level, mixed-effects procedure. In the first level, single-subject BOLD responses were modeled by a design matrix comprising the onsets of each event within the videos (see stimulus material) of all seven experimental conditions. As additional factor each video phase was modeled as mini-bock with 5 s duration. To control for condition specific differences in speech and gesture duration these stimulus characteristics were used as parameters of no interest on single trial level. The hemodynamic response was modeled by the canonical hemodynamic response function (HRF). Parameter estimate (β-) images for the HRF were calculated for each condition and each subject. Parameter estimates for the four relevant conditions were entered into a within-subject flexible factorial ANOVA.

A Monte Carlo simulation of the brain volume was employed to establish an appropriate voxel contiguity threshold (Slotnick and Schacter, 2004). This correction has the advantage of higher sensitivity to smaller effect sizes, while still correcting for multiple comparisons across the whole brain volume. Assuming an individual voxel type I error of $P < 0.001$, a cluster extent of 50 contiguous resampled voxels was indicated as necessary to correct for multiple voxel comparisons at $P < 0.05$. This cluster threshold (based on the whole brain volume) has been applied to all contrasts. The reported voxel coordinates of activation peaks are located in MNI space. For the anatomical localization, functional data were referenced to probabilistic cytoarchitectonic maps (Eickhoff et al., 2005) and the AAL toolbox (Tzourio-Mazoyer et al., 2002).

CONTRASTS OF INTEREST

The neural processing of abstract information was isolated by computing the difference contrast of abstract-social vs. concrete-object-related sentences [AS > CS] and gestures [AG > CG], whereas the opposite contrasts were applied to reveal brain regions sensitive for the processing of concrete information communicated by speech [CS > AS] and gesture [CG > AG].

In order to find regions that are commonly activated by both processes, contrasts were entered into a conjunction analysis (abstract: [AS > CS ∩ AG > CG]; concrete: [CS > AS ∩ CG > AG]), testing for independently significant effects compared at the same threshold (conjunction null, see Nichols et al., 2005).

The identical approach has been applied to demonstrate the effect of modality by calculating the following conjunctional analyses, for gesture [AG > AS ∩ CG > CS] and for speech semantics [AS > AG ∩ CS > CG].

Finally, interaction analyses were performed ([AS vs. AG] vs. [CS vs. CG]) to explore modality specific effects with regard to the processing of abstract vs. concrete information. Masking procedure has been used to ensure that all interactions are based on significant differences of the first contrast (e.g., [CG > CS] > [AG > AS] inclusively masked by [CG > CS]).

RESULTS

BEHAVIORAL RESULTS

Subjects were instructed to indicate via button press whether the actor in the video described a socially related action or an object-related action. Correct responses and their reaction times were analyzed each with a Two-Way within-subjects ANOVA with the repeated measurement factors modality (gesture vs. speech) and abstractness (abstract vs. social).

Correct responses showed a significant main effect for modality with videos depicting gesture with Russian speech receiving slightly lower scores than videos depicting German speech only [21.8 vs. 22.95 out of 26, respectively; $F_{(1, 19)} = 8.369$, $P < 0.05$, partial-eta-squared = 0.31]. A significant main effect for abstractness clearly indicated that videos describing abstract social content were less often identified correctly than videos showing concrete object-related content [20.3 vs. 24.45 out of 26, respectively; $F_{(1, 19)} = 15.361$, $P < 0.001$, partial-eta-squared = 0.45]. The factors modality and abstractness also showed a modest significant interaction effect on correct responses [$F_{(1, 19)} = 4.572$, $P < 0.05$, partial-eta-squared = 0.19] stemming from the fact that for videos depicting abstract content the difference between gesture with Russian speech and German speech was more pronounced than for videos showing concrete object-related content (**Figure 2A**).

For each participant the median reaction time for each condition was computed from all correct responses of that condition. A significant interaction effect of modality and abstractness [$F_{(1, 19)} = 5.227$, $P < 0.05$, partial-eta-squared = 0.22] indicated that while there was no difference for videos depicting concrete content, participants reacted slightly faster to videos depicting abstract content with gesture and slightly slower to videos of abstract content with German speech (**Figure 2B**).

fMRI RESULTS

Effects of modality

For the effect of gesture in contrast to speech semantics independent of the abstractness [AG > AS ∩ CG > CS] we found activation in bilateral occipital, parietal, and right frontal brain regions (see **Table 2**, and **Figure 3C**, yellow). By contrast, for the processing of speech semantics independent of abstractness [AS > AG ∩ CS > CG] we found activations in the left anterior

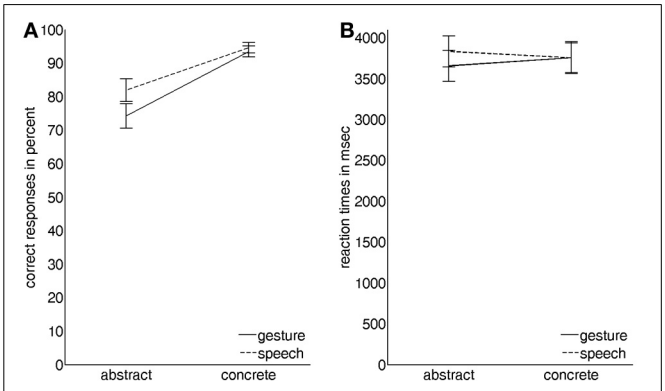


FIGURE 2 | Graphical illustration of the interaction effects of the two factors modality (gesture vs. speech) and abstractness (abstract vs. concrete) on (A) the number of correct responses in percent and on (B) the corresponding reaction times in ms (vertical lines indicate standard errors of the mean).

Table 2 | Activation peaks and anatomical regions comprising activated clusters for the conjunction contrasts representing effects of modality (speech vs. gesture and vice versa).

Contrast	Anatomical regions/hem.	No. voxels	Peak MNI coordinates			*t*-value
			x	*y*	*z*	
AS > AG ∩ CS > CG	Middle temporal gyrus L	673	−52	−12	−20	5.61
	Middle temporal pole L					
	Angular gyrus L	166	−54	−68	34	4.81
	Precuneus L	69	−4	−56	34	3.78
AG > AS ∩ CG > CS	Middle occipital gyrus L	6691	−48	−74	4	19.91
	Inferior temporal gyrus L					
	Middle temporal gyrus R	9536	50	−62	0	19.62
	Fusiform gyrus R					
	Superior occipital gyrus R					
	IFG, pars opercularis R	1313	44	10	28	6.72
	Middle frontal gyrus R					
	Precentral gyrus R					
	Supramarginal gyrus L	202	−62	−36	32	4.56
	Superior parietal lobe L	299	−38	−54	60	4.22
	Inferior parietal lobe L					

Table lists the respective contrast, anatomical regions, cluster size, MNI coordinates, and t-values for each significant activation ($p < 0.05$ corrected for multiple comparisons). MNI, Montreal Neurological Institute; AS, abstract speech; AG, abstract gesture; CS, concrete speech; CG, concrete gesture; IFG, inferior frontal gyrus; L, left; R, right.

temporal lobe and the supramarginal gyrus (see **Table 2**, and **Figure 3D**, yellow).

The exploration of general activation for each condition in contrast to low-level baseline (gray background) indicates that other regions are commonly activated in all conditions (**Figures 3A,B**). Most interestingly, the IFG seems to be activated bilaterally in the gesture conditions (**Figure 3A**) and left lateralized in the speech conditions (**Figure 3B**).

Within modality effects of abstractness

Analyses targeting at within-modality processing of abstractness in language semantics [AS > CS] showed activation in a mainly left-lateralized network encompassing an extended fronto-temporal cluster (IFG, precentral gyrus, middle, inferior, and superior temporal gyrus) as well as medial frontal regions and the right anterior middle temporal gyrus (**Table 3** and **Figure 4** top, blue). We obtained a comparable activation pattern for the within-modality processing of abstractness in gesture semantics ([AG > CG] see **Figure 4** top, yellow). The opposite contrasts revealed activation in clusters encompassing the left cerebellum, fusiform, and inferior temporal gyrus in the language contrast (CS > AS; see **Figure 4** bottom, blue) and the bilateral occipital lobe for the gesture contrast (CG > AG; see **Figure 4** bottom, yellow).

Common activations for abstractness contained in gestures and spoken language

Processing of abstract information independent of input modality as disclosed by the conjunction of [AS > CS ∩ AG > CG] was related to a left-sided frontal cluster including the temporal pole, the IFG (pars triangularis and orbitalis), the middle temporal and angular as well as the medial superior frontal gyrus (**Table 3** and **Figure 4** top middle/right, green). The opposite conjunction analyses [CS > AS ∩ CG > AG] revealed no significant common activation for the processing of concrete in contrast to abstract information.

Interaction

No significant activation could be identified in the interaction analyses on the selected significance threshold. However, by applying a different cluster size to voxel level threshold proportion to correct for multiple comparisons ($p < 0.005$ and 86 voxels) as indicated by an additional Monte Carlo simulation, we found an interaction in occipital (MNI *x, y, z*: −20, −90, −8, $t = 3.63$, $p < 0.001$, 140 voxels), parietal (MNI *x, y, z*: −34, −48, 68, $t = 3.80$, $p < 0.001$, 143 voxels; MNI *x, y, z*: −34, −40, 48, $t = 3.11$, $p < 0.001$, 88 voxels) and premotor (MNI *x, y, z*: −34, −4, 62, $t = 3.55$, $p < 0.001$, 129 voxels) regions reflecting an specific increase of activation in these regions for the processing of concrete-object-related gesture meaning ([CG > CS] > [AG > AS] inclusively masked by [CG > CS]).

DISCUSSION

We hypothesized that the processing of abstract semantic information of spoken language and symbolic emblematic gestures is based on a common neural network. Our study design tailored the comparison to the level of abstract semantics, controlling for processing of general semantic meaning of speech and gesture by using highly meaningful concrete object-related information as control condition. The results demonstrate that the pathways engaged in the processing of semantics contained in both abstract spoken language and

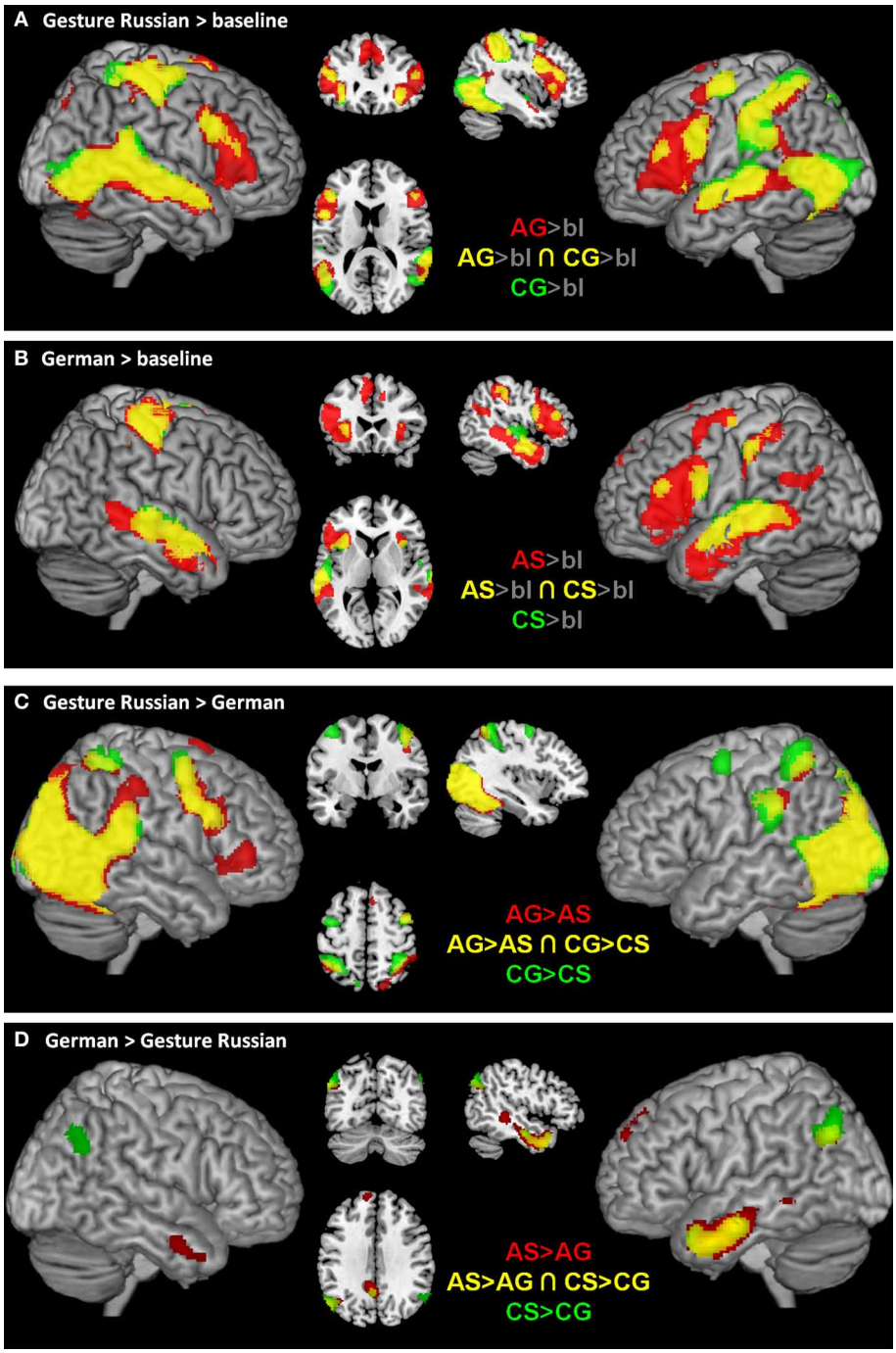


FIGURE 3 | Illustrates the fMRI results for abstract semantics (red), concrete semantics (green), and common neural structures (yellow) for each condition in contrast to low-level baseline (gray background; A, Gesture; B, German), for gesture conditions in contrast to German conditions (C) and for German in contrast to the gesture conditions (D). Results were rendered on brain slices and surface using the MRIcron toolbox (http://www.mccauslandcenter.sc.edu/mricro/mricron/install.html).

abstract-social gestures comprise the temporal pole, the IFG (pars triangularis and orbitalis), the middle temporal, angular and the superior frontal gyri. Thus, in line with our hypothesis we found modality-independent activation in a left hemispheric fronto-temporal network for the processing of abstract information. The strongly left lateralized activation pattern supports the theory that abstract semantics is independent of communication modality represented in language (at least on neural level represented in language-related brain regions).

Table 3 | Activation peaks and anatomical regions comprising activated clusters for the contrasts representing effects of abstractness (abstract vs. concrete and vice versa) dependent of modality (speech or gesture).

Contrast	Anatomical regions/hem.	No. Voxels	Peak MNI coordinates			*t*-value
			x	*y*	*z*	
AS > CS	Middle temporal gyrus L	3150	−52	−34	−6	5.98
	IFG, pars orbitalis L					
	Medial superior frontal gyrus L	1441	−8	56	34	5.72
	Middle temporal pole R	289	48	12	−34	5.16
	Middle temporal gyrus R					
	Angular gyrus L	458	−42	−58	24	4.41
	Precentral gyrus L	248	−38	0	62	4.36
	Precuneus L	195	−8	−50	34	4.22
AG > CG	Superior temporal pole L	910	−36	18	−24	5.43
	IFG, pars triangularis L					
	IFG, pars orbitalis L					
	Medial superior frontal gyrus L	3215	−4	30	54	5.10
	Angular gyrus L	682	−60	−60	30	4.64
	Caudate nucleus R	786	12	2	8	4.54
	Thalamus L					
	Middle temporal gyrus L	209	−48	−16	−18	4.04
AS > CS ∩ AG > CG	Medial superior frontal gyrus L	1015	−8	56	30	4.97
	Superior temporal pole L	779	−36	18	−22	4.93
	IFG, pars triangularis L					
	IFG, pars orbitalis L					
	Middle temporal gyrus L	161	−48	−14	−20	3.99
	Angular gyrus L	253	−54	−56	26	3.95
CS > AS	Cerebellum L	580	−32	−36	−28	5.95
	Inferior temporal gyrus L					
	Fusiform gyrus L					
CG > AG	Middle occipital gyrus L	1046	−44	−76	8	5.86
	Middle temporal gyrus R	285	50	−62	2	4.73
CS > AS ∩ CG > AG						n.s.

Table lists the respective contrast, anatomical regions, cluster size, MNI coordinates, and t-values for each significant activation ($p < 0.05$ corrected for multiple comparisons). MNI, Montreal Neurological Institute; AS, abstract speech; AG, abstract gesture; CS, concrete speech; CG, concrete gesture; IFG, inferior frontal gyrus; L, left; R, right.

EFFECTS OF MODALITY

The results of the speech [CS > CG ∩ AS > AG] and gesture contrasts [CG > CS ∩ AG > AS] clearly demonstrate that communication modality affects neural processing in the brain independent of the communication content (abstract/concrete). In line with other studies that contrasted the processing of a native against an unknown foreign language (Perani et al., 1996; Schlosser et al., 1998; Pallier et al., 2003; Straube et al., 2012), we found activation along the left temporal lobe (including STG, MTG, and ITG) for German speech contrasted with Russian speech and gesture. This strongly left-lateralized pattern has been found in all of the above mentioned studies. Apart from these studies with conditions very similar to ours, temporal as well as inferior frontal regions have been frequently implicated in various language tasks (for reviews see Bookheimer, 2002; Vigneau et al., 2006; Price, 2010). The lack of IFG activation in our study is probably dependent on the fact that we compared a native language (CS, AS) with a foreign language which was accompanied by a meaningful gesture (CG, AG). Thus, motoric or semantic processes of the left IFG might be equally involved in the speech and gesture conditions as indicated by baseline contrasts (see **Figures 3A,B**).

In line with studies on action observation (e.g., Decety et al., 1997; Decety and Grèzes, 1999; Grèzes and Decety, 2001; Filimon et al., 2007) and co-verbal gesture processing (e.g., Green et al., 2009; Kircher et al., 2009b; Straube et al., 2011a), we found for the processing of gesture in contrast to speech information a bilaterally distributed network of activation including occipital, parietal, posterior temporal, and right frontal brain regions.

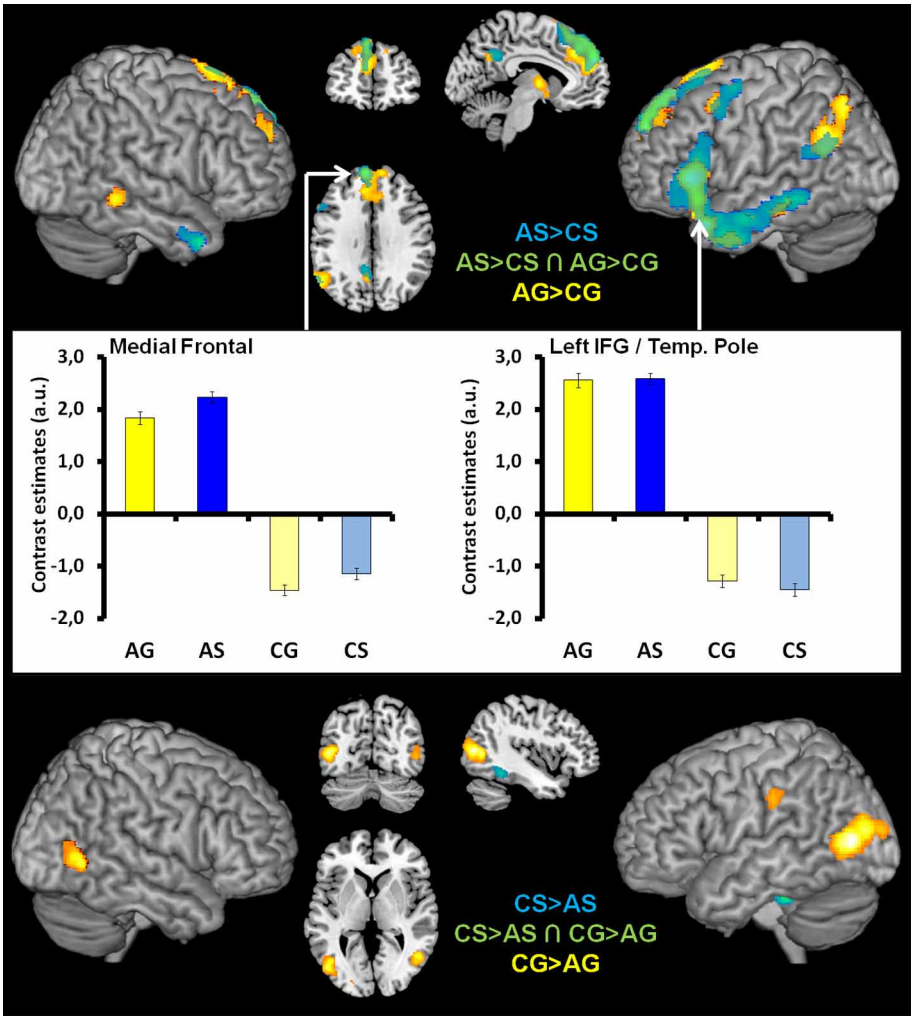


FIGURE 4 | Top illustrates the within-modality processing of abstractness in language semantics ([AS > CS], blue), gesture semantics ([AG > CG], yellow), and in common neural structures (green, overlapping regions). Bar graphs in the **middle** of figure illustrate the contrast estimates (extracted eigenvariates) for the commonly activated (green) medial superior frontal (left) and temporal pole/IFG cluster (right). These are representative for all overlapping activation clusters. The within-modality processing of concrete in contrast to abstract language semantics ([CS > AS], blue) and gesture semantics ([CG > AG], yellow) is illustrated at the **bottom** of figure. Here we found no overlap between activation patterns.

SUPRAMODAL PROCESSING OF ABSTRACT SEMANTICS OF SPEECH AND GESTURE

The processing of abstract spoken language semantics (AS > CS) and abstract semantic information conveyed through abstract-social in contrast to concrete-object-related gestures (AG > CG) activated an overlapping network of brain regions. These include a cluster in the left inferior frontal cortex (BA 44, 45) which expanded into the temporal pole, the left inferior, and middle temporal gyrus as well as a cluster in the left medial superior frontal gyrus. Those findings support the model of a supramodal semantic network for the processing of abstract information. By contrast, for concrete vs. abstract information we obtained no overlapping activation.

These results extend studies from both the gesture and the language domain (see above) in showing a common neural representation of specific speech and gesture semantics. Furthermore, the findings go beyond previous reports about common activation for symbolic gestures and speech semantics (Xu et al., 2009), in showing a specific effects for abstract but not concrete speech and gesture information. Interestingly, we previously found similar activation of the left IFG and temporal brain regions for the processing of concrete speech and gesture semantics of iconic gestures (Straube et al., 2012). Whereas iconic gestures are not symbolic and usually occur in a concrete sentence context (e.g., "The ball is round," using both hands to indicate a round shape), they might implicate rather abstract information without speech, since any concrete meaning can be revealed from these iconic

gestures in this context. Thus, the left IFG activation in our previous study could also be explained by an abstract interpretation of isolated iconic gestures (Straube et al., 2012).

The left-lateralization of our findings is congruent with the majority of fMRI studies on language (see Bookheimer, 2002; Price, 2010, for reviews). Left fronto-temporal activations have been frequently observed for semantic processing [e.g., Gaillard et al., 2004; for a review see Vigneau et al. (2006)], the decoding of meaningful actions (e.g., Decety et al., 1997; Grèzes and Decety, 2001) and also with regard to co-verbal gesture processing (Willems et al., 2007, 2009; Holle et al., 2008, 2010; Kircher et al., 2009b; Straube et al., 2011a).

With regard to the inferior frontal activations, functional imaging studies have underlined the importance of this region in the processing of language semantics. The junction of the precentral gyrus and the pars opercularis of the left IFG has been involved in controlled semantic retrieval (Thompson-Schill et al., 1997; Wiggs et al., 1999; Wagner et al., 2001), semantic priming (Sachs et al., 2008a,b, 2011; Kircher et al., 2009a; Sass et al., 2009a,b) and a supramodal network for semantic processing of words and pictures (Kircher et al., 2009a). The middle frontal gyrus (MFG) was found activated by intramodal semantic priming (e.g., Tivarus et al., 2006). However, medial frontal activation in our study might be better explained by differences in social-emotional content between conditions, which have been often found for social functioning, social cognition, theory of mind, or mentalizing (e.g., Uchiyama et al., 2006, 2012; Krach et al., 2009; Straube et al., 2010).

Since semantic memory represents the basis of semantic processing, an amodal semantic memory (Patterson et al., 2007) is a likely explanation for how speech and gesture semantics could activate a common neural network. Our findings suggest supramodal semantic processing in regions including the left temporal pole, which has been described as best candidate for a supramodal semantic "hub" (Patterson et al., 2007). Thus, abstract semantic information contained in speech and gestures might have activated supramodal semantic knowledge in our study more strongly than concrete information communicated by speech and gesture.

Our data also partially coincide with Binder and Desai's (2011) neuroanatomical model of semantic processing: in this model, low level (concrete) sensory, action and emotion semantics are processed in brain areas that are located near corresponding perceptual networks; higher-level semantics (abstract semantics), on the contrary, converges at temporal, and inferior parietal regions (Binder and Desai, 2011). Additionally, as a next step, inferior prefrontal cortices are responsible for the selection of the information stored in temporo-parietal cortices. In the current experiment, abstract information activates both temporal and inferior frontal cortices, and this could be considered as evidence supporting the role of fronto-temporal pathways in the processing of higher-level semantics. More importantly, our results suggest that this processing of abstract information is independent of input modality.

As for the processing of concrete semantics, our results are somewhat surprising because we did not find an overlap between gestural and verbal-auditory input. This result falls beyond the prediction of both strict embodiment theories (Barsalou, 1999; Gallese and Lakoff, 2005; Pulvermüller and Fadiga, 2010) and theories which propose less strict embodiment: all these theories would predict that the concrete semantics in our experiment, being predominantly action-driven, would activate motoric brain regions such as (pre-)motor and parietal cortices, and this activation pattern should be independent of the input modality. However, previous support for these theories is based on studies using single words (e.g., Willems et al., 2010; Moseley et al., 2012) instead of sentences, which might increase the task effort and specifically trigger motoring simulation. Thus, one explanation for the discrepancy between studies could be that we investigated the processing of tool-use information in a sentence context (see Tremblay and Small, 2011). Here, motoric simulation might not be necessary since contextual information facilitates semantic access (e.g., the blacksmith primes the hammer).

Our results are also in line with a recent mathematically-motivated language-cognition model proposed by Perlovsky and Ilin (2013). This model suggests that high-level abstract thinking relies on the language system and low-level and concrete thinking does not necessarily have to. Transferred to a neural perspective, both abstract meaning (irrespective of input modality) and language (processing) would recruit similar neural networks. In our experiment, the left-lateralized network for abstract meaning comprehension fits perfectly to this prediction. Although it still remains unclear how language and higher-level thinking are related at a functional level, our study provides initial neural evidence, which closely connects the two different domains.

CONCLUSION

Language is not only a communication device, but also a fundamental part of cognition and learning concepts, especially with respect to abstract concepts (Perlovsky and Ilin, 2013). In the last years the understanding of speech and gesture processing has increased; both communication channels have been disentangled and brought together again. Here we investigated the neural correlates of abstractness (abstract vs. concrete) and modality (speech vs. gestures), to demonstrate the existence of an abstractness specific supramodal neural network.

In fact, we could demonstrate the activation of a supramodal network for abstract speech and abstract gestures semantics. The identified left lateralized fronto-temporal network not only maps sound patterns and their corresponding abstract meanings in the auditory domain, but also combines gestures and their abstract meanings in the gestural-visual domain. This modality-independent network most likely gets input from modality-specific areas in the superior temporal (speech) and occipito-temporal brain regions (gestures), where the main characteristics of the spoken and gestured signals are decoded. The inferior frontal regions are responsible for the process of selection and integration, relying on more general world knowledge distributed throughout the brain (Xu et al., 2009). The challenge for future studies will be the identification of specific aspects of speech and gesture semantics or the respective format relevant for the understanding of natural receptive and productive communicative behavior and its dysfunctions in patients, for

example with schizophrenia or autism (Hubbard et al., 2012; Straube et al., 2013a,b).

ACKNOWLEDGMENTS

This research project is supported by a grant from the "Von Behring-Röntgen-Stiftung" (project no. 59-0002) and by the "Deutsche Forschungsgemeinschaft" (project no. DFG: Ki 588/6-1). Yifei He and Helge Gebhardt are supported by the "Von Behring-Röntgen-Stiftung" (project no. 59-0002). Arne Nagels and Miriam Steines are supported by the DFG (project no. Ki 588/6-1). Benjamin Straube is supported by the BMBF (project no. 01GV0615).

REFERENCES

Andric, M., and Small, S. L. (2012). Gesture's neural language. *Front. Psychol.* 3:99. doi: 10.3389/fpsyg.2012.00099

Andric, M., Solodkin, A., Buccino, G., Goldin-Meadow, S., Rizzolatti, G., and Small, S. L. (2013). Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia* 51, 1619–1629. doi: 10.1016/j.neuropsychologia.2013.03.022

Arbib, M. A. (2008). From grasp to language: embodied concepts and the challenge of abstraction. *J. Physiol. Paris* 102, 4–20. doi: 10.1016/j.jphysparis.2008.03.001

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–609. discussion: 610–660. doi: 10.1017/S0140525X99002149

Biagi, L., Cioni, G., Fogassi, L., Guzzetta, A., and Tosetti, M. (2010). Anterior intraparietal cortex codes complexity of observed hand movements. *Brain Res. Bull.* 81, 434–440. doi: 10.1016/j.brainresbull.2009.12.002

Binder, J. R., and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn. Sci.* 15, 527–536. doi: 10.1016/j.tics.2011.10.001

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.* 25, 151–188. doi: 10.1146/annurev.neuro.25.112701.142946

Buxbaum, L. J., Kyle, K. M., and Menon, R. (2005). On beyond mirror neurons: internal representations subserving imitation and recognition of skilled object-related actions in humans. *Brain Res. Cogn. Brain Res.* 25, 226–239. doi: 10.1016/j.cogbrainres.2005.05.014

Cardillo, E. R., Schmidt, G. L., Kranjec, A., and Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behav. Res. Methods* 42, 651–664. doi: 10.3758/BRM.42.3.651

Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., López, V., et al. (2009). Gesture and metaphor comprehension: electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain Cogn.* 70, 42–52. doi: 10.1016/j.bandc.2008.12.005

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385. doi: 10.1016/j.cub.2009.01.017

Davare, M., Rothwell, J. C., and Lemon, R. N. (2010). Causal connectivity between the human anterior intraparietal area and premotor cortex during grasp. *Curr. Biol.* 20, 176–181. doi: 10.1016/j.cub.2009.11.063

Decety, J., and Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends Cogn. Sci.* 3, 172–178. doi: 10.1016/S1364-6613(99)01312-1

Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain* 120(Pt 10), 1763–1777. doi: 10.1093/brain/120.10.1763

Desai, R. H., Binder, J. R., Conant, L. L., Mano, Q. R., and Seidenberg, M. S. (2011). The neural career of sensory-motor metaphors. *J. Cogn. Neurosci.* 23, 2376–2386. doi: 10.1162/jocn.2010.21596

Diaz, M. T., Barrett, K. T., and Hogstrom, L. J. (2011). The influence of sentence novelty and figurativeness on brain activity. *Neuropsychologia* 49, 320–330. doi: 10.1016/j.neuropsychologia.2010.12.004

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Emmorey, K., Xu, J., Gannon, P., Goldin-Meadow, S., and Braun, A. (2010). CNS activation and regional connectivity during pantomime observation: no engagement of the mirror neuron system for deaf signers. *Neuroimage* 49, 994–1005. doi: 10.1016/j.neuroimage.2009.08.001

Eviatar, Z., and Just, M. A. (2006). Brain correlates of discourse processing: an fMRI investigation of irony and conventional metaphor comprehension. *Neuropsychologia* 44, 2348–2359. doi: 10.1016/j.neuropsychologia.2006.05.007

Faillenot, I., Toni, I., Decety, J., Grégoire, M. C., and Jeannerod, M. (1997). Visual pathways for object-oriented action and object recognition: functional anatomy with PET. *Cereb. Cortex* 7, 77–85. doi: 10.1093/cercor/7.1.77

Filimon, F., Nelson, J. D., Hagler, D. J., and Sereno, M. I. (2007). Human cortical representations for reaching: mirror neurons for execution, observation, and imagery. *Neuroimage* 37, 1315–1328. doi: 10.1016/j.neuroimage.2007.06.008

Fischer, M. H., and Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. *Q. J. Exp. Psychol. (Hove)* 61, 825–850. doi: 10.1080/17470210701623605

Gaillard, W. D., Balsamo, L., Xu, B., McKinney, C., Papero, P. H., Weinstein, S., et al. (2004). fMRI language task panel improves determination of language dominance. *Neurology* 63, 1403–1408. doi: 10.1212/01.WNL.0000141852.65175.A7

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310

Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., et al. (2009). Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. *Hum. Brain Mapp.* 30, 3309–3324. doi: 10.1002/hbm.20753

Grèzes, J., and Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp.* 12, 1–19. doi: 10.1002/1097-0193(200101)12:1<1::AID-HBM10>3.0.CO;2-V

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307. doi: 10.1016/S0896-6273(03)00838-9

Hauk, O., and Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Hum. Brain Mapp.* 21, 191–201. doi: 10.1002/hbm.10157

Holle, H., Gunter, T. C., Ruschemeyer, S. A., Hennenlotter, A., and Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage* 39, 2010–2024. doi: 10.1016/j.neuroimage.2007.10.055

Holle, H., Obleser, J., Rueschemeyer, S. A., and Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage* 49, 875–884. doi: 10.1016/j.neuroimage.2009.08.058

Hubbard, A. L., McNealy, K., Scott-Van Zeeland, A. A., Callan, D. E., Bookheimer, S. Y., and Dapretto, M. (2012). Altered integration of speech and gesture in children with autism spectrum disorders. *Brain Behav.* 2, 606–619. doi: 10.1002/brb3.81

Husain, F. T., Patkin, D. J., Thai-Van, H., Braun, A. R., and Horwitz, B. (2009). Distinguishing the processing of gestures from signs in deaf individuals: an fMRI study. *Brain Res.* 1276, 140–150. doi: 10.1016/j.brainres.2009.04.034

Ibáñez, A., Toro, P., Cornejo, C., Urquina, H., Hurquina, H., Manes, F., et al. (2011). High contextual sensitivity of metaphorical expressions and gesture blending: a video event-related potential design. *Psychiatry Res.* 191, 68–75. doi: 10.1016/j.pscychresns.2010.08.008

Jastorff, J., Begliomini, C., Fabbri-Destro, M., Rizzolatti, G., and Orban, G. A. (2010). Coding observed motor acts: different organizational principles in the parietal and premotor cortex of humans. *J. Neurophysiol.* 104, 128–140. doi: 10.1152/jn.00254.2010

Kircher, T., Sass, K., Sachs, O., and Krach, S. (2009a). Priming words with pictures: neural correlates of semantic associations in a cross-modal priming task using fMRI. *Hum. Brain Mapp.* 30, 4116–4128. doi: 10.1002/hbm.20833

Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., et al. (2009b). Neural interaction of speech and gesture: differential activations of metaphoric coverbal gestures. *Neuropsychologia* 47, 169–179. doi: 10.1016/j.neuropsychologia.2008.08.009

Kircher, T. T., Leube, D. T., Erb, M., Grodd, W., and Rapp, A. M. (2007). Neural correlates of metaphor processing in schizophrenia. *Neuroimage* 34, 281–289. doi: 10.1016/j.neuroimage.2006.08.044

Kita, S., de Condappa, O., and Mohr, C. (2007). Metaphor explanation attenuates the right-hand preference for depictive co-speech gestures that imitate actions. *Brain Lang.* 101, 185–197. doi: 10.1016/j.bandl.2006.11.006

Krach, S., Blumel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., et al. (2009). Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI. *BMC Neurosci.*10:9. doi: 10.1186/1471-2202-10-9

Lee, S. S., and Dapretto, M. (2006). Metaphorical vs. literal word meanings: fMRI evidence against a selective role of the right hemisphere. *Neuroimage* 29, 536–544. doi: 10.1016/j.neuroimage.2005.08.003

Leube, D., Straube, B., Green, A., Blümel, I., Prinz, S., Schlotterbeck, P., et al. (2012). A possible brain network for representation of cooperative behavior and its implications for the psychopathology of schizophrenia. *Neuropsychobiology* 66, 24–32. doi: 10.1159/000337131

Lotze, M., Heymans, U., Birbaumer, N., Veit, R., Erb, M., Flor, H., et al. (2006). Differential cerebral activation during observation of expressive gestures and motor acts. *Neuropsychologia* 44, 1787–1795. doi: 10.1016/j.neuropsychologia.2006.03.016

Mainieri, A. G., Heim, S., Straube, B., Binkofski, F., and Kircher, T. (2013). Differential role of mentalizing and the mirror neuron system in the imitation of communicative gestures. *NeuroImage* 81, 294–305. doi: 10.1016/j.neuroimage.2013.05.021

Mashal, N., Faust, M., Hendler, T., and Jung-Beeman, M. (2007). An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain Lang.* 100, 115–126. doi: 10.1016/j.bandl.2005.10.005

Mashal, N., Faust, M., Hendler, T., and Jung-Beeman, M. (2009). An fMRI study of processing novel metaphoric sentences. *Laterality* 14, 30–54. doi: 10.1080/13576500802049433

Molnar-Szakacs, I., Wu, A. D., Robles, F. J., and Iacoboni, M. (2007). Do you see what I mean? Corticospinal excitability during observation of culture-specific gestures. *PLoS ONE* 2:e626. doi: 10.1371/journal.pone.0000626

Moseley, R., Carota, F., Hauk, O., Mohr, B., and Pulvermüller, F. (2012). A role for the motor system in binding abstract emotional meaning. *Cereb. Cortex* 22, 1634–1647. doi: 10.1093/cercor/bhr238

Nakamura, A., Maess, B., Knösche, T. R., Gunter, T. C., Bach, P., and Friederici, A. D. (2004). Cooperation of different neuronal systems during hand sign recognition. *Neuroimage* 23, 25–34. doi: 10.1016/j.neuroimage.2004.04.034

Nichols, T., Brett, M., Andersson, J., Wager, T., and Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage* 25, 653–660. doi: 10.1016/j.neuroimage.2004.12.005

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Pallier, C., Dehaene, S., Poline, J. B., LeBihan, D., Argenti, A. M., Dupoux, E., et al. (2003). Brain imaging of language plasticity in adopted adults: can a second language replace the first? *Cereb. Cortex* 13, 155–161. doi: 10.1093/cercor/13.2.155

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci* 8, 976–987. doi: 10.1038/nrn2277

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., et al. (1996). Brain processing of native and foreign languages. *Neuroreport* 7, 2439–2444. doi: 10.1097/00001756-199611040-00007

Perlovsky, L. I., and Ilin, R. (2010). Neurally and mathematically motivated architecture for language and thought. *Open Neuroimag. J.* 4, 70–80. doi: 10.2174/1874440001004010070

Perlovsky, L. I., and Ilin, R. (2013). Mirror neurons, language, and embodied cognition. *Neural Netw.* 41, 15–22. doi: 10.1016/j.neunet.2013.01.003

Pierno, A. C., Tubaldi, F., Turella, L., Grossi, P., Barachino, L., Gallo, P., et al. (2009). Neurofunctional modulation of brain regions by the observation of pointing and grasping actions. *Cereb. Cortex* 19, 367–374. doi: 10.1093/cercor/bhn089

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in (2009). *Ann. N. Y. Acad. Sci.* 1191, 62–88. doi: 10.1111/j.1749-6632.2010.05444.x

Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360. doi: 10.1038/nrn2811

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., and Kircher, T. T. (2004). Neural correlates of metaphor processing. *Brain Res. Cogn. Brain Res.* 20, 395–402. doi: 10.1016/j.cogbrainres.2004.03.017

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., and Kircher, T. T. (2007). Laterality in metaphor processing: lack of evidence from functional magnetic resonance imaging for the right hemisphere theory. *Brain Lang.* 100, 142–149. doi: 10.1016/j.bandl.2006.04.004

Sachs, O., Weis, S., Krings, T., Huber, W., and Kircher, T. (2008a). Categorical and thematic knowledge representation in the brain: neural correlates of taxonomic and thematic conceptual relations. *Neuropsychologia* 46, 409–418. doi: 10.1016/j.neuropsychologia.2007.08.015

Sachs, O., Weis, S., Zellagui, N., Huber, W., Zvyagintsev, M., Mathiak, K., et al. (2008b). Automatic processing of semantic relations in fMRI: neural activation during semantic priming of taxonomic and thematic categories. *Brain Res.* 1218, 194–205. doi: 10.1016/j.brainres.2008.03.045

Sachs, O., Weis, S., Zellagui, N., Sass, K., Huber, W., Zvyagintsev, M., et al. (2011). How different types of conceptual relations modulate brain activation during semantic priming. *J. Cogn. Neurosci.* 23, 1263–1273. doi: 10.1162/jocn.2010.21483

Sass, K., Krach, S., Sachs, O., and Kircher, T. (2009a). Lion - tiger - stripes: neural correlates of indirect semantic priming across processing modalities. *Neuroimage* 45, 224–236. doi: 10.1016/j.neuroimage.2008.10.014

Sass, K., Sachs, O., Krach, S., and Kircher, T. (2009b). Taxonomic and thematic categories: neural correlates of categorization in an auditory-to-visual priming task using fMRI. *Brain Res.* 1270, 78–87. doi: 10.1016/j.brainres.2009.03.013

Schlosser, M. J., Aoyagi, N., Fulbright, R. K., Gore, J. C., and McCarthy, G. (1998). Functional MRI studies of auditory comprehension. *Hum. Brain Mapp.* 6, 1–13. doi: 10.1002/(SICI)1097-0193(1998)6:1<1::AID-HBM1>3.0.CO;2-7

Schmidt, G. L., Kranjec, A., Cardillo, E. R., and Chatterjee, A. (2010). Beyond laterality: a critical assessment of research on the neural basis of metaphor. *J. Int. Neuropsychol. Soc.* 16, 1–5. doi: 10.1017/S1355617709990543

Schmidt, G. L., and Seger, C. A. (2009). Neural correlates of metaphor processing: the roles of figurativeness, familiarity and difficulty. *Brain Cogn.* 71, 375–386. doi: 10.1016/j.bandc.2009.06.001

Shibata, M., Abe, J., Terao, A., and Miyamoto, T. (2007). Neural mechanisms involved in the comprehension of metaphoric and literal sentences: an fMRI study. *Brain Res.* 1166, 92–102. doi: 10.1016/j.brainres.2007.06.040

Slotnick, S. D., and Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nat. Neurosci.* 7, 664–672. doi: 10.1038/nn1252

Straube, B., Green, A., Bromberger, B., and Kircher, T. (2011a). The differentiation of iconic and metaphoric gestures: common and unique integration processes. *Hum. Brain Mapp.* 32, 520–533. doi: 10.1002/hbm.21041

Straube, B., Green, A., Chatterjee, A., and Kircher, T. (2011b). Encoding social interactions: the neural correlates of true and false memories. *J. Cogn. Neurosci.* 23, 306–324. doi: 10.1162/jocn.2010.21505

Straube, B., Wolk, D., and Chatterjee, A. (2011c). The role of the right parietal lobe in the perception of causality: a tDCS study. *Exp. Brain Res.* 215, 315–325. doi: 10.1007/s00221-011-2899-1

Straube, B., Green, A., Jansen, A., Chatterjee, A., and Kircher, T. (2010). Social cues, mentalizing and the neural processing of speech accompanied by gestures. *Neuropsychologia* 48, 382–393. doi: 10.1016/j.neuropsychologia.2009.09.025

Straube, B., Green, A., Sass, K., Kirner-Veselinovic, A., and Kircher, T. (2013a). Neural integration of

speech and gesture in schizophrenia: evidence for differential processing of metaphoric gestures. *Hum. Brain Mapp.* 34, 1696–1712. doi: 10.1002/hbm.22015

Straube, B., Green, A., Sass, K., and Kircher, T. (2013b). Superior temporal sulcus disconnectivity during processing of metaphoric gestures in schizophrenia. *Schizophr. Bull.* doi: 10.1093/schbul/sbt110. [Epub ahead of print]. Available online at: http://schizophreniabulletin.oxfordjournals.org/content/early/2013/08/16/schbul.sbt110.full.pdf

Straube, B., Green, A., Weis, S., Chatterjee, A., and Kircher, T. (2009). Memory effects of speech and gesture binding: cortical and hippocampal activation in relation to subsequent memory performance. *J. Cogn. Neurosci.* 21, 821–836. doi: 10.1162/jocn.2009.21053

Straube, B., Green, A., Weis, S., and Kircher, T. (2012). A supramodal neural network for speech and gesture semantics: an fMRI study. *PLoS ONE* 7:e51207. doi: 10.1371/journal.pone.0051207

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., and Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 14792–14797. doi: 10.1073/pnas.94.26.14792

Tivarus, M. E., Ibinson, J. W., Hillier, A., Schmalbrock, P., and Beversdorf, D. Q. (2006). An fMRI study of semantic priming: modulation of brain activity by varying semantic distances. *Cogn. Behav. Neurol.* 19, 194–201.

Tremblay, P., and Small, S. L. (2011). From language comprehension to action understanding and back again. *Cereb. Cortex* 21, 1166–1177. doi: 10.1093/cercor/bhq189

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Uchiyama, H., Seki, A., Kageyama, H., Saito, D. N., Koeda, T., Ohno, K., et al. (2006). Neural substrates of sarcasm: a functional magnetic-resonance imaging study. *Brain Res.* 1124, 100–110. doi: 10.1016/j.brainres.2006.09.088

Uchiyama, H. T., Saito, D. N., Tanabe, H. C., Harada, T., Seki, A., Ohno, K., et al. (2012). Distinction between the literal and intended meanings of sentences: a functional magnetic resonance imaging study of metaphor and sarcasm. *Cortex* 48, 563–583. doi: 10.1016/j.cortex.2011.01.004

Ungerleider, L. G., and Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165. doi: 10.1016/0959-4388(94)90066-3

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., et al. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30, 1414–1432. doi: 10.1016/j.neuroimage.2005.11.002

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., and Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31, 329–338. doi: 10.1016/S0896-6273(01)00359-2

Wiggs, C. L., Weisberg, J., and Martin, A. (1999). Neural correlates of semantic and episodic memory retrieval. *Neuropsychologia* 37, 103–118. doi: 10.1016/S0028-3932(98)00044-X

Willems, R. M., and Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: a review. *Brain Lang.* 101, 278–289. doi: 10.1016/j.bandl.2007.03.004

Willems, R. M., Hagoort, P., and Casasanto, D. (2010). Body-specific representations of action verbs: neural evidence from right- and left-handers. *Psychol. Sci.* 21, 67–74. doi: 10.1177/0956797609354072

Willems, R. M., Ozyurek, A., and Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* 17, 2322–2333. doi: 10.1093/cercor/bhl141

Willems, R. M., Ozyurek, A., and Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage* 47, 1992–2004. doi: 10.1016/j.neuroimage.2009.05.066

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., and Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20664–20669. doi: 10.1073/pnas.0909197106

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 July 2013; accepted: 24 August 2013; published online: 13 September 2013.

Citation: Straube B, He Y, Steines M, Gebhardt H, Kircher T, Sammer G and Nagels A (2013) Supramodal neural processing of abstract information conveyed by speech and gesture. Front. Behav. Neurosci. 7:120. doi: 10.3389/fnbeh.2013.00120

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 19 November 2013
doi: 10.3389/fnbeh.2013.00163

Reinforcement and inference in cross-situational word learning

Paulo F. C. Tilles and José F. Fontanari*

Departamento de Física e Informática, Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, Brazil

Edited by:
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA

Reviewed by:
Kenny Smith, University of Edinburgh, UK
Angelo Cangelosi, University of Plymouth, UK
George Kachergis, Leiden University, Netherlands

****Correspondence:***
José F. Fontanari, Departamento de Física e Informática, Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560-970 São Carlos, Brazil
e-mail: fontanari@ifsc.usp.br

Cross-situational word learning is based on the notion that a learner can determine the referent of a word by finding something in common across many observed uses of that word. Here we propose an adaptive learning algorithm that contains a parameter that controls the strength of the reinforcement applied to associations between concurrent words and referents, and a parameter that regulates inference, which includes built-in biases, such as mutual exclusivity, and information of past learning events. By adjusting these parameters so that the model predictions agree with data from representative experiments on cross-situational word learning, we were able to explain the learning strategies adopted by the participants of those experiments in terms of a trade-off between reinforcement and inference. These strategies can vary wildly depending on the conditions of the experiments. For instance, for fast mapping experiments (i.e., the correct referent could, in principle, be inferred in a single observation) inference is prevalent, whereas for segregated contextual diversity experiments (i.e., the referents are separated in groups and are exhibited with members of their groups only) reinforcement is predominant. Other experiments are explained with more balanced doses of reinforcement and inference.

Keywords: statistical learning, word learning, cross-situational learning, associative learning, mutual exclusivity

1. INTRODUCTION

A desirable goal of a psychological theory is to offer explanations grounded on elementary principles to the data available from psychology experiments (Newell, 1994). Although most of these quantitative psychological data are related to mental chronometry and memory accuracy, recent explorations on the human performance to acquire an artificial lexicon in controlled laboratory conditions have paved the way to the understanding of the learning strategies humans use to infer a word-object mapping (Yu and Smith, 2007; Kachergis et al., 2009; Smith et al., 2011; Kachergis et al., 2012; Yu and Smith, 2012a). These experiments are based on the cross-situational word-learning paradigm which avers that a learner can determine the meaning of a word by finding something in common across all observed uses of that word (Gleitman, 1990; Pinker, 1990). In that sense, learning takes place through the statistical sampling of the contexts in which a word appears in accord with the classical associationist stance of Hume and Locke that the mechanism of word learning is sensitivity to covariation: if two events occur at the same time, they become associated (Bloom, 2000).

In a typical cross-situational word-learning experiment, participants are exposed repeatedly to multiple unfamiliar objects concomitantly with multiple spoken pseudo-words, such that a word and its correct referent (object) always appear together on a learning trial. Different trials exhibiting distinct word-object pairs will eventually allow the disambiguation of the word-object associations and the learning of the correct mapping (Yu and Smith, 2007). However, it is questionable whether this scenario is suitable to describe the actual word learning process by children even in the unambiguous situation where the single novel object is followed by the utterance of its corresponding pseudo-word. In fact, it was shown that young children will only make the connection between the object and the word provided they have a reason to believe that they are in presence of an act of naming and for this the speaker has to be present (Baldwin et al., 1996; Bloom, 2000; Waxman and Gelman, 2009). Adults could learn those associations either because they were previously instructed by the experimenter that they would be learning which words go with which objects or because they could infer that the disembodied voice is an act of naming by a concealed person. Although there have been claims that cross-situational statistical learning is part of the repertoire of young word learners (Yu and Smith, 2008), the effect of individual differences in attention and vocabulary development of the infants complicates considerably this issue which is still a matter for debate (Yu and Smith, 2012b; Smith and Yu, 2013).

There are several other alternative or complementary approaches to the statistical learning formulation of language acquisition considered in this paper. For instance, the social-pragmatic hypothesis claims that the child makes the connections between words and their referents by understanding the referential intentions of others. This approach, which seems to be originally due to Augustine, implies that children use intuitive psychology to "read" the adults' minds (Bloom, 2000). A more recent approach that explores the grounding of language in perception and action has been proved effective in the design of

linguistic capabilities in humanoid cognitive robots (Cangelosi et al., 2007; Cangelosi, 2010; Pezzulo et al., 2013) as well as in the support of word learning by toddlers through the stabilization of their attention on the selected object (Yu and Smith, 2012b). In contrast with the unsupervised cross-situational learning scheme, the scenario known as operant conditioning involves the active participation of the agents in the learning process, with exchange of non-linguistic cues to provide feedback on the learner inferences. This supervised learning scheme has been applied to the design of a system for communication by autonomous robots in the Talking Heads experiments (Steels, 2003). We note that a comparison between the cross-situational and operant conditioning learning schemes indicates that they perform similarly in the limit of very large lexicon sizes (Fontanari and Cangelosi, 2011).

As our goal is to interpret the learning performance of adults using a few plausible reasoning tenets, here we assume that in order to learn a word-object mapping within the cross-situational word-learning scenario the learner should be able to (i) recall at least a fraction of the word-object pairings that appeared in the learning trials, (ii) register both co-occurrences and non-co-occurrences of words and objects and (iii) apply the mutual exclusivity principle which favors the association of novel words to novel objects (Markman and Wachtel, 1988). Of course, we note that a hypothetical learner could achieve cross-situational learning solely by registering and recalling co-occurrences of words and objects without carrying out any inferential reasoning (Blythe et al., 2010; Tilles and Fontanari, 2012a), but we find it implausible that human learners would not reap the benefits (e.g., fast mapping) of employing mutual exclusivity (Vogt, 2012; Reisenauer et al., 2013).

In this paper we offer an adaptive learning algorithm that comprises two parameters which regulate the associative reinforcement of pairings between concurrent words and objects, and the non-associative inference process that handles built-in biases (e.g., mutual exclusivity) as well as information of past learning events. By setting the values of these parameters so as to fit a representative selection of experimental data presented in Kachergis et al. (2009, 2012) we are able to identify and explain the learning strategies adopted by the participants of those experiments in terms of a trade-off between reinforcement and inference.

2. CROSS-SITUATIONAL LEARNING SCENARIO

We assume there are N objects $o_1, \ldots, o_N$, N words $w_1, \ldots, w_N$ and a one-to-one mapping between words and objects represented by the set $\Gamma = \{(w_1, o_1), \ldots, (w_N, o_N)\}$. At each learning trial, C word-object pairs are selected from Γ and presented to the learner without providing any clue on which word goes with which object. For instance, pictures of the C objects are displayed in a slide while C pseudo-words are spoken sequentially such that their spatial and temporal arrangements do not give away the correct word-object associations (Yu and Smith, 2007; Kachergis et al., 2009). We refer to the subset of words and their referents (objects) presented to the learner in a learning trial as the context $\Omega = \{w_1, o_1, w_2, o_2, \ldots, w_C, o_C\}$. The context size C is then a measure of the within-trial ambiguity, i.e., the number of co-occurring word-object pairs per learning trial. The selection procedure from the set Γ, which may favor some particular subsets of word-object pairs, determines the different experimental setups discussed in this paper. Although each individual trial is highly ambiguous, repetition of trials with partially overlapping contexts should in principle allow the learning of the N word-object associations.

After the training stage is completed, which typically comprises about two dozen trials, the learning accuracy is measured by instructing the learner to pick the object among the N objects on display which the learner thinks is associated to a particular target word. The test is repeated for all N words and the average learning accuracy calculated as the fraction of correct guesses (Kachergis et al., 2009).

This cross-situational learning scenario does not account for the presence of noise, such as the effect of out-of-context words. This situation can be modeled by assuming that there is a certain probability (noise) that the referent of one of the spoken words is not part of the context (so that word can be said to be out of context). Although theoretical analysis shows that there is a maximum noise intensity beyond which statistical learning is unattainable (Tilles and Fontanari, 2012b), as yet no experiment was carried out to verify the existence of this threshold phenomenon on the learning performance of human subjects.

3. MODEL

We model learning as a change in the confidence with which the algorithm (or, for simplicity, the learner) associates the word w_i to an object o_j that results from the observation and analysis of the contexts presented in the learning trials. More to the point, this confidence is represented by the probability $P_t(w_i, o_j)$ that w_i is associated to o_j at learning trial t. This probability is normalized such that $\sum_{o_j} P_t(w_i, o_j) = 1$ for all w_i and $t > 0$, which then implies that when the word w_i is presented to the learner in the testing stage the learning accuracy is given simply by $P_t(w_i, o_i)$. In addition, we assume that $P_t(w_i, o_j)$ contains information presented in the learning trials up to and including trial t only.

If at learning trial t the learner observes the context $\Omega_t = \{w_1, o_1, w_2, o_2, \ldots, w_C, o_C\}$ then it can infer the existence of two other informative sets. First, the set of the words (and their referents) that appear for the first time at trial t, which we denote by $\tilde{\Omega}_t = \left\{\tilde{w}_1, \tilde{o}_1, \tilde{w}_2, \tilde{o}_2, \ldots, \tilde{w}_{\tilde{C}}, \tilde{o}_{\tilde{C}_t}\right\}$. Clearly, $\tilde{\Omega}_t \subseteq \Omega_t$ and $\tilde{C}_t \leq C$. Second, the set of words (and their referents) that do not appear in Ω_t but that have already appeared in the previous trials, $\bar{\Omega}_t = \left\{\bar{w}_1, \bar{o}_1, \ldots, \bar{w}_{N_t-C}, \bar{o}_{N_t-C}\right\}$ where N_t is the total number of different words that appeared in contexts up to and including trial t. Clearly, $\bar{\Omega}_t \cap \Omega_t = \varnothing$. The update rule of the confidences $P_t(w_i, o_j)$ depends on which of these three sets the word w_i and the object o_j belong to (if $i \neq j$ they may belong to different sets). In fact, our learning algorithm comprises a parameter $\chi \in [0, 1]$ that measures the associative reinforcement capacity and applies only to known words that appear in the current context, and a parameter $\beta \in [0, 1]$ that measures the inference capacity and applies either to known words that do not appear in the current context or to new words in the current context. Before the

experiment begins ($t = 0$) we set $P_0\left(w_i, o_j\right) = 0$ for all words w_i and objects o_j. Next we describe how the confidences are updated following the sequential presentation of contexts.

In the first trial ($t = 1$) all words are new ($\tilde{C}_1 = N_1 = C$), so we set

$$P_1\left(\tilde{w}_i, \tilde{o}_j\right) = \frac{1}{C} \tag{1}$$

for $\tilde{w}_i, \tilde{o}_j \in \tilde{\Omega} = \Omega$. In the second or in an arbitrary trial t we expect to observe contexts exhibiting both novel and repeated words. Novel words must go through an inference preprocessing stage before the reinforcement procedure can be applied to them. This is so because if $\tilde{w}_i$ appears for the first time at trial t then $P_{t-1}\left(\tilde{w}_i, o_j\right) = 0$ for all objects o_j and since the reinforcement is proportional to $P_{t-1}\left(\tilde{w}_i, o_j\right)$ the confidences associated to $\tilde{w}_i$ would never be updated (see Equation 5 and the explanation thereafter). Thus, when a novel word $\tilde{w}_i$ appear at trial $t \geq 2$, we redefine its confidence values at the previous trial (originally set to zero) as

$$P_{t-1}\left(\tilde{w}_i, \tilde{o}_j\right) = \frac{\beta}{\tilde{C}_t} + \frac{1-\beta}{N_{t-1} + \tilde{C}_t}, \tag{2}$$

$$P_{t-1}\left(\tilde{w}_i, o_j\right) = \frac{1-\beta}{N_{t-1} + \tilde{C}_t}, \tag{3}$$

$$P_{t-1}\left(\tilde{w}_i, \bar{o}_j\right) = \frac{1-\beta}{N_{t-1} + \tilde{C}_t}. \tag{4}$$

On the one hand, setting the inference parameter β to its maximum value $\beta = 1$ enforces the mutual exclusivity principle which requires that the new word $\tilde{w}_i$ be associated with equal probability to the $\tilde{C}_t$ new objects $\tilde{o}_j$ in the current context. Hence in the case $\tilde{C}_t = 1$ the meaning of the new word would be inferred in a single presentation. On the other hand, for $\beta = 0$ the new word is associated with equal probability to all objects already seen up to and including trial t, i.e., $N_t = N_{t-1} + \tilde{C}_t$. Intermediate values of β describe a situation of imperfect inference. Note that using Equations 2–4 we can easily verify that $\sum_{\tilde{o}_j} P_{t-1}\left(\tilde{w}_i, \tilde{o}_j\right) + \sum_{o_j} P_{t-1}\left(\tilde{w}_i, o_j\right) + \sum_{\bar{o}_j} P_{t-1}\left(\tilde{w}_i, \bar{o}_j\right) = 1$, in accord with the normalization constraint.

Now we can focus on the update rule of the confidence $P_t\left(w_i, o_j\right)$ in the case both word w_i and object o_j appear in the context at trial t. The rule applies both to repeated and novel words, provided the confidences of the novel words are preprocessed according to Equations 2–4. In order to fulfill automatically the normalization condition for word w_i, the increase of the confidence $P_t\left(w_i, o_j\right)$ with $o_j \in \Omega_t$ must be compensated by the decrease of the confidences $P_t\left(w_i, \bar{o}_j\right)$ with $\bar{o}_j \in \bar{\Omega}_t$. This can be implemented by distributing evenly the total flux of probability out of the latter confidences, i.e., $\sum_{\bar{o}_j \in \bar{\Omega}_t} P_{t-1}\left(w_i, \bar{o}_j\right)$, over the confidences $P_t\left(w_i, o_j\right)$ with $o_j \in \Omega_t$. Hence the net gain of confidence on the association between w_i and o_j is given by

$$r_{t-1}\left(w_i, o_j\right) = \chi P_{t-1}\left(w_i, o_j\right) \frac{\sum_{\bar{o}_j \in \bar{\Omega}_t} P_{t-1}\left(w_i, \bar{o}_j\right)}{\sum_{o_j \in \Omega_t} P_{t-1}\left(w_i, o_j\right)} \tag{5}$$

where, as mentioned before, the parameter $\chi \in [0, 1]$ measures the strength of the reinforcement process. Note that if both o_j and o_k appear in the context together with w_i then the reinforcement procedure should not create any distinction between the associations $\left(w_i, o_j\right)$ and (w_i, o_k). This result is achieved provided that the ratio of the confidence gains equals the ratio of the confidences before reinforcement, i.e., $r_{t-1}\left(w_i, o_j\right) / r_{t-1}(w_i, o_k) = P_{t-1}\left(w_i, o_j\right) / P_{t-1}(w_i, o_k)$. This is the reason that the reinforcement gain of a word-object association given by Equation 5 is proportional to the previous confidence on that association. The total increase in the confidences between w_i and the objects that appear in the context, i.e., $\sum_{o_j \in \Omega_t} r_{t-1}\left(w_i, o_j\right)$, equals the product of χ and the total decrease in the confidences between w_i and the objects that do not appear in the context, i.e., $\sum_{\bar{o}_j \in \bar{\Omega}_t} P_{t-1}\left(w_i, \bar{o}_j\right)$. So for $\chi = 1$ the confidences associated to objects absent from the context are fully transferred to the confidences associated to objects present in the context. Lower values of χ allows us to control the flow of confidence from objects in $\bar{\Omega}_t$ to objects in Ω_t.

Most importantly, in order to implement the reinforcement process the learner should be able to gauge the relevance of the information about the previous trials, which is condensed on the confidence values $P_t\left(w_i, o_j\right)$. The gauging of this information is quantified by the word and trial dependent quantity $\alpha_t(w_i) \in [0, 1]$ that allows for the interpolation between the cases of maximum relevance ($\alpha_t(w_i) = 1$) and complete irrelevancy ($\alpha_t(w_i) = 0$) of the information stored in the confidences $P_t\left(w_i, o_j\right)$. In particular, we assume that the greater the certainty on the association between word w_i and its referent, the more relevant that information is to the learner. A quantitative measure of the uncertainty associated to the confidences regarding word w_i is given by the entropy

$$H_t(w_i) = -\sum_{o_j \in \Omega_t \cup \bar{\Omega}_t} P_t\left(w_i, o_j\right) \log\left[P_t\left(w_i, o_j\right)\right] \tag{6}$$

whose maximum ($\log N_t$) is obtained by the uniform distribution $P_t\left(w_i, o_j\right) = 1/N_t$ for all $o_j \in \Omega_t \cup \bar{\Omega}_t$, and whose minimum (0) by $P_t\left(w_i, o_j\right) = 1$ and $P_t(w_i, o_k) = 0$ for $o_k \neq o_j$. So we define

$$\alpha_t(w_i) = \alpha_0 + (1 - \alpha_0)\left[1 - \frac{H_t(w_i)}{\log N_t}\right], \tag{7}$$

where $\alpha_0 \in [0, 1]$ is a baseline information gauge factor corresponding to the maximum uncertainty about the referent of a target word.

Finally, recalling that at trial t the learner has access to the sets Ω_t, $\bar{\Omega}_t$ as well as to the confidences at trial $t - 1$ we write the update rule

$$P_t\left(w_i, o_j\right) = P_{t-1}\left(w_i, o_j\right) + \alpha_{t-1}(w_i) r_{t-1}\left(w_i, o_j\right) + \left[1 - \alpha_{t-1}\left(w_i\right)\right]\left[\frac{1}{N_t} - P_{t-1}\left(w_i, o_j\right)\right] \tag{8}$$

for $w_i, o_j \in \Omega_t$. Note that if $\alpha_{t-1}(w_i) = 0$ the learner would associate word w_i to all objects that have appeared up to and

including trial t with equal probability. This situation happens only if $\alpha_0 = 0$ and if there is complete uncertainty about the referent of word w_i. Hence the quantity $\alpha_t(w_i)$ determines the extent to which the previous confidences on associations involving word w_i influence the update of those confidences.

Now we consider the update rule for the confidence $P_t(w_i, \bar{o}_j)$ in the case that word w_i appears in the context at trial t but object $\bar{o}_j$ does not. (We recall that object $\bar{o}_j$ must have appeared in some previous trial.) According to the reasoning that led to Equation 5 this confidence must decrease by the amount $\chi P_{t-1}(w_i, \bar{o}_j)$ and so, taking into account the information gauge factor, we obtain

$$P_t(w_i, \bar{o}_j) = P_{t-1}(w_i, \bar{o}_j) - \alpha_{t-1}(w_i)\,\chi P_{t-1}(w_i, \bar{o}_j) + [1 - \alpha_{t-1}(w_i)]\left[\frac{1}{N_t} - P_{t-1}(w_i, \bar{o}_j)\right] \quad (9)$$

which can be easily seen to satisfy the normalization

$$\sum_{o_j \in \Omega_t} P_t(w_i, o_j) + \sum_{\bar{o}_j \in \bar{\Omega}_t} P_t(w_i, \bar{o}_j) = 1. \quad (10)$$

We focus now on the update rule for the confidence $P_t(\bar{w}_i, \bar{o}_j)$ with $\bar{w}_i, \bar{o}_j \in \bar{\Omega}_t$, i.e., both the word $\bar{w}_i$ and the object $\bar{o}_j$ are absent from the context shown at trial t, but they have already appeared, not necessarily together, in previous trials. A similar inference reasoning that led to the expressions for the preprocessing of new words would allow the learner to conclude that a word absent from the context should be associated to an object that is also absent from it. In that sense, confidence should flow from the associations between $\bar{w}_i$ and objects $o_j \in \Omega_t$ to the associations between $\bar{w}_i$ and objects $\bar{o}_j \in \bar{\Omega}_t$. Hence, ignoring the information gauge factor for the moment, the net gain to confidence $P_t(\bar{w}_i, \bar{o}_j)$ is given by

$$\bar{r}_{t-1}(\bar{w}_i, \bar{o}_j) = \beta P_{t-1}(\bar{w}_i, \bar{o}_j)\,\frac{\sum_{o_j \in \Omega_t} P_{t-1}(\bar{w}_i, o_j)}{\sum_{\bar{o}_j \in \bar{\Omega}_t} P_{t-1}(\bar{w}_i, \bar{o}_j)}. \quad (11)$$

The direct proportionality of this gain to $P_{t-1}(\bar{w}_i, \bar{o}_j)$ can be justified by an argument similar to that used to justify Equation 5 in the case of reinforcement. The information relevance issue is also handled in a similar manner so the desired update rule reads

$$P_t(\bar{w}_i, \bar{o}_j) = P_{t-1}(\bar{w}_i, \bar{o}_j) + \alpha_{t-1}(\bar{w}_i)\,\bar{r}_{t-1}(\bar{w}_i, \bar{o}_j) + [1 - \alpha_{t-1}(\bar{w}_i)]\left[\frac{1}{N_t} - P_{t-1}(\bar{w}_i, \bar{o}_j)\right] \quad (12)$$

for $\bar{w}_i, \bar{o}_j \in \bar{\Omega}_t$. To ensure normalization the confidence $P_t(\bar{w}_i, o_j)$ must decrease by an amount proportional to $\beta P_{t-1}(\bar{w}_i, o_j)$ so that

$$P_t(\bar{w}_i, o_j) = P_{t-1}(\bar{w}_i, o_j) - \alpha_{t-1}(\bar{w}_i)\,\beta P_{t-1}(\bar{w}_i, o_j) + [1 - \alpha_{t-1}(\bar{w}_i)]\left[\frac{1}{N_t} - P_{t-1}(\bar{w}_i, o_j)\right] \quad (13)$$

for $\bar{w}_i \in \bar{\Omega}_t$ and $o_j \in \Omega_t$. We can verify that prescriptions (12) and (13) satisfy the normalization

$$\sum_{\bar{o}_j \in \bar{\Omega}_t} P_t(\bar{w}_i, \bar{o}_j) + \sum_{o_j \in \Omega_t} P_t(\bar{w}_i, o_j) = 1, \quad (14)$$

as expected.

In summary, before any trial ($t = 0$) we set all confidence values to zero, i.e., $P_0(w_i, o_j) = 0$, and fix the values of the parameters α_0, χ and β. In the first trial ($t = 1$) we set the confidences of the words and objects in Ω_1 according to Equation (1), so we have the values of $P_1(w_i, o_j)$ for $w_i, o_j \in \Omega_1$. In the second trial, we separate the novel words $\tilde{w}_i \in \tilde{\Omega}_2$ and reset $P_1(\tilde{w}_i, o_j)$ with $o_i \in \Omega_2 \cup \bar{\Omega}_2$ according to Equations 2–4. Only then we calculate $\alpha_1(w_i)$ with $w_i \in \Omega_1 \cup \tilde{\Omega}_2$ using Equation (7). The confidences at trial $t = 2$ then follows from Equations (8), (9), (12), and (13). As before, in the third trial we separate the novel words $\tilde{w}_i \in \tilde{\Omega}_3$, reset $P_2(\tilde{w}_i, o_j)$ with $o_i \in \Omega_3 \cup \bar{\Omega}_3$ according to Equations 2–4, calculate $\alpha_2(w_i)$ with $w_i \in \Omega_1 \cup \Omega_2 \cup \tilde{\Omega}_3$ using Equation (7), and only then resume the evaluation of the confidences at trial $t = 3$. This procedure is repeated until the training stage is completed, say, at $t = t^*$. At this point, knowledge of the confidence values $P_{t^*}(w_i, o_j)$ allows us to answer any question posed in the testing stage.

Our model borrows many features from other proposed models of word learning (Siskind, 1996; Fontanari et al., 2009; Frank et al., 2009; Fazly et al., 2010; Kachergis et al., 2012). In particular, the entropy expression (6) was used by Kachergis et al. (2012) to allocate attention trial-by-trail to the associations presented in the contexts. Here we use that expression to quantify the uncertainty associated to the various confidences in order to determine the extent to which those confidences are updated on a learning trial. A distinctive feature of our model is the update of associations that are not in the current trial according to Equation (12). In particular, we note that whereas *ad hoc* normalization can only decrease the confidences on associations between words and objects that did not appear in the current context, our update rule can increase those associations as well. The extent of this update is weighted by the inference parameter β and it allows the application of mutual exclusivity to associations that are not shown in the current context. In fact, the splitting of mental processes in two classes, namely, reinforcement processes that update associations in the current context and inference processes that update the other associations is the main thrust of our paper. In the next section we evaluate the adequacy of our model to describe a selection of cross-situational word-learning experiments carried out on adult subjects by Kachergis et al. (2009, 2012).

4. RESULTS

The cross-situational word-learning experiments of Kachergis et al. (2009, 2012) aimed to understand how word sampling frequency (i.e., number of trials in which a word appears), contextual diversity (i.e., the co-occurrence of distinct words or groups of words in the learning trials), within-trial ambiguity (i.e., the context size C), and fast-mapping of novel words affect the learning performance of adult subjects. In this section we compare

the performance of the algorithm described in the previous section with the performance of adult subjects reported in Kachergis et al. (2009, 2012). In particular, once the conditions of the training stage are specified, we carry out 10^4 runs of our algorithm for fixed values of the three parameters α_0, β, χ, and then calculate the average accuracy at trial $t = t^*$ over all those runs for that parameter setting. Since the algorithm is deterministic, what changes in each run is the composition of the contexts at each learning trial. As our goal is to model the results of the experiments, we search the space of parameters to find the setting such that the performance of the algorithm matches that of humans within the error bars (i.e., one standard deviation) of the experiments.

4.1. WORD SAMPLING FREQUENCY

In these experiments the number of words (and objects) is $N = 18$ and the training stage totals $t^* = 27$ learning trials, with each trial comprising the presentation of 4 words together with their referents ($C = 4$). Following Kachergis et al. (2009), we investigate two conditions which differ with respect to the number of times a word is exhibited in the training stage. In the two-frequency condition, the 18 words are divided into two subsets of 9 words each. The words in the first subset appear 9 times and those in the second only 3 times. In the three-frequency condition, the 18 words are divided into three subsets of 6 words each. Words in the first subset appear 3 times, in the second, 6 times and in the third, 9 times. In these two conditions, the same word was not allowed to appear in two consecutive learning trials.

Figures 1, **2** summarize our main results for the two-frequency and three-frequency conditions, respectively. The left panels show the regions (shaded areas) in the (χ, β) plane for fixed α_0 where the algorithm describes the experimental data. We note that if those regions are located left to the diagonal $\chi = \beta$ then the inference process is dominant whereas if they are right to the diagonal then reinforcement is the dominant process. The middle panels show the accuracy of the best fit as function of the parameter α_0 and the right panels exhibit the values of χ and β corresponding to that fit. The broken horizontal lines and the shaded zones around them represent the means and standard deviations of the results of experiments carried out with 33 adult subjects (Kachergis et al., 2009).

It is interesting that although the words sampled more frequently are learned best in the two-frequency condition as expected, this advantage practically disappears in the three-frequency condition in which case all words are learned at equal levels within the experimental error. Note that the average accuracy for the words sampled 3 times is actually greater than the accuracy for the words sampled 6 times, but this inversion is not statistically significant, although, most surprisingly, the algorithm does reproduce it for $\alpha_0 \in [0.7, 0.8]$. According to Kachergis et al. (2009), the reason for the observed sampling frequency insensitivity might be because the high-frequency words are learned quickly and once they are learned subsequent trials containing those words will exhibit an effectively smaller within-trial ambiguity. In this vein, the inversion could be explained if by chance the words less frequently sampled were generally paired with the highly sampled words. Thus, contextual diversity seems to play a key role in cross-situational word learning.

4.2. CONTEXTUAL DIVERSITY AND WITHIN-TRIAL AMBIGUITY

In the first experiment aiming to probe the role of contextual diversity in the cross-situational learning, the 18 words were divided in two groups of 6 and 12 words each, and the contexts of size $C = 3$ were formed with words belonging to the same group only. Since the sampling frequency was fixed to 6 repetitions for each word, those words belonging to the more numerous group are exposed to a larger contextual diversity (i.e., the variety of different words with which a given word appear in the course of the training stage). The results summarized in **Figure 3** indicate clearly that contextual diversity enhances the learning accuracy. Perhaps more telling is the finding that incorrect responses are

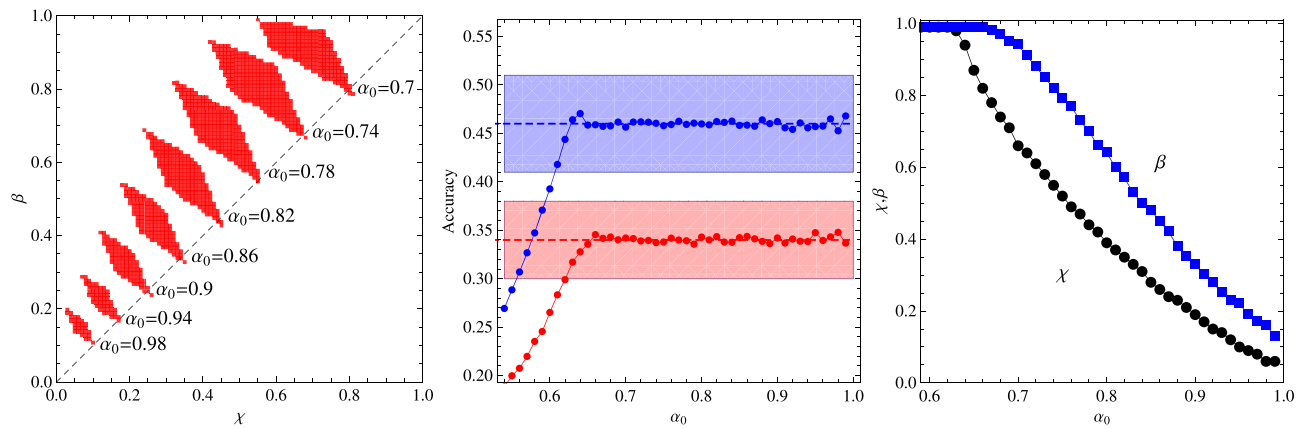


FIGURE 1 | Summary of the results for the two-frequency condition experiment. Left panel: Regions in the plane (χ, β) where the algorithm fits the experimental data for fixed α_0 as indicated in the figure. **Middle panel:** Average accuracy for the best fit to the results of Experiment 1 of Kachergis et al. (2009) represented by the broken horizontal lines (means) and shaded regions around them (one standard deviation). The blue symbols represent the accuracy for the group of words sampled 9 times whereas the red symbols represent the accuracy for the words sampled 3 times. **Right panel:** Parameters χ and β corresponding to the best fit shown in the middle panel. The other parameters are $N = 18$ and $C = 4$.

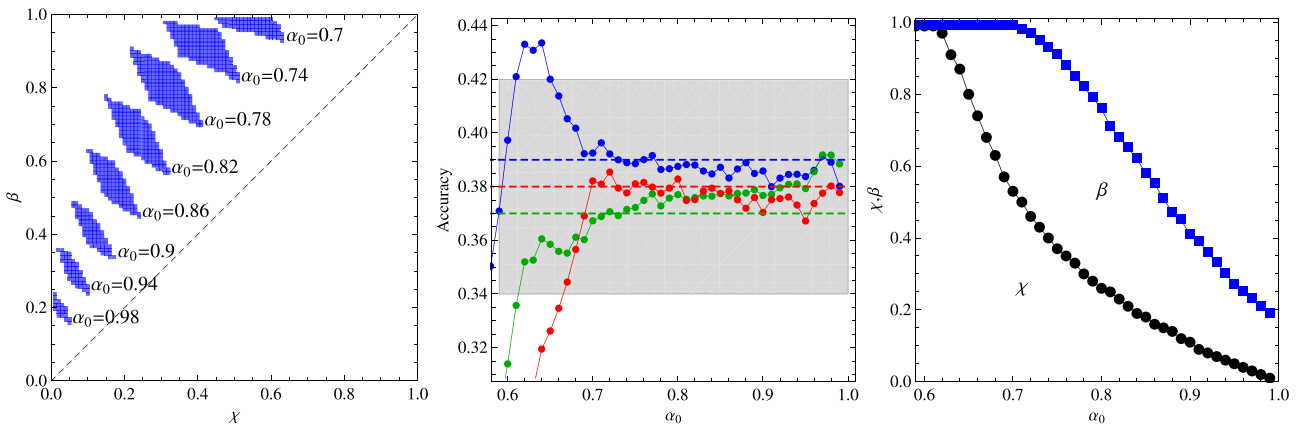


FIGURE 2 | Summary of the results for the three-frequency condition experiment. Left panel: Regions in the plane (χ, β) where the algorithm fits the experimental data for fixed α_0 as indicated in the figure. **Middle panel:** Average accuracy for the best fit to the results of Experiment 1 of Kachergis et al. (2009) represented by the broken horizontal lines (means) and shaded regions around them (one standard deviation). The blue symbols represent the accuracy for the group of words sampled 9 times, the green symbols for the words sampled 6 times, and the red symbols for the words sampled 3 times. **Right panel:** Parameters χ and β corresponding to the best fit shown in the middle panel. The other parameters are $N = 18$ and $C = 4$.

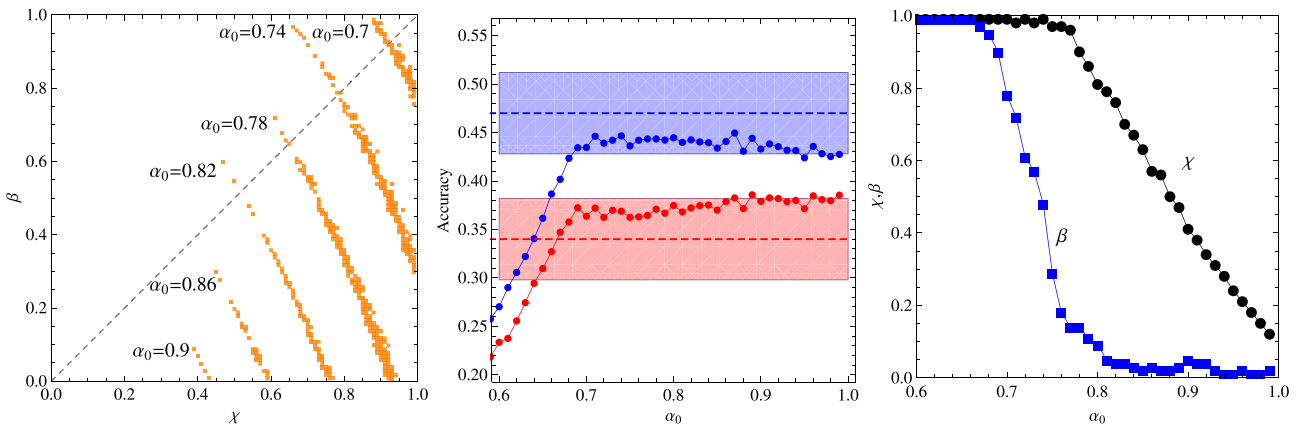


FIGURE 3 | Summary of the results of the two-level contextual diversity experiment. Left panel: Regions in the plane (χ, β) where the algorithm fits the experimental data for fixed α_0 as indicated in the figure. **Middle panel:** Average accuracy for the best fit to the results of Experiment 2 of Kachergis et al. (2009) represented by the broken horizontal lines (means) and shaded regions around them (one standard deviation). The blue symbols represent the accuracy for the group of words belonging to the 12-components subgroup and the red symbols for the words belonging to the 6-components subgroup. All words are repeated exactly 6 times during the $t^* = 27$ learning trials. **Right panel:** Parameters χ and β corresponding to the best fit shown in the middle panel. The other parameters are $N = 18$ and $C = 3$.

largely due to misassignments to referents whose words belong to the same group of the test word. In particular, Kachergis et al. (2009) found that this type of error accounts for 56% of incorrect answers when the test word belongs to the 6-components subgroup and for 76% when it belongs to the 12-components subgroup. The corresponding statistics for our algorithm with the optimal parameters set at $\alpha_0 = 0.9$ are 43% and 70%, respectively. The region in the space of parameters where the model can be said to describe the experimental data is greatly reduced in this experiment and even the best fit is barely within the error bars. It is interesting that, contrasting with the previous experiments, in this case the reinforcement procedure seems to play the more important role in the performance of the algorithm.

The effect of the context size or within-trial ambiguity is addressed by the experiment summarized in **Figure 4**, which is similar to the previous experiment, except that the words that compose the context are chosen uniformly from the entire repertoire of $N = 18$ words. Two context sizes are considered, namely, $C = 3$ and $C = 4$. In both cases, there is a large selection of parameter values that explain the experimental data, yielding results indistinguishable from the experimental average accuracies. This is the reason we do not exhibit a graph akin to those shown in the right panels of the previous figures. Since a perfect fitting can be obtained both for $\chi > \beta$ and for $\chi < \beta$, this experiment is uninformative with respect to these two abilities. As expected, increase of the within-trial ambiguity difficilitate

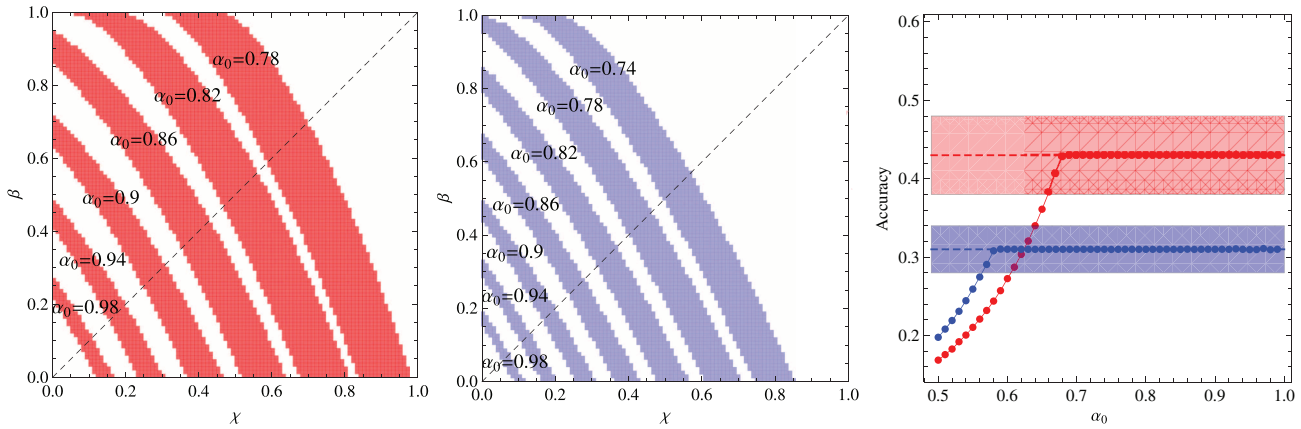


FIGURE 4 | Summary of the results of the experiments where all words co-occur without constraint and the $N = 18$ words are repeated exactly 6 times during the $t^* = 27$ learning trials. Left panel: Regions in the plane (χ, β) where the algorithm fits the experimental data for fixed α_0 and context size $C = 3$. **Middle panel:** Same as the left panel but for context size $C = 4$. **Right panel:** Average accuracy for the best fitting of the results of Experiment 2 of Kachergis et al. (2009) represented by the broken horizontal lines (means) and shaded regions around them (one standard deviation). The red symbols are for $C = 3$ and the blue symbols for $C = 4$.

learning. In addition, the (experimental) results for $C = 3$ yield a learning accuracy value that is intermediary to those measured for the 6 and 12-components subgroups, which is in agreement with the conclusion that the increase of the contextual diversity enhances learning, since the mean number of different co-occurring words is 4.0 in the 6-components subgroup, 9.2 in the 12-components subgroup and 8.8 in the uniformly mixed situation (Kachergis et al., 2009).

4.3. FAST MAPPING

The experiments carried out by Kachergis et al. (2012) were designed to elicit participants' use of the mutual exclusivity principle (i.e., the assumption of one-to-one mappings between words and referents) and to test the flexibility of a learned word-object association when new evidence is provided in support to a many-to-many mapping. To see how mutual exclusivity implies fast mapping assume that a learner who knows the association (w_1, o_1) is exposed to the context $\Omega = \{w_1, o_1, w_2, o_2\}$ in which the word w_2 (and its referent) appears for the first time. Then it is clear that a mutual-exclusivity-biased learner would infer the association (w_2, o_2) in this single trial. However, a purely associative learner would give equal weights to o_1 and o_2 if asked about the referent of w_2.

In the specific experiment we address in this section, $N = 12$ words and their referents are split up into two groups of 6 words each, say $A = \{(w_1, o_1), \ldots, (w_6, o_6)\}$ and $B = \{(w_7, o_7), \ldots, (w_{12}, o_{12})\}$. The context size is set to $C = 2$ and the training stage is divided in two phases. In the early phase, only the words belonging to group A are presented and the duration of this phase is set such that each word is repeated 3, 6 or 9 times. In the late phase, the contexts consist of one word belonging to A and one belonging to B forming fixed couples, i.e., whenever w_i appears in a context, w_{i+6}, with $i = 1, \ldots, 6$, must appear too. The duration of the late phase depends on the number of repetitions of each word that can be 3, 6, or 9 as in the early phase (Kachergis et al., 2012). The combinations of the sampling frequencies yield 9 different training conditions but here we will consider only the case that the late phase comprises 6 repetitions of each word.

The testing stage comprises the play of a single word, say w_1, and the display of 11 of the 12 trained objects (Kachergis et al., 2012). Each word was tested twice with a time lag between the tests: once without its corresponding early object (o_1 in the case) and once without its late object (o_7 in the case). This procedure requires that we renormalize the confidences for each test. For instance, in the case o_1 is left out of the display, the renormalization is

$$P_{t^*}{}'(w_1, o_j) = P_{t^*}(w_1, o_j) \Big/ \sum_{o_k \neq o_1} P_{t^*}(w_1, o_k) \qquad (15)$$

with $j = 2, \ldots, 12$ so that $\sum_{o_j \neq o_1} P_{t^*}{}'(w_1, o_j) = 1$. Similarly, in the case o_7 is left out the renormalization becomes

$$P_{t^*}{}'(w_1, o_j) = P_{t^*}(w_1, o_j) \Big/ \sum_{o_k \neq o_7} P_{t^*}(w_1, o_k) \qquad (16)$$

with $j = 1, \ldots, 6, 8, \ldots, 12$ so that $\sum_{o_j \neq o_7} P_{t^*}{}'(w_1, o_j) = 1$. We are interested on the (renormalized) confidences $P_{t^*}{}'(w_1, o_1)$, $P_{t^*}{}'(w_1, o_7)$, $P_{t^*}{}'(w_7, o_7)$, and $P_{t^*}{}'(w_7, o_1)$, which are shown in **Figures 5**, **6** for the conditions where words w_i, $i = 1, \ldots, 6$ are repeated 3 (left panel), 6 (middle panel), and 9 (right panel) times in the early learning phase, and the words w_i, $i = 1, \ldots, 12$ are repeated 6 times in the late phase. The figures exhibit the performance of the algorithm for the set of parameters χ and β that fits best the experimental data of Kachergis et al. (2012) for fixed α_0. This optimum set is shown in **Figure 7** for the 6 early repetition condition, which is practically indistinguishable from the optima of the other two conditions. The conditions with the different word repetitions in the early phase intended to produce

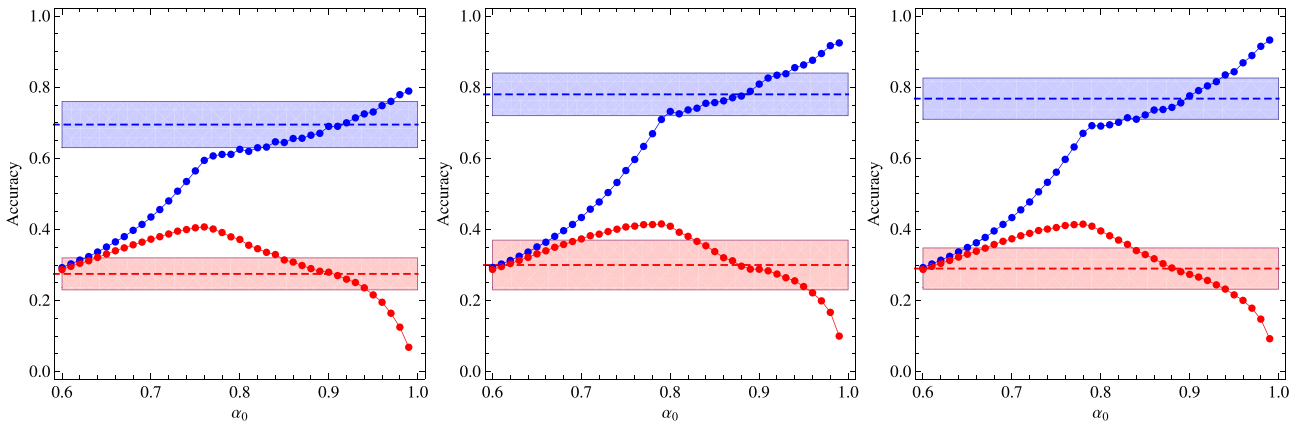


FIGURE 5 | Results of the experiments on mutual exclusivity in the case the late phase of the training process comprises 6 repetitions of each word. The blue symbols represent the probability that the algorithm picks object o_1 as the referent of word w_1 whereas the red symbols represent the probability it picks o_7. The broken horizontal lines and the shaded zones around them represent the experimental means and standard deviations (Kachergis et al., 2012) represented by the broken horizontal lines (means) and shaded regions around them (one standard deviation). The left panel shows the results for 3 repetitions of w_1 in the early training phase, the middle panel for 6 repetitions and the right panel for 9 repetitions. The results correspond to the parameters χ and β that best fit the experimental data for fixed α_0.

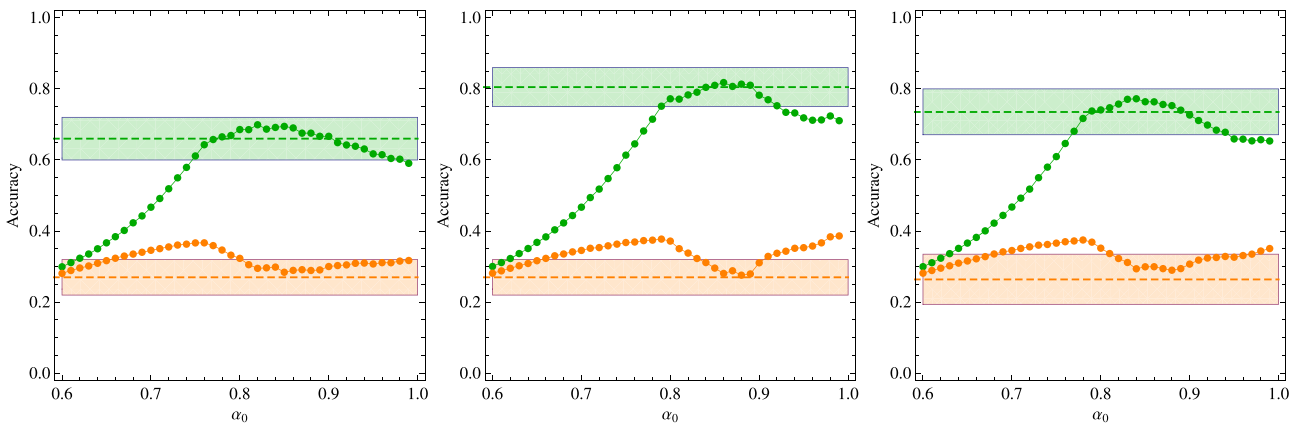


FIGURE 6 | Results of the experiments on mutual exclusivity in the case the late phase of the training process comprises 6 repetitions of each word. The green symbols represent the probability that the algorithm picks object o_7 as the referent of word w_7 whereas the orange symbols represent the probability it picks o_1. The broken horizontal lines and the shaded zones around them represent the experimental means and standard deviations (Kachergis et al., 2012) represented by the broken horizontal lines (means) and shaded regions around them (one standard deviation). The left panel shows the results for 3 repetitions of w_1 in the early training phase, the middle panel for 6 repetitions and the right panel for 9 repetitions. The results correspond to the parameters χ and β that best fit the experimental data for fixed α_0.

distinct confidences on the learned association (w_1, o_1) before the onset of the late phase in the training stage. The insensitivity of the results to these conditions probably indicates that association was already learned well enough with 3 repetitions only. Finally, we note that, though the testing stage focused on words w_1 and w_7 only, all word pairs w_i and w_{i+6} with $i = 1, \ldots, 6$ are strictly equivalent since they appear the same number of times during the training stage.

The experimental results exhibited in **Figure 6** offer indirect evidence that the participants have resorted to mutual exclusivity to produce their word-object mappings. In fact, from the perspective of a purely associative learner, word w_7 should be associated to objects o_1 or o_7 only, but since in the testing stage one of those objects was not displayed, such a learner would surely select the correct referent. However, the finding that $P_{t^*}{}'(w_7, o_7)$ is considerably greater than $P_{t^*}{}'(w_7, o_1)$ (they should be equal for an associative learner) indicates that there is a bias against the association (w_7, o_1) which is motivated, perhaps, from the previous understanding that o_1 was the referent of word w_1. In fact, a most remarkable result revealed by **Figure 6** is that $P_{t^*}{}'(w_7, o_7) < 1$. Since word w_7 appeared only in the late phase context $\Omega = \{w_1, o_1, w_7, o_7\}$ and object o_1 was not displayed in the testing stage, we must conclude that the participants produced spurious associations between words and objects that never appeared together in a context. Our algorithm accounts for these associations through Equation (4) in the case of new

words and, more importantly, through eqs. (9) and (13) due to the effect of the information efficiency factor $\alpha_t(w_i)$. The experimental data is well described only in the narrow range $\alpha_0 \in [0.85, 0.9]$.

Figure 8 exhibits the developmental timeline of the cross-situational learning history of the algorithm with the optimal set of parameters (see the figure caption) for the three different training conditions in the early training phase. This phase is characterized by the steady growth of the confidence on the association (w_1, o_1) (blue symbols) accompanied by the decrease of the confidence on association (w_1, o_7) (red symbols). As the word w_7 does not appear in the early training phase, the confidences on its association with any object remain constant corresponding to the accuracy value 1/11 (we recall that o_1 is left out of the display in the testing stage). The beginning of the late training stage is marked by a steep increase of the confidence on the association (w_7, o_7) (green symbols) whereas the confidence on (w_1, o_1) decreases gradually. A similar gradual increase is observed on the confidence on the association (w_7, o_1) (orange symbols). As expected, for large t all confidences presented in this figure tend to the same value, since the words w_1 and w_7 always appear together in the context $\Omega = \{w_1, o_1, w_7, o_7\}$. Finally, we note that this developmental timeline is qualitatively similar to that produced by the algorithm proposed by Kachergis et al. (2012).

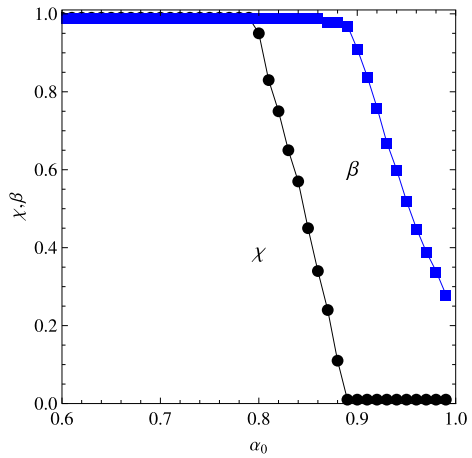


FIGURE 7 | Parameters χ (reinforcement strength) and β (inference strength) corresponding to the best fit shown in Figures 5 and 6 in the case word w_1 is repeated 6 times in the early training phase.

5. DISCUSSION

The chief purpose of this paper is to understand and model the mental processes used by human subjects to produce their word-object mappings in the controlled cross-situational word-learning scenarios devised by Yu and Smith (2007) and Kachergis et al. (2009, 2012). In other words, we seek to analyze the psychological phenomena involved in the production of those mappings. Accordingly, we assume that the completion of that task requires the existence of two cognitive abilities, namely, the associative capacity to create and reinforce associations between words and referents that co-occur in a context, and the non-associative capacity to infer word-object associations based on previous learning events, which accounts for the mutual exclusivity principle, among other things. In order to regulate the effectiveness of these two capacities we introduce the parameters $\chi \in [0, 1]$, which yields the reinforcement strength, and $\beta \in [0, 1]$, which determines the inference strength.

In addition, since the reinforcement and inference processes require storage, use and transmission of past and present information (coded mainly on the values of the confidences $P_t(w_i, o_j)$) we introduce a word-dependent quantity $\alpha_t(w_i) \in [0, 1]$ which gauges the impact of the confidences at trial $t - 1$ on the update of the confidences at trial t. In particular, the greater the certainty

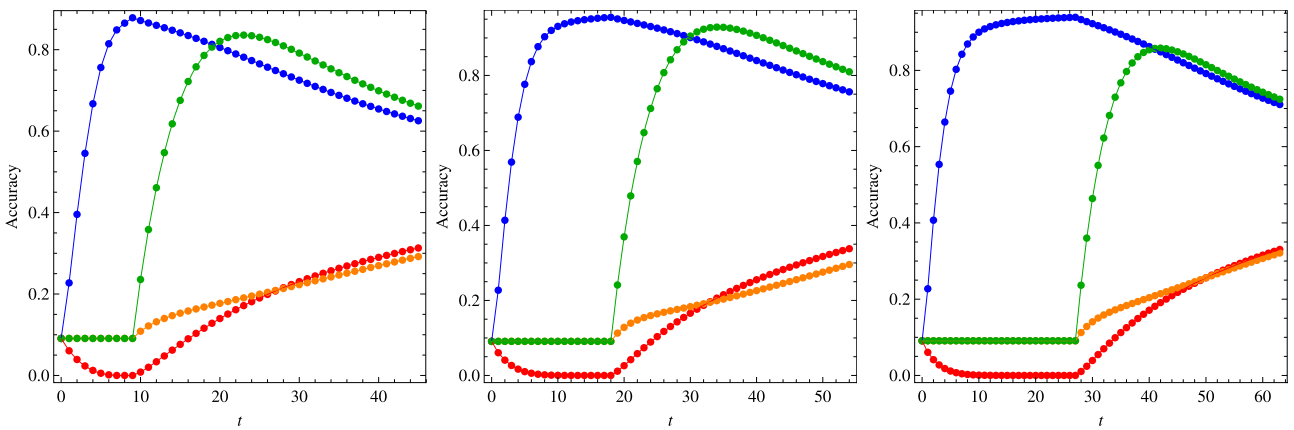


FIGURE 8 | Knowledge development for the model parameters that best fit the results of the mutual exclusivity experiments summarized in Figures 5, 6 in the case the late phase of the training process comprises 6 repetitions of each word. The symbol colors follow the convention used in those figures, i.e., the blue symbols represent the confidence on association (w_1, o_1), the red symbols on association (w_1, o_7), the green symbols on association (w_7, o_7) and the orange symbols on association (w_7, o_1). The left panel shows the results for 3 repetitions of w_1 in the early training phase ($\alpha_0 = 0.85$, $\chi = 0.25$, $\beta = 0.95$), the middle panel for 6 repetitions ($\alpha_0 = 0.85$, $\chi = 0.45$, $\beta = 0.99$) and the right panel for 9 repetitions ($\alpha_0 = 0.85$, $\chi = 0.4$, $\beta = 0.95$). For each trial t the symbols represent the average over 10^5 realizations of the learning process.

about the referent of word w_i, the greater the relevance of the previous confidences. However, there is a baseline information gauge factor $\alpha_0 \in [0, 1]$ used to process words for which the uncertainty about their referents is maximum. The adaptive expression for $\alpha_t(w_i)$ given in Equation (7) seems to be critical for the fitting of the experimental data. In fact, our first choice was to use a constant information gauge factor (i.e., $\alpha_t(w_i) = \alpha\ \forall t, w_i$) with which we were able to describe only the experiments summarized in **Figures 1**, **4** (data not shown). Note that a consequence of prescription (7) is that once the referent of a word is learned with maximum confidence (i.e., $P_t\big(w_i, o_j\big) = 1$ and $P_t(w_i, o_k) = 0$ for $o_k \neq o_j$) it is never forgotten.

The algorithm described in Section 3 comprises three free parameters χ, β and α_0 which are adjusted so as to fit a representative selection of the experimental data presented in Kachergis et al. (2009, 2012). A robust result from all experiments is that the baseline information gauge factor is in the range $0.7 < \alpha_0 < 1$. Actually, the fast mapping experiments narrow this interval down to $0.85 < \alpha_0 < 0.9$. This is a welcome result because we do not have a clear-cut interpretation for α_0—it encompasses storage, processing and transmission of information—and so the fact that this parameter does not vary much for wildly distinct experimental settings is evidence that, whatever its meaning, it is not relevant to explain the learning strategies used in the different experimental conditions. Fortunately, this is not the case for the two other parameters χ and β.

For instance, in the fast mapping experiments discussed in Subsection 4.3 the best fit of the experimental data is achieved for $\beta \approx 1$ indicating thus the extensive use of mutual exclusivity, and inference in general, by the participants of those experiments. Moreover, in that case the best fit corresponds to a low (but non-zero) value of χ, which is expected since for contexts that exhibit two associations ($C = 2$) only, most of the disambiguations are likely to be achieved solely through inference. This contrasts with the experiments on variable word sampling frequencies discussed in Subsection 4.1, for which the best fit is obtained with intermediate values of β and χ so the participants' use of reinforcement and inference was not too unbalanced. The contextual diversity experiment of Subsection 4.2, in which the words are segregated in two isolated groups of 12 and 6 components, offers another extreme learning situation, since the best fit corresponds to $\chi \approx 1$ and $\beta \approx 0$ in that case. To understand this result, first we recall that most of the participants' errors were due to misassignments of referents belonging to the same group of the test word, and those confidences were strengthened mainly by the reinforcement process. Second, in contrast to the inference process, which creates and strengthens spurious intergroup associations via Equation (12), the reinforcement process solely weakens those associations via Equation (9). Thus, considering the learning conditions of the contextual diversity experiment it is no surprise that reinforcement was the participants' choice strategy.

It is interesting to note that the optimal set of parameters that describe the fast mapping experiments (see **Figures 5–8**) indicate that there is a trade-off in the values of those parameters, in the sense that high values of the inference parameter β require low values of the reinforcement parameter χ. Since this is not an artifact of the model which poses no constrain on those values (e.g., they are both large for small α_0), the trade-off may reveal a limitation on the amount of attentional resources available to the learner to distribute among the two distinct mental processes.

Our results agree with the findings of Smith et al. (2011) that participants use various learning strategies, which in our case are determined by the values of the parameters χ and β, depending on the specific conditions of the cross-situational word-learning experiment. In particular, in the case of low within-trial ambiguity those authors found that participants generally resorted to a rigorous eliminative approach to infer the correct word-object mapping. This is exactly the conclusion we reached in the analysis of the fast mapping experiment for which the within-trial ambiguity takes the lowest possible value ($C = 2$).

Although the adaptive learning algorithm presented in this paper reproduced the performance of adult participants in cross-situational word-learning experiments quite successfully, the deterministic nature of the algorithm hindered somewhat the psychological interpretation of the information gauge factor $\alpha_t(w_i)$. In fact, not only learning and behavior are best described as stochastic processes (Atkinson et al., 1965) but also the modeling of those processes requires (and facilitates) a precise interpretation of the model parameters, since they are introduced in the model as transition probabilities.

ACKNOWLEDGMENTS

The work of José F. Fontanari was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Paulo F. C. Tilles was supported by grant # 2011/11386-1, São Paulo Research Foundation (FAPESP).

REFERENCES

Atkinson, R. C., Bower, G. H., and Crothers, E. J. (1965). *An introduction to mathematical learning theory*. New York, NY: Wiley.

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., and Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Dev.* 67, 3135–3153. doi: 10.2307/1131771

Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT Press.

Blythe, R. A., Smith, K., and Smith, A. D. M. (2010). Learning times for large lexicons through cross-situational learning. *Cogn. Sci.* 34, 620–642. doi: 10.1111/j.1551-6709.2009.01089.x

Cangelosi, A. (2010). Grounding language in action and perception: from cognitive agents to humanoid robots. *Phys. Life Rev.* 7, 139–151. doi: 10.1016/j.plrev.2010.02.001

Cangelosi, A., Tikhanoff, V., Fontanari, J. F., and Hourdakis, E. (2007). Integrating language and cognition: a cognitive robotics approach. *IEEE Comput. Intell. M.* 2, 65–70. doi: 10.1109/MCI.2007.385366

Fazly, A., Alishahi, A., and Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cogn. Sci.* 34, 1017–1063. doi: 10.1111/j.1551-6709.2010.01104.x

Fontanari, J. F., and Cangelosi, A. (2011). Cross-situational and supervised learning in the emergence of communication. *Interact. Stud.* 12, 119–133. doi: 10.1075/is.12.1.05fon

Fontanari, J. F., Tikhanoff, V.. Cangelosi, A., Ilin, R., and Perlovsky, L. I. (2009). Cross-situational learning of object-word mapping using Neural Modeling Fields. *Neural Netw.* 22, 579–585. doi: 10.1016/j.neunet.2009.06.010

Frank, M. C., Goodman, N. D., and Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychol. Sci.* 20, 578–585. doi: 10.1111/j.1467-9280.2009.02335.x

Gleitman, L. (1990). The structural sources of verb meanings. *Lang. Acquis.* 1, 1–55. doi: 10.1207/s15327817la0101_2

Kachergis, G., Yu, C., and Shiffrin, R. M. (2009). "Frequency and contextual diversity effects in cross-situational word learning," in *Proceedings of 31st Annual Meeting of the Cognitive Science Society*. (Austin, TX: Cognitive Science Society), 755–760.

Kachergis, G., Yu, C., and Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word-referent mappings. *Psychon. Bull. Rev.* 19, 317–324. doi: 10.3758/s13423-011-0194-6

Markman, E. M., and Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cogn. Psychol.* 20, 121–157. doi: 10.1016/0010-0285(88)90017-5

Newell, A. (1994). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., and Spivey, M. J. (2013). Computational Grounded Cognition: a new alliance between grounded cognition and computational modeling. *Front. Psychol.* 3:612. doi: 10.3389/fpsyg.2012.00612

Pinker, S. (1990). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Reisenauer, R., Smith, K., and Blythe, R. A. (2013). Stochastic dynamics of lexicon learning in an uncertain and nonuniform world. *Phys. Rev. Lett.* 110, 258701. doi: 10.1103/PhysRevLett.110.258701

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61, 39–91. doi: 10.1016/S0010-0277(96)00728-7

Smith, K., Smith, A. D. M., and Blythe, R. A. (2011). Cross-situational learning: an experimental study of word-learning mechanisms. *Cogn. Sci.* 35, 480–498. doi: 10.1111/j.1551-6709.2010.01158.x

Smith, L. B., and Yu, C. (2013). Visual attention is not enough: individual differences in statistical word-referent learning in infants *Lang. Learn. Dev.* 9, 25–49. doi: 10.1080/15475441.2012.707104

Steels, L. (2003). Evolving grounded communication for robots. *Trends Cogn. Sci.* 7, 308–312. doi: 10.1016/S1364-6613(03)00129-3

Tilles, P. F. C., and Fontanari, J. F. (2012a). Minimal model of associative learning for cross-situational lexicon acquisition. *J. Math. Psychol.* 56, 396–403. doi: 10.1016/j.jmp.2012.11.002

Tilles, P. F. C., and Fontanari, J. F. (2012b). Critical behavior in a cross-situational lexicon learning scenario. *Europhys. Lett.* 99, 60001. doi: 10.1209/0295-5075/99/60001

Vogt, P. (2012). Exploring the robustness of cross-situational learning under Zipfian distributions. *Cogn. Sci.* 36, 726–739. doi: 10.1111/j.1551-6709.2011.1226.x

Waxman, S. R., and Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends Cogn. Sci.* 13, 258–263. doi: 10.1016/j.tics.2009.03.006

Yu, C., and Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* 18, 414–420. doi: 10.1111/j.1467-9280.2007.01915.x

Yu, C., and Smith, L. B. (2008). "Statistical cross-situational learning in adults and infants," in *Proceedings of the 32th Annual Boston University Conference on Language Development* (Somerville, MA: Cascadilla Press), 562–573.

Yu, C., and Smith, L. B. (2012a). Modeling cross-situational word-referent learning: prior questions. *Psychol. Rev.* 119, 21–39. doi: 10.1037/a0026182

Yu, C., and Smith, L. B. (2012b). Embodied attention and word learning by toddlers. *Cognition* 125, 244–262. doi: 10.1016/j.cognition.2012.06.016

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 July 2013; paper pending published: 01 October 2013; accepted: 28 October 2013; published online: 19 November 2013.

Citation: Tilles PFC and Fontanari JF (2013) Reinforcement and inference in cross-situational word learning. Front. Behav. Neurosci. 7:163. doi: 10.3389/fnbeh.2013.00163

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 18 December 2013
doi: 10.3389/fnbeh.2013.00181

The role of semantic abstractness and perceptual category in processing speech accompanied by gestures

Arne Nagels[1]*, Anjan Chatterjee[2], Tilo Kircher[1] and Benjamin Straube[1]

[1] *Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Marburg, Germany*
[2] *Department of Neurology and the Center for Cognitive Neuroscience, The University of Pennsylvania, Philadelphia, PA, USA*

Edited by:
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA

Reviewed by:
D. Caroline Blanchard, University of Hawaii at Manoa, USA
Mar Sanchez, Emory University, USA

***Correspondence:**
Arne Nagels, Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Rudolf-Bultmann-Straße 8, 35039 Marburg, Germany
e-mail: nagels@med.uni-marburg.de

Space and shape are distinct perceptual categories. In language, perceptual information can also be used to describe abstract semantic concepts like a "rising income" (space) or a "square personality" (shape). Despite being inherently concrete, co-speech gestures depicting space and shape can accompany concrete or abstract utterances. Here, we investigated the way that abstractness influences the neural processing of the perceptual categories of space and shape in gestures. Thus, we tested the hypothesis that the neural processing of perceptual categories is highly dependent on language context. In a two-factorial design, we investigated the neural basis for the processing of gestures containing shape (SH) and spatial information (SP) when accompanying concrete (c) or abstract (a) verbal utterances. During fMRI data acquisition participants were presented with short video clips of the four conditions (cSP, aSP, cSH, aSH) while performing an independent control task. Abstract (a) as opposed to concrete (c) utterances activated temporal lobes bilaterally and the left inferior frontal gyrus (IFG) for both shape-related (SH) and space-related (SP) utterances. An interaction of perceptual category and semantic abstractness in a more anterior part of the left IFG and inferior part of the posterior temporal lobe (pTL) indicates that abstractness strongly influenced the neural processing of space and shape information. Despite the concrete visual input of co-speech gestures in all conditions, space and shape information is processed differently depending on the semantic abstractness of its linguistic context.

Keywords: iconic gestures, deictic gestures, metaphoric gestures, functional magnetic resonance imaging, speech-associated gestures, cognition

INTRODUCTION

In face-to-face communication people often use gestures to complement the content of their verbal message. People produce different kinds of gestures (McNeill, 1992), such as iconic gestures illustrating shape (e.g., "The ball is round") or deictic gestures referring to spatial information in our physical environment (e.g., "The cat is sitting on the roof"; pointing gesture). Shape gestures resemble the information they convey, as when someone draws a circle in the air to indicate a round shape ("The table in the kitchen is round," circle gesture). Space and shape gestures typically refer to concrete entities in the world. However, they can also make abstract references depending on the nature of the verbal message (McNeill, 1992; McNeill et al., 1993a,b). For instance, shape-related gestures can illustrate a deep connection between twins when the speaker touches the fingertips of both hands ("The twins had a spiritual bond between them"). Similarly space-related gestures can refer to abstract relationships or locations such as lifting the hand when saying that the discussion occurred at a very "high level."

In direct face-to-face communication people use gestures (Ozyurek and Kelly, 2007), regardless of whether the utterances are concrete or abstract. In line with theories suggesting gestures may represent the phylogenetic origin of human speech (Corballis, 2003, 2009, 2010; Gentilucci and Corballis, 2006; Gentilucci et al., 2006; Bernardis et al., 2008), gestures might represent the basis of spatial or action representations in human language [for example, see Tettamanti and Moro (2011)]. Such spatial elements transferred into speech and gestures could be an expression of how our language is rooted in embodied experiences (Gibbs, 1996; Lakoff, 1987). Following this idea perceptual elements and the sensory-motor system might both contribute to the processing and comprehension of figurative abstract language (particularly in the context of metaphors such as "grasp an idea"), as suggested by the embodiment theory (Gallese and Lakoff, 2005; Arbib, 2008; Fischer and Zwaan, 2008; D'Ausilio, 2009; Pulvermüller and Fadiga, 2010). Thus, the investigation of the neural substrates underlying the processing of perceptual categories such as shape or space in the context of concrete vs. abstract language semantics would give an answer to this hypothesis.

Recent fMRI investigations have focused on the processing of speech and gesture for different gesture types beat gestures: (Hubbard et al., 2009); iconic gestures: (Willems et al., 2007, 2009); and metaphoric gestures: (Kircher et al., 2009; Straube et al., 2009, 2011a). In general, left hemispheric posterior temporal (Holle et al., 2008, 2010; Green et al., 2009) and inferior frontal brain regions (Willems et al., 2007; Kircher et al., 2009; Straube et al., 2009, 2011a) are commonly found for the semantic processing of speech and gesture. The left posterior temporal

lobe (pTL) seems to be involved during the apprehension of co-verbal gestures, whereas the left inferior frontal gyrus (IFG) seems to be additionally recruited when processing gestures in an abstract sentence context (Kircher et al., 2009; Straube et al., 2011a, 2013) or when accompanying incongruent ("The fisherman has caught a huge fish," while the actor is angling his arms) concrete speech (Willems et al., 2007; Green et al., 2009; Willems et al., 2009). However, these studies do not examine the neural effects of processing of concrete or abstract utterances with different perceptual categories, such as gestures referring to shape (e.g., "The ball is round") or space (e.g., "The shed is next to the building").

In a previous study, we compared brain activation in response to object-related (non-social) and person-related (social) co-verbal gestures (Straube et al., 2010). Person-related as opposed to object-related gestures activated anterior brain regions including the medial and bilateral frontal cortex as well as the temporal lobes. These data indicate that dependent of speech and gesture content (person-related vs. object-related) different brain regions are activated during comprehension. However, in the aforementioned study the content of the verbal utterances was confounded by differences in the level of abstractness, since person-related gestures are not only social, but also more abstract symbolic than object-related gestures (e.g., "The actor did a good job in the play"). Therefore, the specific influence of person-related and object-related content independent of abstractness was not disentangled.

Beside this evidence for a posterior to anterior gradient of processing for concrete to abstract speech-gesture information, it is generally assumed that specific regions of the brain are specialized for the processing of specific kinds of contents (Patterson et al., 2007). Information about shapes of objects are processed in lateral occipital and inferior temporal brain areas (e.g., Kourtzi and Kanwisher, 2000; Grill-Spector et al., 2001; Kourtzi and Kanwisher, 2001; Kourtzi et al., 2003; Panis et al., 2008; Karnath et al., 2009, whereas the parietal lobe is involved in processing of spatial information (Rizzolatti et al., 1997, 2006; Koshino et al., 2000, 2005; Rizzolatti and Matelli, 2003; Chica et al., 2011; Gillebert et al., 2011). Although gestures can be distinguished by perceptual category [e.g., deictic gestures convey spatial information and iconic gestures predominantly convey shape information (McNeill, 1992)] there is insufficient knowledge about the neural processing of these different perceptual categories in the context of abstract and concrete sentence contexts.

Here we investigate the way in which perceptual category and semantic abstractness of co-verbal gestures interact. Our experiment aims at the question whether different perceptual categories are processed in the same or in distinct brain regions, irrespective of their linguistic abstractness. To approach this research question, we applied a naturalistic approach comparing shape-related and space-related gestures in the context of concrete and abstract sentences.

On a cognitive level (concrete physical) gesture content has to be aligned with the content of speech, regardless of whether the message is concrete or abstract. We hypothesize that the effort to incorporate both abstract speech with concrete gestures will likely result in enhanced neural responses in the left inferior frontal cortex (Willems et al., 2007) and in bilateral temporal brain regions (Kircher et al., 2009) as compared to the concrete conditions, independent of perceptual category. With regard to shape-related and space-related gestural information we expected differential activation within the inferior temporal and parietal lobe, respectively. For the interaction of perceptual (space and shape) and semantic category (concreteness and abstractness) two alternative results were hypothesized: (1) If the same neural processes are engaged when processing shape and space information regardless of the abstractness of the message, we will find no significant activation in interaction analyses. In this case, conjunction analyses (e.g., aSP > aSH ∩ cSP > cSH) will result in common activation patterns in the parietal cortex for space and inferior temporal cortex for shape. (2) If abstractness influences the processing of shape-related and space-related gesture information, interaction analyses will show differential activations between conditions. Here, we expected an interaction since language content may differentially influence the interpretation of perceptual categories and consequently the neural processing predominantly in the left IFG and pTL. Enhanced neural responses in classical "language regions" would strengthen the assumption that perceptual categories are differentially processed if embedded into an abstract vs. concrete language context.

MATERIALS AND METHODS

PARTICIPANTS

Seventeen male right handed (Oldfield, 1971) healthy volunteers, all native speakers of German (mean age = 23.8 ± 2.7 years, range: 20–30 years, mean years of school education = 12.65 ± 0.86, range: 10–13 years), without impairments of vision or hearing, participated in the study. None of the participants had any serious medical, neurological or psychiatric illness, past or present. All participants gave written informed consent and were paid 20 Euro for participation. The study was approved by the local ethics committee. Because of technical problems one fMRI-data set was excluded from the analyses.

STIMULUS CONSTRUCTION

A set of 388 short video clips depicting an actor was initially created, consisting of 231 concrete and 157 abstract sentences, each accompanied by co-verbal gestures.

Iconic gestures refer to the concrete content of sentences, whereas metaphoric gestures illustrate abstract information in sentences. For example in the sentences "To get down to business" (drop of the hand) or "The politician builds a bridge to the next topic" (depicting an arch with the hand), abstract information is illustrated using metaphoric gestures. By contrast, the same gestures can be iconic (drop of the right hand or depicting an arch with the right hand) with the sentences "The man goes down the hill" or "There is a bridge over the river" when they illustrate concrete physical features of the world. Thus, concrete utterances are those containing referents that are perceptible to the senses ("The man ascends to the top of the mountain"). Abstract sentences, on the other hand, contain referents that are not directly perceptible ("The man ascends to the top of the company"), where the spatial or shape terms in the utterance are being used figuratively. For the distinction between concrete and abstract concepts see Holmes and Rundle (1985).

Here we were interested in the neural processing of the following types of sentences accompanied by gestures: (1) utterances with concrete content and space-related perceptual information (cSP; "deictic gesture"); (2) utterances with concrete content and shape-related perceptual information (cSH; "iconic gesture"); (3) utterances with an abstract content and space-related perceptual information (aSP; "abstract deictic gestures"); and (4) utterances with an abstract content and shape-related perceptual information (aSH; "metaphoric gestures").

All sentences accompanying gestures had a length of 5–10 words, with an average duration of 2.37 s ($SD = 0.35$) and a similar grammatical form (subject—predicate—object). The speech and gestures were performed by the same male actor in a natural, spontaneous way. This procedure was continuously supervised by two of the authors (Benjamin Straube, Tilo Kircher) and timed digitally. All video clips had the same length of 5 s with at least 0.5 s before and after the sentence onset and offset, respectively, where the actor did not speak or move.

STIMULUS SELECTION: RATING / MATERIAL SELECTION/MATCHING

For stimulus validation, 17 raters not participating in the fMRI study evaluated each video on a scale ranging from 1 to 7 (1 = very low to 7 = very high) according to three content dimensions (space, shape and action information) and familiarity. Other general parameters like "understandability" and "naturalness" were previously validated and controlled for (for detailed information see (Green et al., 2009; Kircher et al., 2009; Straube et al., 2011a,b).

Material was selected to address our manipulations of interest (cf. above):

1. cSP = **C**oncrete content and **SP**ace-related information
2. cSH = **C**oncrete content and **SH**ape-related information
3. aSP = **A**bstract content and **SP**ace-related information
4. aSH = **A**bstract content and **SH**ape-related information

For each condition 30 sentences were selected to differentiate both factors. Therefore, co-verbal gestures conveying space-related perceptual information (cSP, aSP) were selected to have similar spatial rating scores independent of the level of the abstractness of the utterance (c vs. a). Abstract co-verbal gestures (aSP, aSH) were selected to be similarly abstract independent of the perceptual category of information (space or shape; see **Table 1**).

To confirm that our stimuli met our design criteria, we calculated analyses of variances for the factors perceptual (space-, shape related) and semantic category (concrete, abstract) as represented in the 2 × 2 experimental design.

As intended we found for the rating of spatial information a significant main effect for perceptual category [SP > SH; $F_{(1,\ 116)} = 72.532$, $p < 0.001$], but no significant effects for the main effect of semantic category [a vs. c; $F_{(1,\ 116)} = 0.149$, $p = 0.603$] or the interaction of perceptual and semantic category [$F_{(1,\ 116)} = 3.250$, $p = 0.074$].

For the rating of shape information we obtained again a significant main effect for perceptual category [SH > SP; $F_{(1,\ 120)} = 98.466$, $p < 0.001$], but no significant effects for the main effect of abstractness [a vs. c; $F_{(1,\ 120)} = 0.001$, $p = 0.988$] or the interaction of perceptual category and abstractness [$F_{(1,\ 120)} = 2.053$, $p = 0.155$].

For the rating of abstractness we obtained a significant main effect for abstractness [a > c; $F_{(1,\ 116)} = 116.124$, $p < 0.001$], but no significant effects for the main effect of perceptual category [SP vs. SH; $F_{(1,\ 116)} = 0.005$, $p = 0.942$] or the interaction of perceptual category and abstractness [$F_{(1,\ 116)} = 2.975$, $p = 0.087$]. For means and confidence intervals see **Table 1**. Together, these analyses confirm that stimulus selection worked out and stimulus characteristics for each condition met our design criteria.

For the control variables familiarity, naturalness and action information we found no significant main effects or interactions (for all $p > 0.10$). However, we found significant effects for understandability [main effect perceptual category: SP > SH: $F_{(1,120)} < 4.960$, $p = 0.028$; interaction: $F_{(1,116)} < 17.704$, $p < 0.001$], speech duration [main effect abstractness: a > c: $F_{(1,\ 116)} = 9.024$, $p < 0.003$] and gesture duration [main effect abstractness: c > a: $F_{(1,116)} < 10.821$, $p < 0.001$]. However, differences in understandability were small (<0.22 rating points) and most likely because of ceiling effects in the aSP (skewness = −1.68; kurtosis = 4.31) and cSH (skewness = −1.40; kurtosis =1.80) conditions. For means and confidence intervals of the control variables see **Table 2**.

In the event-related fMRI study design focusing on the co-occurrence of speech and gesture, differences in speech or gesture duration should not have a crucial impact on our results. However, we included differences in speech and gesture duration for each event as a covariate of no interest in our single-subject design matrix.

Table 1 | Experimental manipulations.

Condition	Space-related				Abstract				Shape-related			
	Mean	*SD*	95% confidence interval for mean		Mean	*SD*	95% confidence interval for mean		Mean	*SD*	95% confidence interval for mean	
cSP	5.71	0.60	5.49	5.93	2.18	0.82	1.87	2.49	2.97	0.97	2.60	3.33
cSH	3.91	0.91	3.57	4.25	2.44	0.78	2.15	2.74	5.17	1.39	4.65	5.69
aSP	5.40	0.53	5.20	5.59	4.19	0.90	3.85	4.53	3.25	0.87	2.92	3.57
aSH	4.08	0.86	3.76	4.41	3.90	1.00	3.53	4.27	4.90	0.95	4.54	5.25
Total	4.78	1.08	4.58	4.97	3.18	1.24	2.95	3.40	4.07	1.44	3.81	4.33

Rating results for the conditions cSP, cSH, aSP, and aSH for space-relatedness, abstractness and shape-relatedness.

Table 2 | Control variables: understandability, familiarity and naturalness.

Condition	Understandability				Familiarity				Naturalness			
	Mean	*SD*	95% confidence interval for mean		Mean	*SD*	95% confidence interval for mean		Mean	*SD*	95% confidence interval for mean	
cSP	6.63	0.21	6.55	6.71	4.98	0.97	4.62	5.34	4.87	0.48	4.69	5.05
cSH	6.84	0.14	6.79	6.89	4.75	0.77	4.46	5.04	4.97	0.48	4.79	5.15
aSP	6.82	0.17	6.75	6.88	5.10	0.91	4.76	5.44	4.80	0.59	4.58	5.02
aSH	6.75	0.20	6.68	6.83	4.80	0.89	4.47	5.13	4.80	0.49	4.62	4.98
Total	6.76	0.20	6.73	6.80	4.91	0.89	4.75	5.07	4.86	0.51	4.77	4.95

Rating results for the conditions cSP, cSH, aSP, and aSH for the dimensions understandability, familiarity and naturalness.

Apart from the aforementioned factors, further differences in movement characteristics were found between the conditions. For all four conditions predominantly right (cSP = 19; cSH = 13; aSP = 16; aSH = 11) or bimanual movements were performed (cSP = 11; cSH = 17; aSP = 14; aSH = 19). To ensure that none of the patterns of neural activation were produced by differences in hand movements (right hand vs. both hands) and speech length, a separate control analysis was run accounting for the aforementioned dimensions. A set of 11 exactly paired video clips for each condition was used for the additional analysis.

To account for differences in the size of movements between conditions, we coded each video clip with regard to the extent of the hand movement. We divided the video screen into small rectangles that corresponded to the gesture space described by McNeill (1992); McNeill (2005) and counted the number of rectangles in which gesture movements occurred see Straube et al. (2011a). For each video the number of rectangles was also included as covariate of no interest in the single subject model.

EXPERIMENTAL DESIGN AND PROCEDURE

During the fMRI scanning procedure, videos were presented via MR-compatible video goggles (VisuaStim©, Resonance Technology, Inc.) and non-magnetic headphones (audio presenting systems for stereophonic stimuli: Commander; Resonance Technology, Inc.), which additionally dampened scanner noise.

Thirty items of each of the four conditions were presented in an event-related design, in a pseudo-randomized order and counterbalanced across subjects. Each video was followed by a baseline condition (gray background with a fixation cross) with a variable duration of 3750–6750 ms (average: 5000 ms) see **Figure 1**.

During scanning participants were instructed to watch the videos and to indicate via left hand key presses at the beginning of each video whether the spot displayed on the actor's sweater was light or dark colored. This task was chosen to focus participants' attention on the middle of the screen and enabled us to investigate implicit speech and gesture processing without possible instruction-related attention biases. Performance rates and reaction times were recorded. Prior to scanning, each participant received at least 10 practice trials outside the scanner, which were different from the stimuli used in the main experiment. During the preparation scans additional clips were presented to adjust the volume of the headphone. Each participant performed two runs with 60 video clips and a total duration of 10.5 min each.

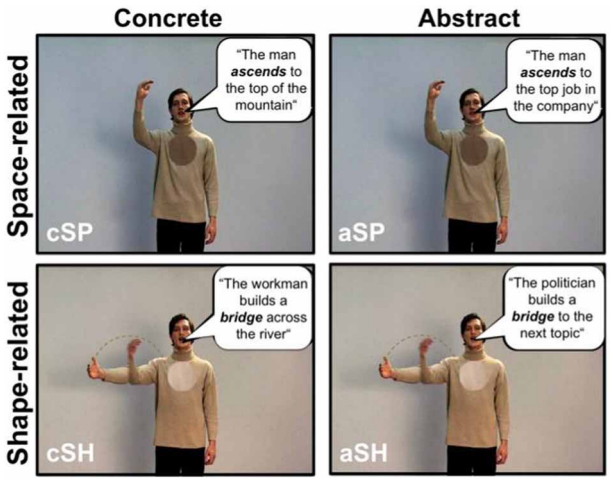


FIGURE 1 | Examples of the different speech and gesture video-clips. The stimulus material consisted of video clips of an actor performing either space-related **(top)** or shape-related **(bottom)** gestures to corresponding sentences with an concrete **(left)** or abstract content **(right)**. One screen shot of an example video is shown for each condition (cSP, concrete space-related; cSH, concrete shape-related; aSP, abstract space-related; aSH, abstract shape-related). In order to exemplify the stimulus material German sentences are translated into English, and written in speech bubbles for illustration (unlike in the actual stimuli).

fMRI DATA ACQUISITION

MRI was performed on a 3T Siemens scanner (Siemens MRT Trio series). Functional data were acquired with echo planar images in 38 transversal slices (repetition time [TR] = 2000 ms; echo time [TE] = 30 ms; flip angle = 90°; slice thickness = 3 mm; inter-slice gap = 0.30 mm; field of view [FoV] = 220 × 199 mm, voxel resolution = 3.44 × 3.44 mm, matrix dimensions 64 × 58 mm). Slices were positioned to achieve whole brain coverage. During each functional run 315 volumes were acquired.

DATA ANALYSIS

MR images were analyzed using Statistical Parametric Mapping (SPM2; www.fil.ion.ucl.ac.uk) implemented in MATLAB 6.5 (Mathworks Inc., Sherborn, MA). The first five volumes of every functional run were discarded from the analysis to minimize T1-saturation effects. To correct for different acquisition times, the

signal measured in each slice was shifted relative to the acquisition time of the middle slice using a slice interpolation in time. All images of one session were realigned to the first image of a run to correct for head movement and normalized into standard stereotaxic anatomical MNI-space by using the transformation matrix calculated from the first EPI-scan of each subject and the EPI-template. Afterwards, the normalized data with a resliced voxel size of 3.5 × 3.5 × 3.5 mm were smoothed with a 6 mm FWHM isotropic Gaussian kernel to accommodate intersubject variation in brain anatomy. Proportional scaling with high-pass filtering was used to eliminate confounding effects of differences in global activity within and between subjects.

The expected hemodynamic response at the defined "points of integration" for each event-type was modeled by two response functions, a canonical hemodynamic response function (HRF; Friston et al., 1998) and its temporal derivative. The temporal derivative was included in the model to account for the residual variance resulting from small temporal differences in the onset of the hemodynamic response, which is not explained by the canonical HRF alone. The functions were convolved with the event sequence, with fixed event duration of 1 s, for the onsets corresponding to the integration points of gesture stroke and sentence keyword to create the stimulus conditions in a general linear model (Green et al., 2009; Kircher et al., 2009; Straube et al., 2010, 2011b). The fixed event duration of 1 s was chosen to get a broader range of data around the assumed time point of integration. This methodological approach was also applied successfully in previous studies of co-verbal gesture processing (Kircher et al., 2009; Straube et al., 2010, 2011b).

A group analysis was performed by entering contrast images into a flexible factorial analysis as implemented in SPM5 in which subjects are treated as random variables. A Monte Carlo simulation of the brain volume of the current study was conducted to establish an appropriate voxel contiguity threshold (Slotnick et al., 2003). Assuming an individual voxel type I error of $p < 0.005$, a cluster extent of 8 contiguous re-sampled voxels was necessary to correct for multiple voxel comparisons at $p < 0.05$. Thus, voxels with a significance level of $p < 0.005$ uncorrected, belonging to clusters with at least eight voxels are reported (Straube et al., 2010). Activation peaks of some of the activation clusters also hold a family wise error (FWE) correction. Corresponding corrected p-values for each activation peak were included in the tables. The reported voxel coordinates of activation peaks are located in MNI space. Statistical analyses of data other than fMRI were performed using SPSS version 14.0 for Windows (SPSS Inc., Chicago, IL, USA). Greenhouse–Geisser correction was applied whenever necessary.

CONTRASTS OF INTEREST

To test our hypothesis on the neural processing of different perceptual categories in concrete vs. abstract sentence contexts (cf. Introduction section), baseline contrasts (main effects of condition), conjunction analysis and interaction analysis were run.

At first, baseline contrasts were calculated in order to detect general activations with regard to the four main conditions (aSP, cSP, aSH, cSH) as compared to baseline (fixation cross).

In a next step, main effects (SH vs. SP and a vs. c) as well as the interaction were calculated (t-contrasts) to show brain regions involved in the processing of different factors (directed general effects).

To test the hypothesis that perceptual category is processed in the same neural structures regardless of the language context we performed conjunction analyses of difference contrasts (aSP > aSH ∩ cSP > cSH and aSH > aSP ∩ cSH > cSP). To test for general effects of abstractness independent of both space-related as well as shape-related contents the same approach was used (aSP > cSP ∩ aSH > cSH and cSH > aSH ∩ cSP > aSP).

Finally, we performed two interaction analyses to test the hypothesis that abstractness significantly changes the processing of perceptual categories, space and shape: (1) = (aSP > cSP) > (aSH > cSH) masked for (aSP > cSP) and aSP; (2) = (aSH > cSH) > (aSP > cSP) masked for (aSH > cSH) and aSH. The masking procedure was applied to avoid the interpretation of deactivation in the concrete conditions and restrict the effects to increased activity for aSP vs. low-level baseline and its concrete derivative (cSP). Based on our hypothesis, this methodological approach enables us to find specific neural responses for semantic category (concrete/abstract) in space-related (1) and shape-related (2) perceptual contexts.

RESULTS

BEHAVIORAL RESULTS

The average reaction time for the control task ("indicate the color of the spot on the actor's sweater") did not differ with regard to color or gesture condition [color: $F_{(1,15)} = 0.506$, $P = 0.488$; condition: $F_{(4,60)} = 0.604$, $P = 0.604$; interaction: $F_{(4,60)} = 1.256$, $P = 0.301$; within-subjects two-factorial ANOVA; mean = 1.23 sec, $SD = 0.94$]. The participants showed an average accuracy rate of 99% which did not differ across conditions [$F_{(4,60)} = 0.273$, $P = 0.841$, within-subjects ANOVA]. Thus, the attention control task indicated that participants did pay attention to the video clips.

fMRI RESULTS

Baseline contrasts (aSP, cSP, aSH, cSH)

To explore the general processing mechanisms for each condition and the high comparability between conditions baseline contrasts were calculated (**Figure 2**, **Table 3**). We found comparable activation patterns as in previous studies on speech and gesture stimuli (Straube et al., 2011a).

Main effects for perceptual category

To identify the general effect of speech-gesture information, the main effect for the factors perception category [space-related (SP) vs. shape-related (SH)] were calculated.

For the effect of space-related vs. shape-related information (SP > SH) we found an extended network of activations including left middle [Brodmann Area (BA) 6] and superior frontal (BAs 6/8) as well as temporo-parietal (BAs 21/39/40) brain regions (**Table 4**).

The processing of shape-related vs. space-related information (SH > SP) resulted in enhanced neural responses in bilateral occipital-parietal (BAs 18/37) and middle (BA 11) as well as

inferior frontal (BA 45) gyri and left parietal (BA 40) brain region (**Table 4**).

Main effects for abstractness

Abstract vs. concrete speech-gesture information (a > c) revealed a widespread pattern of activation. A large cluster of activation was found in the left IFG extending to the temporal lobe, including the temporal pole and the middle temporal gyrus. Activations were also found in the right superior temporal gyrus, in the left precuneus and right cuneus as well as in the left precentral and superior medial gyri (BAs 6/9). Enhanced neural responses were also found in the middle cingulate, the left superior frontal and superior medial cortex as well as in the left angular gyrus (BA 39/40) (see **Table 5, Figure 3**).

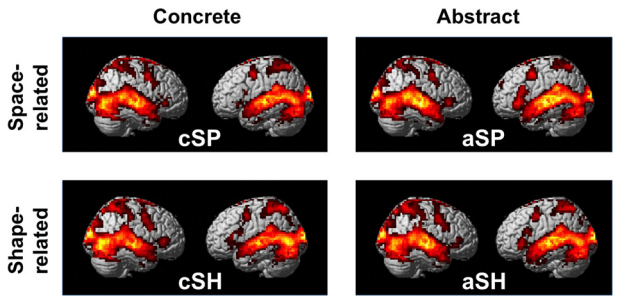


FIGURE 2 | Activation pattern in contrast to baseline (whole-brain, $p < 0.005$, cluster extend threshold = 8 voxels; MC corrected $p < 0.05$).

For the reverse contrast (c > a) we found activations in the left and right parahippocampal and fusiform gyri (BA 36/37), in the left inferior frontal (BA 46) and in the temporo-occipital region (BA 37) as well as in the left superior occipital gyrus (BA 19) (see **Table 5**). Smaller clusters of activation were found in the right cerebellum, the middle frontal (BA 11) and in the precentral gyrus (BA 4).

Interaction of perceptual categories and abstractness

For the interaction of perceptual category and abstractness (aSP > cSP)>(aSH > cSH) we found superior medial frontal, left inferior frontal (BA45/44) and middle temporal and superior parietal brain regions (see **Table 6**).

Table 3 | Gesture conditions in contrast to low level baseline (fixation cross).

Contrast	Anatomical region	Hem.	BA	Coordinates			*t*-value	*Uncor	*FWE	No. voxels
				x	*y*	*z*				
aSP	Superior temporal gyrus	L	22	−60	−21	0	16.71	< 0.001	< 0.001	7102
	Middle frontal gyrus	L	6	−46	4	53	7.87	< 0.001	< 0.001	127
	Precentral gyrus	R	6	53	0	49	6.57	< 0.001	< 0.001	106
	Superior parietal lobe	L	7	−25	−67	60	4.73	< 0.001	0.224	147
	Cerebellum	R		14	−25	−32	3.75	< 0.001	0.999	10
	Parahippocampal gyrus	L	34	−14	−11	−21	3.75	< 0.001	0.999	19
	Superior frontal gyrus	L	9	−14	53	25	3.68	< 0.001	1.000	20
	Cingulate gyrus	L	31	−11	−46	39	3.50	< 0.001	1.000	10
	Cerebellum	L		−11	−25	−32	3.40	0.001	1.000	15
cSP	Superior temporal gyrus	L	22	−60	−21	0	14.87	< 0.001	< 0.001	7196
	Inferior parietal lobe	L	40	−46	−39	63	6.20	< 0.001	0.001	225
	Caudate (sub-gyral)	L		0	4	18	4.34	< 0.001	0.737	29
	Cingulate gyrus	R	24	21	4	39	4.05	< 0.001	0.950	17
	Amygdala	L		−18	−4	−25	3.92	< 0.001	0.985	17
	Cerebellum	L		−14	−25	−32	3.83	< 0.001	0.995	10
	Precentral gyrus	R		28	−21	42	3.78	< 0.001	0.998	19
aSH	Superior temporal gyrus	L	22	−60	−21	0	16.16	< 0.001	< 0.001	8172
	Precentral gyrus	L	6	−49	0	53	7.03	< 0.001	< 0.001	104
	Middle frontal gyrus	R	6	53	4	49	6.54	< 0.001	< 0.001	182
	Caudate (sub-gyral)	L		−4	−4	21	4.31	< 0.001	0.764	18
	Precuneus	L	7	−25	−74	39	3.96	< 0.001	0.977	23
cSH	Middle occipital gyrus	L	19	−46	−77	0	14.91	< 0.001	< 0.001	7512
	Postcentral gyrus	L	2	−42	−39	63	6.67	< 0.001	< 0.001	371
	Inferior frontal gyrus	R	47	49	35	0	5.39	< 0.001	0.019	76
	Amygdala	L		−18	−4	−25	4.15	< 0.001	0.895	22
	Putamen	L		−21	11	−7	3.06	0.002	1.000	9

Significance level (t-value), size of the respective activation cluster (No. voxels; number of voxels > 8) at p < 0.005 MC corrected for multiple comparisons. Coordinates are listed in MNI space. BA is the Brodmann area nearest to the coordinate and should be considered approximate. (cSP, concrete spatial; cSH, concrete descriptive; AS, abstract spatial; aSH, abstract descriptive).

Table 4 | Main effects for space-related and shape-related semantic contents.

Contrast	Anatomical region	Hem.	BA	Coordinates			*t*-value	*Uncor	*FWE	No. voxels
				x	*y*	*z*				
SP > SH	SMA	L	6	−11	7	74	4.24	< 0.001	0.829	28
	Precentral gyrus	L	44	−42	7	49	3.92	< 0.001	0.986	20
	Inferior parietal lobe	L	39	−46	−81	28	3.81	< 0.001	0.996	10
	Superior frontal gyrus	L	8	−11	39	49	3.73	< 0.001	0.999	21
	Angular gyrus	L	39	−46	−60	32	3.59	< 0.001	1.000	23
	Superior temporal gyrus	L	40	−56	−46	21	3.48	< 0.001	1.000	10
	Middle temporal gyrus	L	21	−53	−21	−15	3.43	0.001	1.000	11
	Fusiform gyrus	R	20	42	−14	−21	3.30	0.001	1.000	9
SH > SP	Inferior occipital gyrus	L	37	−39	−70	−4	5.84	< 0.001	0.003	536
	Superior parietal lobe	R	7	28	−53	56	4.65	< 0.001	0.298	214
	Hippocampus	L		−21	−28	0	4.51	< 0.001	0.497	14
	Inferior frontal gyrus	L	45	−42	35	7	4.50	< 0.001	0.522	10
	Middle orbital gyrus	L	11	−28	39	−14	4.13	< 0.001	0.908	24
	Inferior parietal lobe	L	40	−32	−42	35	3.95	< 0.001	0.979	56
	Inferior frontal gyrus	R	46	46	39	7	3.94	< 0.001	0.982	23
	Middle occipital gyrus	R	17	32	−77	7	3.90	< 0.001	0.988	24
	Precentral gyrus	R	9	53	4	28	3.79	< 0.001	0.997	23
	Inferior occipital gyrus	R	18	28	−88	−11	3.68	< 0.001	1.000	10

Significance level (t-value), size of the respective activation cluster (No. voxels; number of voxels > 8) at $p < 0.005$ *MC corrected for multiple comparisons. Coordinates are listed in MNI space. BA is the Brodmann area nearest to the coordinate and should be considered approximate.*

Table 5 | Main effect for abstractness and concreteness.

Contrast	Anatomical Region	Hem.	BA	Coordinates			*t*-value	Uncor	No. voxels
				x	*y*	*z*			
a > c	Temporal pole	L	38	−49	10	−24	7.88	< 0.001	981
	Superior temporal gyrus	R	22	49	−14	0	5.29	< 0.001	313
	Precuneus	L	7	−10	−60	38	4.51	< 0.001	82
	Precentral gyrus	L	9	−38	7	42	4.18	< 0.001	31
	Superior medial gyrus	L	9	−10	60	32	4.05	< 0.001	32
	Cuneus	R	7	21	−60	35	3.90	< 0.001	17
	Superior medial gyrus	L	6	−4	28	60	3.70	< 0.001	10
	Angular gyrus	L	39	−42	−60	32	3.65	< 0.001	28
	Middle cingulate cortex	L	23	−7	−24	32	3.21	0.001	8
c > a	Fusiform gyrus	L	37	−32	−38	−14	5.46	< 0.001	126
	Inferior frontal gyrus	L	46	−42	32	10	4.03	< 0.001	18
	Cerebellum	R		32	−42	−24	3.90	< 0.001	9
	Inferior occipital gyrus	L	37	−49	−70	−7	3.82	< 0.001	25
	Fusiform gyrus	R	36	32	−35	−14	3.81	< 0.001	15
	Middle occipital gyrus	L	19	−35	−88	28	3.70	< 0.001	16
	Middle frontal gyrus	R	11	28	32	−18	3.63	< 0.001	13
	Precentral gyrus	R	4	28	−24	52	3.05	0.002	11

For the contrast in the opposite direction (aSH > cSH) > (aSP > cSP) we found a more distributed predominantly right hemispheric activation pattern including the occipital lobe, the middle frontal gyrus, the inferior parietal lobe, the precuneus, the IFG (BA44/45), the middle occipital gyrus and the bilateral fusiform gyri (see **Table 6**).

Specific contrasts of interest

Brain areas sensitive for shape-related and space-related perceptual contents independent of abstractness. A conjunction analysis for shape-related form descriptive perceptual contents irrespective of the level of abstractness (aSH > aSP ∩ cSH > cSP) revealed enhanced neural responses in the left middle occipital gyrus (BA 37; see supplementary material **Table 7**).

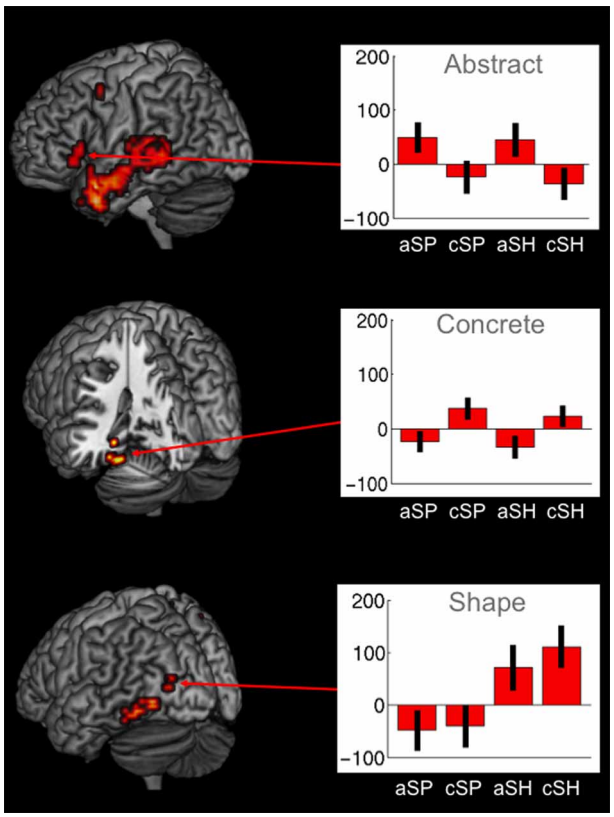


FIGURE 3 | Significant brain activations for abstractness, concreteness as well as for shape-related co-verbal gesture processing (whole-brain, $p < 0.005$, cluster extend threshold = 8 voxels; MC corrected $p < 0.05$) (cSP, concrete spatial; cSH, concrete shape; AS, abstract spatial; aSH, abstract shape).

No region was found to be significantly activated for space vs. shape-related processing on concrete and abstract level (aSP > aSH ∩ cSP > cSH) (see supplementary material **Table 8**).

Brain areas sensitive for abstractness independent of perceptual category (shape/space). Common activations for abstract as opposed to concrete co-verbal gestures, irrespective of descriptive or spatial information (aSH > cSH ∩ aSP > cSP), resulted in a large cluster of activation encompassing the left temporal pole and the middle temporal gyrus. Another cluster of activation was found in the right superior temporal gyrus and in the left IFG, including the pars Orbitalis as well as the pars Triangularis (BA 44; see supplementary material **Table 9**).

The imaging results for concreteness independent of the shape-related or space-related perceptual content (cSH > aSH ∩ cSP > aSP) revealed enhanced BOLD responses in the left parahippocampal gyrus (BA 35; see supplementary material **Table 10**).

Specific neural responses for abstractness in space-related (1) as well as in shape-related (2) content domains. The specifically masked interaction analyses (see Contrast of Interest section) revealed distinct activation for abstractness on space-related information [(sSP > cSP) > (aSH > cSH) masked for (aSP > cSP) and aSP] within the left IFG (MNIxyz: −53, 28, 0; $t = 4.77$; 42 voxels) and the left pTL (MNIxyz: −60, −46, 4; $t = 3.07$; 10 voxels; see **Figure 4**). The other direction of contrasts did not reveal any significant results.

Taken together, significant main effects and interactions of brain activation with regard to the manipulated factors [type of communicated perceptual information (SP, SH) and abstractness (c, a)] revealed different patterns of activation. The specific contrasts indicated that subregions of the left IFG and the left

Table 6 | Interaction of semantic categories and abstractness.

Contrast	Anatomical region	Hem.	BA	Coordinates			*t*-value	Uncor	No. voxels
				x	*y*	*z*			
(aSP>cSP)	Superior medial gyrus	L	8	−7	46	52	5.00	< 0.001	36
>	Inferior frontal gyrus	L	47	−52	28	0	4.77	< 0.001	57
(aSH>cSH)	Superior medial gyrus	L	9	−7	49	24	4.27	< 0.001	60
	Inferior frontal gyrus	L	47	−35	28	−18	3.61	< 0.001	8
	Superior parietal lobule	L	7	−28	−66	56	3.52	< 0.001	8
	Middle temporal gyrus	L	22	−60	−46	4	3.07	0.002	11
(aSH>cSH)	Fusiform gyrus	L	19	−28	−46	−10	5.03	< 0.001	48
>	Fusiform gyrus	R	19	28	−49	−10	4.48	< 0.001	23
(aSP>cSP)	Middle frontal gyrus	R	8	38	14	46	4.23	< 0.001	35
	Middle frontal gyrus	R	10	38	52	7	4.16	< 0.001	27
	Calcarine gyrus	R	18	14	−77	4	4.13	< 0.001	81
	Inferior parietal lobule	R	40	49	−49	46	3.67	< 0.001	31
	Inferior frontal gyrus	R	44	46	10	10	3.50	< 0.001	11
	Middle occipital gyrus	R	39	42	−77	32	3.28	0.001	9
	Precuneus	R	7	7	−56	56	3.23	0.001	12
	Paracentral lobule	L	6	−4	−28	52	3.22	0.001	20

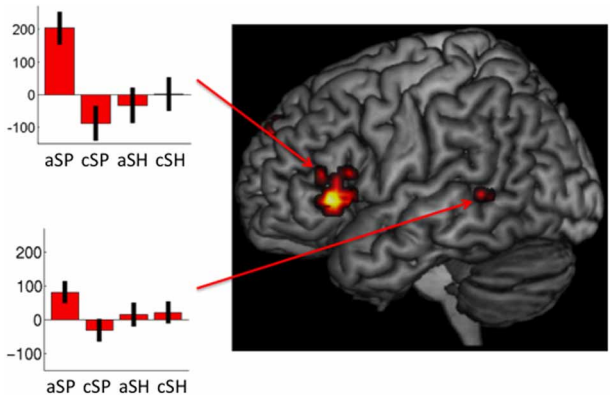


FIGURE 4 | Interaction of space-related co-verbal information processing and abstractness (whole-brain, $p < 0.005$, cluster extend threshold = 8 voxels; MC corrected $p < 0.05$) (cSP, concrete spatial; cSH, concrete shape; AS, abstract spatial; aSH, abstract shape).

pTL have common [conjunction analyses: IFG [MNIxyz: −39, 28, −4; $t = 3.27$; 11 voxels], pTL (MNIxyz: −53, −38, 0; $t = 4.26$; 196)] and distinct functions [interaction: IFG (MNIxyz: −53, 28, 0; $t = 4.77$; 42 voxels], pTL [MNIxyz: −60, −46, 4; $t = 3.07$; 10 voxels)] with regard to perceptual type and abstractness.

The same analysis, including only right-handed gesture stimuli of equal length (speech duration) revealed the same pattern of activation encompassing the left IFG as well as the left middle temporal gyrus, indicating that this effect is not based on irrelevant differences in stimulus material.

DISCUSSION

Space and shape are distinct perceptual categories. Words referring to space and shape also describe abstract concepts like "rising income" (space) or a "square personality" (shape). Gestures are an important part of human communication that underpin verbal utterances and can convey shape or space information even when accompanying abstract sentences. Recent studies have investigated the neural processing of speech and gesture (Willems and Hagoort, 2007; Willems et al., 2007, 2009; Dick et al., 2009, 2012; Green et al., 2009; Hubbard et al., 2009; Kelly et al., 2010; Kircher et al., 2009; Skipper et al., 2009; Straube et al., 2009; Holle et al., 2010). Despite the fact that the investigation of perceptual categories used in speech and gesture could give important answers with regard to the effect of abstractness on particular neural networks relevant for the processing of such perceptual information, the related effect is not known. Thus, the purpose of the current fMRI study was to investigate the neural processing of shape-related vs. space-related co-speech gesture information when presented with abstract or concrete utterances aiming at the question whether similar or distinct neural networks are involved.

In line with previous findings (Straube et al., 2011a) we found enhanced cortical activations for abstract (a) as opposed to concrete (c) utterances in the bilateral temporal lobes and in the left IFG for both, space as well as shape-related sentences (aSP > cSP and aSH > cSH). The interaction of perceptual category and abstractness in a more anterior part of the left IFG and inferior part of the pTL indicates that abstractness strongly influenced the neural processing of space and shape information. Only the effect of shape- vs. space-related information revealed activation in a single cluster of the left inferior occipital gyrus independent of abstractness (cSH > cSP ∩ aSH c> aSP). By contrast, the interaction resulted in enhanced BOLD responses in a more anterior part of the left IFG and inferior part of the pTL. Thus, we demonstrate the interaction of perceptual category and abstractness on the neural processing of speech accompanied by gestures. These data suggest a functional division of the pTL and left IFG being sensitive to the processing of both the level of abstractness and the type of categorical information. These imaging results further offer neural support for the traditional categorization of co-verbal gestures with regard to their content and abstractness (McNeill, 1992, 2005).

The imaging results for the abstract co-verbal gesture condition revealed BOLD enhancements in the left inferior frontal and the bilateral temporal regions, respectively. This finding is consistent with previous evidence of involvement of the left IFG and bilateral temporal lobes in the integration of gestures with abstract sentences (Kircher et al., 2009; Straube et al., 2009, 2011a). With regard to the underlying neuro-cognitive processes, we assume that the concrete visual gesture information (e.g., illustrating an arch of a bridge) is being interpreted in context of the abstract sentence meaning ("the politician builds a bridge to the next topic"). Thus, correspondence of gesture and sentence meaning must be identified and figurative components of speech and gesture must be translated from their literal/concrete meanings. To build this relation between speech and gesture information on the level of abstractness, additional online unification processes within the IFG seem to be relevant (Straube et al., 2011a). Such processes might be similar to those responsible for making inferences (e.g., Bunge et al., 2009, relational reasoning (e.g., Wendelken et al., 2008), the building of analogies (e.g., Luo et al., 2003; Bunge et al., 2005; Green et al., 2006; Watson and Chatterjee, 2012), and unification (Hagoort et al., 2009; Straube et al., 2011a). Those processes may also be involved in the comprehension of novel metaphoric or ambiguous communications and consistently activate the left IFG (Rapp et al., 2004, 2007; Stringaris et al., 2007; Chen et al., 2008; Cardillo et al., 2012). Consequently, enhanced neural responses in the fronto-temporal network may be evoked by the higher cognitive demand in an abstract metaphoric context which may have resulted in the recruitment of the left inferior frontal and middle temporal region (Kircher et al., 2009; Straube et al., 2011a).

Concrete speech accompanied by gestures revealed a pattern of enhanced BOLD responses in parahippocampal regions bilaterally as well as in the left superior occipital gyrus. Concrete co-verbal utterances such as, "the workman builds a bridge over the river," evokes a comparatively transparent connection/relation to a familiar everyday event. Accordingly, an experienced-based understanding of a scene may have resulted in the recruitment of the parahippocampal regions, whereas the direct imagery of concrete objects or actions may have resulted in enhanced neural responses in the left superior occipital region (Green et al., 2009) facilitating the understanding of the concrete co-verbal content.

The shape-related sentences accompanied by shape-related gestures revealed activations in the left middle occipital region. Similar to the activations found for the concrete condition (c > a), imagery of an experience-based perceptual representation resulted in the activations of the left occipital area. However, we did not observe common activation for the processing of spatial information in a concrete and abstract sentence context. Together these data do not support a universal neural processing of space and shape in a multimodal communication context.

By contrast, we found an interaction for perceptual category and abstractness, as spatial information on an abstract level (aSP) specifically (in contrast to all other conditions) activated a particular part of the left IFG and the left superior temporal region. This finding was robust and independent of both hand movement and speech duration. Thus, BOLD enhancements in these regions suggest that predominantly spatial information is processed differently in an abstract vs. concrete sentence context. Additional semantic information is retrieved from the left superior temporal region. The higher cognitive load together with the resulting enhanced effort with regard to information-specific abstract and spatial lexical retrieval may account for the recruitment of the fronto-temporal network. However, specific activation of the left IFG could also represent competition between meanings of spatial terms in the aSP condition, including at a minimum the concrete/literal and the abstract/metaphoric interpretations (Chatterjee, 2008; Chen et al., 2008).

For the processing of shape-related information we found common activation within the inferior temporal gyrus and the occipital lobe for concrete and abstract utterances, suggesting a common perceptual representation activated during comprehension of shape information. This perceptual representation probably compensated for the need of additional resources of the IFG and pTL, which were activated for space-related information in an abstract sentence context. Thus, this finding suggests that a concrete representation of shape is also activated in an abstract sentence context. This might have further facilitated the processing of the abstract representation of shape. For the processing of space-related information we found no common activation for concrete and abstract utterances, indicating different neural processing mechanism for both types of communications. The transformation of space-related gesture information in an abstract sentence context probably required higher order semantic processing mechanisms (Straube et al., 2011a) which probably inhibited the actual perceptual spatial representation of these gestures.

A limitation of this study is that the specific effects of gesture as well as integration processes cannot be disentangled. Distinguishing between speech and gesture was not the purpose of the current study. The problem with regard to the interpretation of our results for the main effect of abstractness, irrespective of perceptual category, might be that the activation patterns found for abstract speech accompanied by gestures in the left IFG and bilateral temporal lobes is produced by differences in the abstractness between the sentences, as demonstrated by several studies about metaphoric speech processing (Rapp et al., 2004, 2007; Eviatar and Just, 2006; Mashal et al., 2007, 2009; Nagels et al., 2013; Stringaris et al., 2007; Chen et al., 2008). However, in a previous study we observed increased activation in the left IFG for metaphoric co-verbal gestures in contrast to control sentences with the identical abstract semantic content (Kircher et al., 2009). Furthermore, there is evidence that activation of the left IFG is specifically related to the processing of novel and therefore unconventional metaphoric sentences (Rapp et al., 2004, 2007; Cardillo et al., 2012), in which abstract information must be interpreted online in terms of its non-literal meaning. However, the abstract sentences used in the current study were conventional and part of everyday communication, e.g., "The talk was on a high level." This is supported by our rating results, which revealed no differences between the conditions with regard to familiarity. Despite the fact, that we cannot exclude that differences between conditions might be explained by differences in difficulty due to our language manipulation (concrete vs. abstract), the lack of commonalities (e.g., Spa > SHa ∩ SPc > SHc) cannot be explained by these potential differences. The robustness of the imaging results in the aforementioned regions is further supported by the separate control analyses encompassing a carefully matched subset of paired (hand movements and speech length) stimuli.

A further limitation is that the distinction between space- and shape-related information in the current experiment is artificial and do not represent independent factors. Shape gestures include some spatial information. However, despite this intrinsic connection between space and shape, our data demonstrate that these perceptual categories can be distinguished by independent raters and produce distinct interacting activation patterns with regard to abstractness. Therefore, our data support the validity of this separation, which has been traditionally applied in terms of deictic or abstract deictic gestures (which refer to space) in contrast to iconic and metaphoric gestures (which rather refer to form or shape; e.g., McNeill, 1992).

With this study we demonstrate the interaction of perceptual category and abstractness in the neural processing of speech-gesture utterances. Besides abstractness, the type of information was relevant to the neural processing of speech accompanied by gestures. This finding illustrates the relevance of the interaction between language and cognition, which characterizes the complexity of natural interpersonal communication. Future studies should therefore consider the importance of perceptual type and abstractness for the interpretation of their imaging results. Our data suggest a functional subdivision of the pTL and left IFG with regard to the processing of space and shape-related information in an abstract sentence context. Such differences support the theoretically based traditional categorization of co-verbal gestures with regard to information type and abstractness (McNeill, 1992). Most likely the investigation of other types of co-verbal gestures will demonstrate further important differences in the processing of specific co-verbal gesture types, which will enlighten the fine-grained differences of processing mechanisms, which underlie the comprehension of multimodal natural communication.

ACKNOWLEDGMENTS

This research project is supported by a grant from the Interdisciplinary Center for Clinical Research "BIOMAT" (IZKF VV N68). Arne Nagels is supported by a grant form the "Deutsche Forschungsgemeinschaft" (DFG: Ki 588/6-1), Benjamin Straube

is supported by the BMBF (project no. 01GV0615). We thank Katharina Augustin, Bettina Freese and Simone Schröder for the preparation and evaluation of the stimulus material.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fnbeh.2013.00181/abstract

Table 7 | Brain areas sensitive for shape-related contents (independent of abstractness). Significance level (t-value), size of the respective activation cluster (No. voxels; number of voxels > 8) at $p < 0.005$ MC corrected for multiple comparisons. Coordinates are listed in MNI space. BA is the Brodmann area nearest to the coordinate and should be considered approximate. (cSP, concrete spatial; cSH, concrete shape; aSP, abstract spatial; aSH, abstract shape).

Table 8 | Brain areas sensitive for space-related contents (independent of abstractness). Significance level (t-value), size of the respective activation cluster (No. voxels; number of voxels > 8) at $p < 0.005$ MC corrected for multiple comparisons. Coordinates are listed in MNI space. BA is the Brodmann area nearest to the coordinate and should be considered approximate. (cSP, concrete spatial; cSH, concrete shape; aSP, abstract spatial; aSH, abstract shape).

Table 9 | Brain areas sensitive for abstractness (independent of content). Significance level (t-value), size of the respective activation cluster (No. voxels; number of voxels > 8) at $p < 0.005$ MC corrected for multiple comparisons. Coordinates are listed in MNI space. BA is the Brodmann area nearest to the coordinate and should be considered approximate. (cSP, concrete spatial; cSH, concrete shape; aSP, abstract spatial; aSH, abstract shape).

Table 10 | Brain areas sensitive for concreteness (independent of content). Significance level (t-value), size of the respective activation cluster (No. voxels; number of voxels > 8) at $p < 0.005$ MC corrected for multiple comparisons. Coordinates are listed in MNI space. BA is the Brodmann area nearest to the coordinate and should be considered approximate. (cSP, concrete spatial; cSH, concrete shape; aSP, abstract spatial; aSH, abstract shape).

REFERENCES

Arbib, M. A., (2008). From grasp to language: embodied concepts and the challenge of abstraction. *J. Physiol. Paris* 102, 4–20.

Bernardis, P., Bello, A., Pettenati, P., Stefanini, S., and Gentilucci, M. (2008). Manual actions affect vocalizations of infants. *Exp. Brain. Res.* 184, 599–603. doi: 10.1007/s00221-007-1256-x

Bunge, S. A., Helskog, E. H., and Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *Neuroimage* 46, 338–342. doi: 10.1016/j.neuroimage.2009.01.064

Bunge, S. A., Wendelken, C., Badre, D., and Wagner, A. D. (2005). Analogical reasoning and prefrontal cortex: evidence for separable retrieval and integration mechanisms. *Cereb. Cortex* 15, 239–249. doi: 10.1093/cercor/bhh126

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., and Chatterjee, A. (2012). From novel to familiar: tuning the brain for metaphors. *Neuroimage* 59, 3212–3221. doi: 10.1016/j.neuroimage.2011.11.079

Chatterjee, A. (2008). The neural organization of spatial thought and language. *Semin. Speech Lang.* 29, 226–238; quiz C226. doi: 10.1055/s-0028-1082886

Chen, E., Widick, P., and Chatterjee, A. (2008). Functional-anatomical organization of predicate metaphor processing. *Brain Lang.* 107, 194–202. doi: 10.1016/j.bandl.2008.06.007

Chica, A. B., Bartolomeo, P., and Valero-Cabré, A. (2011). Dorsal and ventral parietal contributions to spatial orienting in the human brain. *J. Neurosci.* 31, 8143–8149. doi: 10.1523/JNEUROSCI.5463-10.2010

Corballis, M.C. (2003). From mouth to hand: gesture, speech, and the evolution of right-handedness. *Behav. Brain Sci.* 26, 199–208; discussion 208–260. doi: 10.1017/S0140525X03000062

Corballis, M. C. (2009). Language as gesture. *Hum. Mov. Sci.* 28, 556–565. doi: 10.1016/j.humov.2009.07.003

Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain Lang.* 112, 25–35. doi: 10.1016/j.bandl.2009.02.002

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L., (2009). The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385. doi: 10.1016/j.cub.2009.01.017

Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I., and Small, S. L.,(2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Hum. Brain. Mapp.* 30, 3509–3526. doi: 10.1002/hbm.20774

Dick, A. S., Goldin-Meadow, S., Solodkin, A., and Small, S. L. (2012). Gesture in the developing brain. *Dev. Sci.* 15, 165–180. doi: 10.1111/j.1467-7687.2011.01100.x

Eviatar, Z., and Just, M. A., (2006). Brain correlates of discourse processing: an fMRI investigation of irony and conventional metaphor comprehension. *Neuropsychologia* 44, 2348–2359. doi: 10.1016/j.neuropsychologia.2006.05.007

Fischer, M. H., and Zwaan, R. A., (2008). Embodied language: a review of the role of the motor system in language comprehension. *Q. J. Exp. Psychol. (Hove).* 61, 825–850. doi: 10.1080/17470210701623605

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. (1998). Event-related fMRI: characterizing differential responses. *Neuroimage* 7, 30–40. doi: 10.1006/nimg.1997.0306

Gallese, V., and Lakoff, G., (2005). The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310

Gentilucci, M., Bernardis, P., Crisi, G., and Dalla Volta, R. (2006). Repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation. *J. Cogn. Neurosci.* 18, 1059–1074. doi: 10.1162/jocn.2006.18.7.1059

Gentilucci, M., and Corballis, M. C. (2006). From manual gesture to speech: a gradual transition. *Neurosci. Biobehav. Rev.* 30, 949–960. doi: 10.1016/j.neubiorev.2006.02.004

Gibbs, R. W., Jr., (1996). Why many concepts are metaphorical. *Cognition* 61, 309–319. doi: 10.1016/S0010-0277(96)00723-8

Gillebert, C. R., Mantini, D., Thijs, V., Sunaert, S., Dupont, P., and Vandenberghe, R. (2011). Lesion evidence for the critical role of the intraparietal sulcus in spatial attention. *Brain* 134, 1694–1709. doi: 10.1093/brain/awr085

Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., et al. (2009). Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. *Hum. Brain Mapp.* 30, 3309–3324. doi: 10.1002/hbm.20753

Green, A. E., Fugelsang, J. A., and Dunbar, K. N. (2006). Automatic activation of categorical and abstract analogical relations in analogical reasoning. *Mem. Cogn.* 34, 1414–1421. doi: 10.3758/BF03195906

Grill-Spector, K., Kourtzi, Z., and Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Res.* 41, 1409–1422. doi: 10.1016/S0042-6989(01)00073-6

Hagoort, P., Baggio, G., and Willems, R. (2009). "Semantic unification," in *The Cognitive Neurosciences* ed G. M. (Cambridge, MNA: MIT Press), 819–836.

Holle, H., Gunter, T. C., Ruschemeyer, S. A., Hennenlotter, A., and Iacoboni, M., (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage* 39, 2010–2024. doi: 10.1016/j.neuroimage.2007.10.055

Holle, H., Obleser, J., Rueschemeyer, S. A., and Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage* 49, 875–884. doi: 10.1016/j.neuroimage.2009.08.058

Holmes, V. M., and Rundle, M. (1985). The role of prior context in the comprehension of abstract and concrete sentences. *Psychol. Res.* 47, 159–171. doi: 10.1007/BF00309266

Hubbard, A. L., Wilson, S. M., Callan, D. E., and Dapretto, M. (2009). Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Hum. Brain Mapp.* 30, 1028–1037. doi: 10.1002/hbm.20565

Karnath, H. O., Rüter, J., Mandler, A., and Himmelbach, M. (2009). The anatomy of object recognition—visual form agnosia caused by medial occipitotemporal stroke. *J. Neurosci.* 29, 5854–5862. doi: 10.1523/JNEUROSCI.5192-08.2009

Kelly, S. D., Creigh, P., and Bartolotti, J., (2010). Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. *J. Cogn. Neurosci.* 22, 683–694. doi: 10.1162/jocn.2009.21254

Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., et al. (2009). Neural interaction of speech and gesture: differential activations of metaphoric co-verbal gestures. *Neuropsychologia* 47, 169–179. doi: 10.1016/j.neuropsychologia.2008.08.009

Koshino, H., Boese, G. A., and Ferraro, F. R. (2000). The relationship between cognitive ability and positive and negative priming in identity and spatial priming tasks. *J. Gen. Psychol.* 127, 372–382. doi: 10.1080/00221300009598591

Koshino, H., Carpenter, P. A., Keller, T. A., and Just, M. A. (2005). Interactions between the dorsal and the ventral pathways in mental rotation: an fMRI study. *Cogn. Affect. Behav. Neurosci.* 5, 54–66. doi: 10.3758/CABN.5.1.54

Kourtzi, Z., Erb, M., Grodd, W., and Bü lthoff, H. H. (2003). Representation of the perceived 3-D object shape in the human lateral occipital complex. *Cereb. Cortex* 13, 911–920. doi: 10.1093/cercor/13.9.911

Kourtzi, Z., and Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *J. Neurosci.* 20, 3310–3318.

Kourtzi, Z., and Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science* 293, 1506–1509. doi: 10.1126/science.1061133

Lakoff, G., (1987). *Women, Fire and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago.

Luo, Q., Perry, C., Peng, D., Jin, Z., Xu, D., Ding, G., et al. (2003). The neural substrate of analogical reasoning: an fMRI study. *Brain Res. Cogn. Brain Res.* 17, 527–534. doi: 10.1016/S0926-6410(03)00167-8

Mashal, N., Faust, M., Hendler, T., and Jung-Beeman, M., (2007). An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain Lang.* 100, 115–126. doi: 10.1016/j.bandl.2005.10.005

Mashal, N., Faust, M., Hendler, T., and Jung-Beeman, M., (2009). An fMRI study of processing novel metaphoric sentences. *Laterality* 14, 30–54. doi: 10.1080/13576500802049433

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought.* Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226514642.001.0001

McNeill, D., Cassell, J., and Levy, E. (1993a). Abstract deixis. *Semiotica* 15, 5–19.

McNeill, D., Justine, C., and Elena, L. (1993b). Abstract deixis. *Semiotica* 15, 5–19.

Nagels, A., Kauschke, C., Schrauf, J., Whitney, C., Straube, B., and Kircher, T., (2013). Neural substrates of figurative language during natural speech perception: an fMRI study. *Front. Behav. Neurosci.* 7:121. doi: 10.3389/fnbeh.2013.00121

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Ozyurek, A., and Kelly, S. D. (2007). Gesture, brain, and language. *Brain Lang.* 101, 181–184. doi: 10.1016/j.bandl.2007.03.006

Panis, S., Vangeneugden, J., Op de Beeck, H. P., and Wagemans, J. (2008). The representation of subordinate shape similarity in human occipitotemporal cortex. *J. Vis.* 8, 9.1–15. doi: 10.1167/8.10.9

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987. doi: 10.1038/nrn2277

Pulvermüller, F., and Fadiga, L., (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360. doi: 10.1038/nrn2811

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., and Kircher, T. T. (2004). Neural correlates of metaphor processing. *Brain Res. Cogn. Brain Res.* 20, 395–402. doi: 10.1016/j.cogbrainres.2004.03.017

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., and Kircher, T. T. (2007). Laterality in metaphor processing: lack of evidence from functional magnetic resonance imaging for the right hemisphere theory. *Brain Lang.* 100, 142–149. doi: 10.1016/j.bandl.2006.04.004

Rizzolatti, G., Ferrari, P. F., Rozzi, S., and Fogassi, L. (2006). The inferior parietal lobule: where action becomes perception. *Novartis Found. Symp.* 270, 129–140; discussion 140–125, 164–129.

Rizzolatti, G., Fogassi, L., and Gallese, V. (1997). Parietal cortex: from sight to action. *Curr. Opin. Neurobiol.* 7, 562–567. doi: 10.1016/S0959-4388(97)80037-2

Rizzolatti, G., and Matelli, M. (2003). Two different streams form the dorsal visual system: anatomy and functions. *Exp. Brain Res.* 153, 146–157. doi: 10.1007/s00221-003-1588-0

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., and Small, S. L. (2009). Gestures orchestrate brain networks for language understanding. *Curr. Biol.* 19, 661–667. doi: 10.1016/j.cub.2009.02.051

Slotnick, S. D., Moo, L. R., Segal, J. B., and Hart, J. Jr. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Res. Cogn. Brain Res.* 17, 75–82. doi: 10.1016/S0926-6410(03)00082-X

Straube, B., Green, A., Bromberger, B., and Kircher, T. (2011a). The differentiation of iconic and metaphoric gestures: common and unique integration processes. *Hum. Brain Mapp.* 32, 520–533. doi: 10.1002/hbm.21041

Straube, B., Green, A., Chatterjee, A., and Kircher, T. (2011b). Encoding social interactions: the neural correlates of true and false memories. *J. Cogn. Neurosci.* 23, 306–324. doi: 10.1162/jocn.2010.21505

Straube, B., Green, A., Jansen, A., Chatterjee, A., and Kircher, T. (2010). Social cues, mentalizing and the neural processing of speech accompanied by gestures. *Neuropsychologia* 48, 382–393. doi: 10.1016/j.neuropsychologia.2009.09.025

Straube, B., Green, A., Sass, K., and Kircher, T., (2013). Superior temporal sulcus disconnectivity during processing of metaphoric gestures in schizophrenia. *Schizophr Bull*. doi: 10.1093/schbul/sbt110. [Epub ahead of print].

Straube, B., Green, A., Weis, S., Chatterjee, A., and Kircher, T. (2009). Memory effects of speech and gesture binding: cortical and hippocampal activation in relation to subsequent memory performance. *J. Cogn. Neurosci.* 21, 821–836. doi: 10.1162/jocn.2009.21053

Stringaris, A. K., Medford, N. C., Giampietro, V., Brammer, M. J., and David, A. S. (2007). Deriving meaning: distinct neural mechanisms for metaphoric, literal, and non-meaningful sentences. *Brain Lang.* 100, 150–162. doi: 10.1016/j.bandl.2005.08.001

Tettamanti, M., and Moro, A. (2011). Can syntax appear in a mirror (system)? *Cortex* 48, 923–935. doi: 10.1016/j.cortex.2011.05.020

Watson, C. E., and Chatterjee, A. (2012). A bilateral frontoparietal network underlies visuospatial analogical reasoning. *Neuroimage* 59, 2831–2838. doi: 10.1016/j.neuroimage.2011.09.030

Wendelken, C., Nakhabenko, D., Donohue, S. E., Carter, C. S., and Bunge, S. A. (2008). "Brain is to thought as stomach is to ?" Investigating the role of rostrolateral prefrontal cortex in relational reasoning. *J. Cogn. Neurosci.* 20, 682–693. doi: 10.1162/jocn.2008.20055

Willems, R. M., and Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: a review. *Brain Lang.* 101, 278–289. doi: 10.1016/j.bandl.2007.03.004

Willems, R. M., Ozyurek, A., and Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* 17, 2322–2333. doi: 10.1093/cercor/bhl141

Willems, R. M., Ozyurek, A., and Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage* 47, 1992–2004. doi: 10.1016/j.neuroimage.2009.05.066

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 July 2013; accepted: 11 November 2013; published online: 18 December 2013.

Citation: Nagels A, Chatterjee A, Kircher T and Straube B (2013) The role of semantic abstractness and perceptual category in processing speech accompanied by gestures. Front. Behav. Neurosci. 7:181. doi: 10.3389/fnbeh.2013.00181

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 04 February 2014
doi: 10.3389/fnbeh.2014.00022

Toward a self-organizing pre-symbolic neural model representing sensorimotor primitives

***Junpei Zhong*[1,2]*, *Angelo Cangelosi*[3] *and Stefan Wermter*[1]**

[1] *Department of Computer Science, University of Hamburg, Hamburg, Germany*
[2] *School of Computer Science, University of Hertfordshire, Hatfield, UK*
[3] *School of Computing and Mathematics, University of Plymouth, Plymouth, UK*

Edited by:
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA

Reviewed by:
Matthew Schlesinger, Southern Illinois University, USA
Stefano Nolfi, Institute of Cognitive Sciences and Technologies CNR, Italy

***Correspondence:**
Junpei Zhong, Department of Computer Science, Knowledge Technology, University of Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany
e-mail: zhong@informatik.uni-hamburg.de

The acquisition of symbolic and linguistic representations of sensorimotor behavior is a cognitive process performed by an agent when it is executing and/or observing own and others' actions. According to Piaget's theory of cognitive development, these representations develop during the sensorimotor stage and the pre-operational stage. We propose a model that relates the conceptualization of the higher-level information from visual stimuli to the development of ventral/dorsal visual streams. This model employs neural network architecture incorporating a predictive sensory module based on an RNNPB (Recurrent Neural Network with Parametric Biases) and a horizontal product model. We exemplify this model through a robot passively observing an object to learn its features and movements. During the learning process of observing sensorimotor primitives, i.e., observing a set of trajectories of arm movements and its oriented object features, the pre-symbolic representation is self-organized in the parametric units. These representational units act as bifurcation parameters, guiding the robot to recognize and predict various learned sensorimotor primitives. The pre-symbolic representation also accounts for the learning of sensorimotor primitives in a latent learning context.

Keywords: pre-symbolic communication, sensorimotor integration, recurrent neural networks, parametric biases, horizontal product

1. INTRODUCTION

Although infants are not supposed to acquire the symbolic representational system at the sensorimotor stage, based on Piaget's definition of infant development, the preparation of language development, such as a pre-symbolic representation for conceptualization, has been set at the time when the infant starts babbling (Mandler, 1999). Experiments have shown that infants have established the concept of animate and inanimate objects, even if they have not yet seen the objects before (Gelman and Spelke, 1981). Similar phenomena also include the conceptualization of object affordances such as the conceptualization of containment (Bonniec, 1985). This conceptualization mechanism is developed at the sensorimotor stage to represent sensorimotor primitives and other object-affordance related properties.

During an infants' development at the sensorimotor stage, one way to learn affordances is to interact with objects using tactile perception, observe the object from visual perception and thus learn the causality relation between the visual features, affordance and movements as well as to conceptualize them. This learning starts with the basic ability to move an arm toward the visual-fixated objects in new-born infants (Von Hofsten, 1982), continues through object-directed reaching at the age of 4 months (Streri et al., 1993; Corbetta and Snapp-Childs, 2009), and can also be found during the object exploration of older infants (c.f. Ruff, 1984; Mandler, 1992). From these interactions leading to visual and tactile percepts, infants gain experience through the instantiated "bottom-up" knowledge about object affordances and sensorimotor primitives. Building on this, infants at the age of around 8–12 months gradually expand the concept of object features, affordances and the possible causal movements in the sensorimotor context (Gibson, 1988; Newman et al., 2001; Rocha et al., 2006). For instance, they realize that it is possible to pull a string that is tied to a toy car to fetch it instead of crawling toward it. An associative rule has also been built that connects conceptualized visual feature inputs, object affordance and the corresponding frequent auditory inputs of words, across various contexts (Romberg and Saffran, 2010). At this stage, categories of object features are particularly learned in different contexts due to their affordance-invariance (Bloom et al., 1993).

Therefore the integrated learning process of the object's features, movements according to the affordances, and other knowledge is a globally conceptualized process through visual and tactile perception. This conceptualized learning is a precursor of a pre-symbolic representation of language development. This learning is the process to form an abstract and simplified representation for information exchange and sharing[1]. To conceptualize from visual perception, it usually includes a planning process: first the speaker receives and segments visual knowledge in the perceptual flow into a number of states on the basis of different criteria, then the speaker selects essential elements, such as the units to be verbalized, and last the speaker constructs certain temporal perspectives when the events have to be anchored and linked (c.f. Habel and Tappe, 1999; von Stutterheim and Nuse, 2003). Assuming this planning process is distributed between ventral and dorsal streams, the conceptualization process should

[1]For comparison of conceptualization between engineering and language perspectives, see (Gruber and Olsen, 1994; Bowerman and Levinson, 2001).

also emerge from the visual information that is perceived in each stream, associating the distributed information in both streams. As a result, the candidate concepts of visual information are statistically associated with the input stimuli. For instance, they may represent a particular visual feature with a particular class of label (e.g., a particular visual stimuli with an auditory wording "circle") (Chemla et al., 2009). Furthermore, the establishment of such links also strengthens the high-order associations that generate predictions and generalize to novel visual stimuli (Yu, 2008). Once the infants have learned a sufficient number of words, they begin to detect a particular conceptualized cue with a specific kind of wording. At this stage, infants begin to use their own conceptualized visual "database" of known words to identify a novel meaning class and possibly to extend their wording vocabulary (Smith et al., 2002). Thus, this associative learning process enables the acquisition and the extension of the concepts of domain-specific information (e.g., features and movements in our experiments) with the visual stimuli.

This conceptualization will further result in a pre-symbolic way for infants to communicate when they encounter a conceptualized object and intend to execute a correspondingly conceptualized well-practised sensorimotor action toward that object. For example, behavioral studies showed that when 8-to-11-month-old infants are unable to reach and pick up an empty cup, they may point it out to the parents and execute an arm movement intending to bring it to their lips. The conceptualized shape of a cup reminds infants of its affordance and thus they can communicate in a pre-symbolic way. Thus, the emergence from the conceptualized visual stimuli to the pre-symbolic communication also gives further rise to the different periods of learning nouns and verbs in infancy development (c.f. Gentner, 1982; Tardif, 1996; Bassano, 2000). This evidence supports that the production of verbs and nouns are not correlated to the same modality in sensory perception: experiments performed by Kersten (1998) suggest that nouns are more related to the movement orientation caused by the intrinsic properties of an object, while verbs are more related to the trajectories of an object. Thus we argue that such differences of acquisitions in lexical classes also relate to the conceptualized visual ventral and dorsal streams. The finding is consistent with Damasio and Tranel (1993)'s hypothesis that verb generation is modulated by the perception of conceptualization of movement and its spatio-temporal relationship.

For this reason, we propose that the conceptualized visual information, which is a prerequisite for the pre-symbolic communication, is also modulated by perception in two visual streams. Although there have been studies of modeling the functional modularity in the development of ventral and dorsal streams (e.g., Jacobs et al., 1991; Mareschal et al., 1999), the bilinear models of visual routing (e.g., Olshausen et al., 1993; Memisevic and Hinton, 2007; Bergmann and von der Malsburg, 2011), in which a set of control neurons dynamically modifies the weights of the "what" pathway on a short time scale, or transform-invariance models (e.g., Földiák, 1991; Wiskott and Sejnowski, 2002) by encouraging the neurons to fire invariantly while transformations are performed in their input stimuli. However, a model that explains the development of conceptualization from both streams and results in an explicit representation of conceptualization of both streams while the visual stimuli is presented is still missing in the literature. This conceptualization should be able to encode the same category for information flows in both ventral and dorsal streams like "object files" in the visual understanding (Fields, 2011) so that they could be discriminated in different contexts during language development.

On the other hand, this conceptualized representation that is distributed in two visual streams is also able to predict the tendency of appearance of an action-oriented object in the visual field, which causes some sensorimotor phenomena such as object permanence (Tomasello and Farrar, 1986) showing the infants' attention usually is driven by the object's features and movements. For instance, when infants are observing the movement of the object, recording showed an increase of the looking times when the visual information after occlusion is violated in either surface features or location (Mareschal and Johnson, 2003). Also the words and sounds play a top–down role in the early infants' visual attention (Sloutsky and Robinson, 2008). This could hint at the different development stages of the ventral and dorsal streams and their effect on the conceptualized prediction mechanism in the infant's consciousness. Accordingly, the model we propose about the conceptualized visual information should also be able to explain the emergence of a predictive function in the sensorimotor system, e.g., the ventral stream attempts to track the object and the dorsal stream processes and predicts the object's spatial location, when the sensorimotor system is involved in an object interaction. We have been aware of that this build-in predictive function in a forward sensorimotor system is essential: neuroimaging research has revealed the existence of internal forward models in the parietal lobe and the cerebellum that predict sensory consequences from efference copies of motor commands (Kawato et al., 2003) and supports fast motor reactions (e.g., Hollerbach, 1982). Since the probable position and the movement pattern of the action should be predicted on a short time scale, sensory feedback produced by a forward model with negligible delay is necessary in this sensorimotor loop.

Particularly, the predictive sensorimotor model we propose is suitable to work as one of the building modules that takes into account the predictive object movement in a forward sensorimotor system to deal with object interaction from visual stimuli input as **Figure 1** shows. This system is similar to Wolpert et al. (1995)'s sensorimotor integration, but it includes an additional sensory estimator (the lower brown block) which takes into account the visual stimuli from the object so that it is able to predict the dynamics of both the end-effector (which is accomplished by the upper brown block) and the sensory input of the object. This object-predictive module is essential in a sensorimotor system to generate sensorimotor actions like tracking and avoiding when dealing with fast-moving objects, e.g., in ball sports. We also assert that the additional inclusion of forward models in the visual perception of the objects can explain some predictive developmental sensorimotor phenomena, such as object permanence.

In summary, we propose a model that establishes links between the development of ventral/dorsal visual streams and the emergence of the conceptualization in visual streams, which further leads to the predictive function of a sensorimotor system. To

validate this proof-of-concept model, we also conducted experiments in a simplified robotics scenario. Two NAO robots were employed in the experiments: one of them was used as a "presenter" and moved its arm along pre-programmed trajectories as motion primitives. A ball was attached at the end of the arm so that another robot could obtain the movement by tracking the ball. Our neural network was trained and run on the other NAO, which was called the "observer". In this way, the observer robot perceived the object movement from its vision passively, so that its network took the object's visual features and the movements into account. Though we could also use one robot and a human presenter to run the same tasks, we used two identical robots, due to the following reasons: (1) the object movement trajectories can be done by a pre-programmed machinery so that the types and parameters of it can be adjusted; (2) the use of two identical robots allows to interchange the roles of the presenter and observer in an easier manner. As other humanoid robots, a sensorimotor cycle that is composed of cameras and motors also exists in NAO robots. Although its physical configurations and parameters of sensory and motor systems are different from those in human beings' or other biological systems, our model only handles the pre-processed information extracted from visual stimuli. Therefore it is sufficient to serve as a neural model that is running in a robot CPU to explain the language development in the cortical areas.

2. MATERIALS AND METHODS

2.1. NETWORK MODEL

A similar forward model exhibiting sensory prediction for visual object perception has been proposed in our recently published work (Zhong et al., 2012b) where we suggested an RNN implementation of the sensory forward model. Together with a CACLA trained multi-layer network as a controller model, the forward model embodied in a robot receiving visual landmark percepts enabled a smooth and robust robot behavior. However, one drawback of this work was its inability to store multiple sets of spatial-temporal input–output mappings, i.e., the learning did not converge if there appeared several spatial-temporal mapping sequences in the training. Consequently, a simple RNN network was not able to predict different sensory percepts for different reward-driven tasks. Another problem was that it assumed only one visual feature appeared in the robot's visual field, and that was the only visual cue it could learn during development. To solve the first problem, we further augment the RNN with parametric bias (PB) units. They are connected like ordinary biases, but the internal values are also updated through back-propagation. Comparing to the generic RNN, the additional PB units in this network act as bifurcation parameters for the non-linear dynamics. According to Cuijpers et al. (2009), a trained RNNPB can successfully retrieve and recognize different types of pre-learned, non-linear oscillation dynamics. Thus, this bifurcation function can be regarded as an expansion of the storage capability of working-memory within the sensory system. Furthermore, it adds the generalization ability of the PB units, in terms of recognizing and generating non-linear dynamics. To tackle the second problem, in order to realize sensorimotor prediction behaviors such as object permanence, the model should be able to learn objects' features and object movements separately in the ventral and dorsal visual streams, as we have shown in Zhong et al. (2012a).

Merging these two ideas, in the context of sensorimotor integration in hand-object interaction, the PB units can be considered as a small set of high-level conceptualized units that describe various types of non-linear dynamics of visual percepts, such as features and movements. This representation is more related to the "natural prototypes" from visual perception, for instance, than a specific language representation (Rosch, 1973).

The development of PB units can also be seen as the pre-symbolic communication that emerges during sensorimotor learning. The conceptualization, on the other hand, could also result in the prediction of future visual percepts of moving objects in sensorimotor integration.

In this model (**Figure 2**), we propose a three-layer, horizontal product Elman network with PB units. Similar to the original

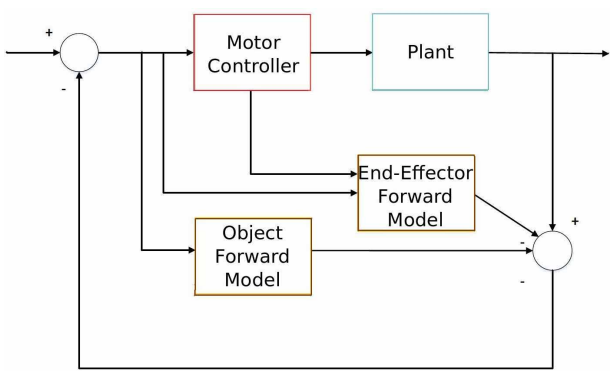


FIGURE 1 | Diagram of sensorimotor integration with the object interaction. The lower forward model predicts the object movement, while the upper forward model extracts the end-effector movement from sensory information in order to accomplish a certain task (e.g., object interaction).

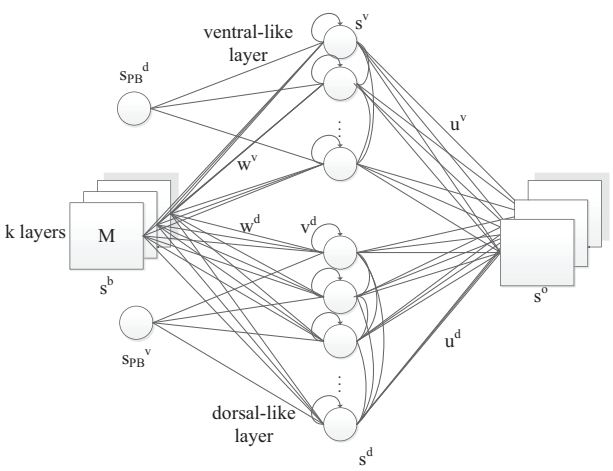


FIGURE 2 | The RNNPB-horizontal network architecture, where *k* layers represent *k* different types of features. Size of *M* indicates the transitional information of the object.

RNNPB model, the network is capable of being executed under three running modes, according to the pre-known conditions of inputs and outputs: learning, recognition and prediction. In learning mode, the representation of object features and movements are first encoded in the weights of both streams, while the bifurcation parameters with a smaller number of dimensions are encoded in the PB units. This is consistent with the emergence of the conceptualization at the sensorimotor stage of infant development.

Apart from the PB units, another novelty in the network is that the visual object information is encoded in two neural streams and is further conceptualized in PB units. Two streams share the same set of input neurons, where the coordinates of the object in the visual field are used as identities of the perceived images. The appearance of values in different layers represents different visual features: in our experiment, the color of the object detected by the yellow filter appears in the first layer whereas the color detected by the green filter appears in the second layer; the other layer remains zero. For instance, the input $((0,0),(x,y))$ represents a green object at (x,y) coordinates in the visual field. The hidden layer contains two independent sets of units representing dorsal-like "d" and ventral-like "v" neurons, respectively. These two sets of neurons are inspired by the functional properties of dorsal and ventral streams: (1) fast responding dorsal-like units predict object position and hence encode movements; (2) slow responding ventral-like units represent object features. The recurrent connection in the hidden layers also helps to predict movements in layer d and to maintain a persistent representation of an object's feature in layer v. The horizontal product brings both pathways together again in the output layer with one-step ahead predictions. Let us denote the output layer's input from layer d and layer v as x^d and x^v, respectively. The network output s^o is obtained via the horizontal product as

$$s^o = x^d \odot x^v \tag{1}$$

where $\odot$ indicates element-wise multiplication, so each pixel is defined by the product of two independent parts, i.e., for output unit k it is $s^o_k = x^d_k \cdot x^v_k$.

2.2. NEURAL DYNAMICS

We use $s^b(t)$ to represent the activation and $PB^{d/v}(t)$ to represent the activation of the dorsal/ventral PB units at the time-step t. In some of the following equations, the time-index t is omitted if all activations are from the same time-step. The inputs to the hidden units y^v_j in the ventral stream and y^d_j in the dorsal stream are defined as

$$y^d_l(t) = \sum_i s^b_i(t) w^d_{li} + \sum_{l'} s^d_l(t-1) v^d_{ll'} + \sum_{n_2} PB^v_{n_2}(t) \bar{w}^d_{ln_2} \tag{2}$$

$$y^v_j(t) = \sum_i s^b_i(t) w^v_{ji} + \sum_{j'} s^v_j(t-1) v^v_{jj'} + \sum_{n_1} PB^d_{n_1}(t) \bar{w}^v_{jn_1} \tag{3}$$

where w^d_{li}, w^v_{ji} represent the weighting matrices between dorsal/ventral layers and the input layer, $\bar{w}^d_{li}$, $\bar{w}^v_{ji}$ represent the weighting matrices between PB units and the two hidden layers, and $v^d_{ll'}$ and $v^v_{jj'}$ indicate the recurrent weighting matrices within the hidden layers.

The transfer functions in both hidden layers and the PB units all employ the sigmoid function recommended by LeCun et al. (1998),

$$s^{d/v}_{l/j} = 1.7159 \cdot \tanh\left(\frac{2}{3} y^{d/v}_{l/j}\right) \tag{4}$$

$$\mathrm{PB}^{d/v}_{n_1/n_2} = 1.7159 \cdot \tanh\left(\frac{2}{3} \rho^{d/v}_{n_1/n_2}\right) \tag{5}$$

where $\rho^{d/v}$ represent the internal values of the PB units.

The terms of the horizontal products of both pathways can be presented as follows:

$$x^v_k = \sum_j s^v_j u^v_{kj}; \quad x^d_k = \sum_l s^d_l u^d_{kl} \tag{6}$$

The output of the two streams composes a horizontal product for the network output as we defined in Equation (1).

2.2.1. Learning mode

The training progress is basically determined by the cost function:

$$C = \frac{1}{2} \sum_t^T \sum_k^N (s^b_k(t+1) - s^o_k(t))^2 \tag{7}$$

where $s^b_i(t+1)$ is the one-step ahead input (as well as the desired output), $s^o_k(t)$ is the current output, T is the total number of available time-step samples in a complete sensorimotor sequence and N is the number of output nodes, which is equal to the number of input nodes. Following gradient descent, each weight update in the network is proportional to the negative gradient of the cost with respect to the specific weight w that will be updated:

$$\Delta w_{ij} = -\eta_{ij} \frac{\partial C}{\partial w_{ij}} \tag{8}$$

where η_{ij} is the adaptive learning rate of the weights between neuron i and j, which is adjusted in every epoch (Kleesiek et al., 2013). To determine whether the learning rate has to be increased or decreased, we compute the changes of the weight $w_{i,j}$ in consecutive epochs:

$$\sigma_{i,j} = \frac{\partial C}{\partial w_{i,j}}(e-1) \frac{\partial C}{\partial w_{i,j}}(e) \tag{9}$$

The update of the learning rate is

$$\eta_{i,j}(e) = \begin{cases} \min(\eta_{i,j}(e-1) \cdot \xi^+, \eta_{\max}) & \text{if } \sigma_{i,j} > 0, \\ \max(\eta_{i,j}(e-1) \cdot \xi^-, \eta_{\min}) & \text{if } \sigma_{i,j} < 0, \\ \eta_{i,j}(e-1) & \text{else.} \end{cases}$$

where $\xi^+ > 1$ and $\xi^- < 1$ represent the increasing/decreasing rate of the adaptive learning rates, with $\eta_{\min}$ and $\eta_{\max}$ as lower

and upper bounds, respectively. Thus, the learning rate of a particular weight increases by ξ^+ to speed up the learning when the changes of that weight from two consecutive epochs have the same sign, and vice versa.

Besides the usual weight update according to back-propagation through time, the accumulated error over the whole time-series also contributes to the update of the PB units. The update for the i-th unit in the PB vector for a time-series of length T is defined as:

$$\rho_i(e+1) = \rho_i(e) + \gamma_i \sum_{t=1}^{T} \delta_{i,j}^{PB} \tag{10}$$

where δ^{PB} is the error back-propagated to the PB units, e is eth time-step in the whole time-series (e.g., epoch), γ_i is PB units' adaptive updating rate which is proportional to the absolute mean value of the back-propagation error at the i-th PB node over the complete time-series of length T:

$$\gamma_i \propto \frac{1}{T} \left\| \sum_{t=1}^{T} \delta_{i,j}^{PB} \right\| \tag{11}$$

The reason for applying the adaptive technique is that it was realized that the PB units converge with difficulty. Usually a smaller learning rate is used in the generic version of RNNPB to ensure the convergence of the network. However, this results in a trade-off in convergence speed. The adaptive learning rate is an efficient technique to overcome this trade-off (Kleesiek et al., 2013).

2.2.2. Recognition mode

The recognition mode is executed with a similar information flow as the learning mode: given a set of the spatio-temporal sequences, the error between the target and the real output is back-propagated through the network to the PB units. However, the synaptic weights remain constant and only the PB units will be updated, so that the PB units are self-organized as the pre-trained values after certain epochs. Assuming the length of the observed sequence is a, the update rule is defined as:

$$\rho_i(e+1) = \rho_i(e) + \gamma \sum_{t=T-a}^{T} \delta_{i,j}^{PB} \tag{12}$$

where δ^{PB} is the error back-propagated from a certain sensory information sequence to the PB units and γ is the updating rate of PB units in recognition mode, which should be larger than the adaptive rate γ_i at the learning mode.

2.2.3. Prediction mode

The values of the PB units can also be manually set or obtained from recognition, so that the network can generate the upcoming sequence with one-step prediction.

3. RESULTS

In this experiment, as we introduced, we examined this network by implementing it on two NAO robots. They were placed face-to-face in a rectangle box of 61.5 cm × 19.2 cm as shown in **Figure 3**. These distances were carefully adjusted so that the observer was able to keep track of movement trajectories in its visual field during all experiments using the images from the lower camera. The NAO robot has two cameras. We use the lower one to capture the images because its installation angle is more suitable to track the balls when they are held in the other NAO's hand.

Two 3.8 cm diameter balls with yellow/green color were used for the following experiments. The presenter consecutively held each of the balls to present the object interaction. The original image, received from the lower camera of the observer, was pre-processed with thresholding in HSV color-space and the coordinates of its centroid in the image moment were calculated. Here we only considered two different colors as the only feature to be encoded in the ventral stream, as well as two sets of movement trajectories encoded in the dorsal stream. Although we have only tested a few categories of trajectories and features, we believe the results can be extrapolated to multiple categories in future applications.

3.1. LEARNING

The two different trajectories are defined as below,
The *cosine* curve,

$$x = 12 \tag{13}$$

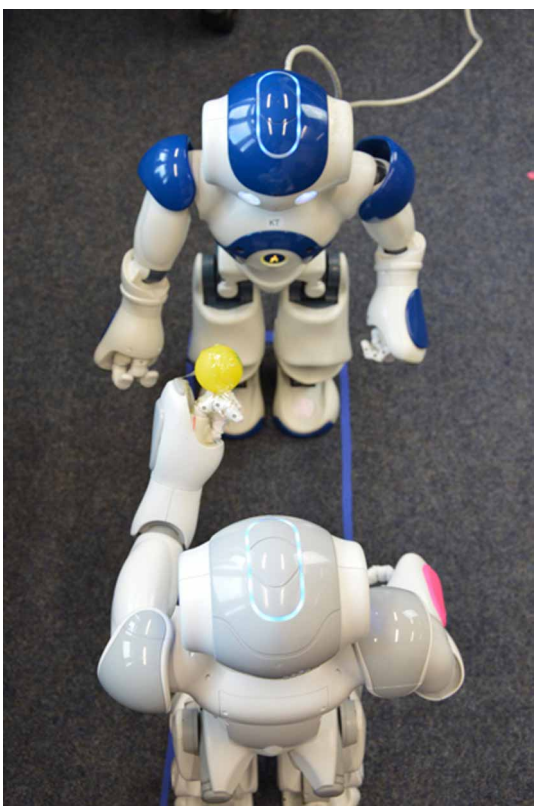


FIGURE 3 | Experimental Scenario: two NAOs are standing face-to-face with in a rectangle box.

$$y = 8 \cdot \left(-\frac{t}{2}\right) + 0.04 \tag{14}$$

$$z = 4 \cdot \cos(2t) + 0.10 \tag{15}$$

and the *square* curve,

$$x = 12 \tag{16}$$

$$y = \begin{cases} 0 & t \leq -\frac{3\pi}{4} \\ \frac{16}{\pi}t + 12 & -\frac{3\pi}{4} < t \leq -\frac{\pi}{4} \\ 8 & -\frac{\pi}{4} < t \leq \frac{\pi}{4} \\ -\frac{16}{\pi}t + 12 & \frac{\pi}{4} < t \leq \frac{3\pi}{4} \\ 0 & t > \frac{3\pi}{4} \end{cases} \tag{17}$$

$$z = \begin{cases} \frac{16}{\pi}t + 20 & t \leq -\frac{3\pi}{4} \\ 14 & -\frac{3\pi}{4} < t \leq -\frac{\pi}{4} \\ -\frac{16}{\pi}t + 10 & -\frac{\pi}{4} < t \leq \frac{\pi}{4} \\ 6 & \frac{\pi}{4} < t \leq \frac{3\pi}{4} \\ \frac{16}{\pi}t - 6 & t > \frac{3\pi}{4} \end{cases} \tag{18}$$

where the 3-dimension tuple (x, y, z) are the coordinates (centimeters) of the ball w.r.t the torso frame of the NAO presenter. t loops between $(-\pi, \pi]$. In each loop, we calculated 20 data points to construct trajectories with 4 s sleeping time between every two data points. Note that although we have defined the optimal desired trajectories, the arm movement was not ideally identical to the optimal trajectories due to the noisy position control of the end-effector of the robot. On the observer side, the (x, y) coordinates of the color-filtered moment of the ball in the visual field were recorded to form a trajectory with sampling time of 0.2 s. Five trajectories, in the form of tuple (x, y, z) w.r.t the torso frame of the NAO observer were recorded with each color and each curve, so total 20 trajectories were available for training.

In each training epoch, these trajectories, in the form of tuples, were fed into the input layer one after another for training, with the tuples of the next time-step serving as a training target. The parameters are listed in **Table 1**. The final PB values were examined after the training was done, and the values were shown in **Figure 4**. It can be seen that the first PB unit, along with the dorsal stream, was approximately self-organized with the color information, while the second PB unit, along with the ventral stream, was self-organized with the movement information.

Table 1 | Network parameters.

Parameters	Parameter's descriptions	Value
$\eta_{ventral}$	Learning rate in ventral stream	1.0×10^{-5}
η_{dorsal}	Learning rate in dorsal stream	1.0×10^{-3}
η_{max}	Maximum value of learning rate	1.0×10^{-1}
η_{min}	Minimum value of learning rate	1.0×10^{-7}
M_γ	Proportionality constant of PB units updating rate	1.0×10^{-2}
n_1	Size of PB unit 1	1
n_2	Size of PB unit 2	1
n_v	Size of ventral-like layer	50
n_d	Size of dorsal-like layer	50
ξ^-	Decreasing rate of learning rate	0.999999
ξ^+	Increasing rate of learning rate	1.000001

3.2. RECOGNITION

Another four trajectories were presented in the recognition experiment, in which the length of the sliding-window is equal to the length of the whole time-series, i.e., $T = a$ in Equation (12). The update of the PB units were shown in **Figure 5**. Although we used the complete time-series sequence for the recognition, it should also be possible to use only part of the sequence, e.g., through the sliding-window approach with a smaller number of a to fulfil the real-time requirement in the future.

3.3. PREDICTION

In this simulation, the obtained PB units from the previous recognition experiment were used to generate the predicted movements using the prior knowledge of a specific object. Then, the one-step prediction from the output units were again applied to the input at the next time-step, so that the whole time-series corresponding to the object's movements and features were obtained. **Figure 6** presents the comparisons between the true values (the same as used in recognition) and the predicted ones.

From **Figure 6** and **Table 2**, it can be observed that the estimation was biased quite largely to the true value within the first few time-steps, as the RNN needs to accumulate enough input values to access its short-term memory. However, the error became smaller and it kept track of the true value in the following time-steps. Considering that the curves are automatically generated given the PB units and the values at the first time-step, the error between the true values and the estimated ones are acceptable. Moreover, this result show clearly that the conceptualization affects the (predictive) visual perception.

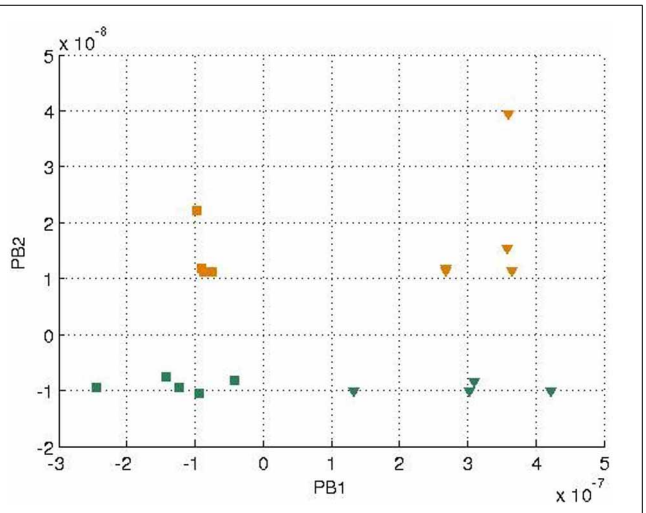


FIGURE 4 | Values of two sets of PB units in the two streams after training. The square markers represent those PB units after the *square* curves training and the triangle markers represent those of the *cosine* curves training. The colors of the markers, yellow and green, represent the colors of the balls used for training.

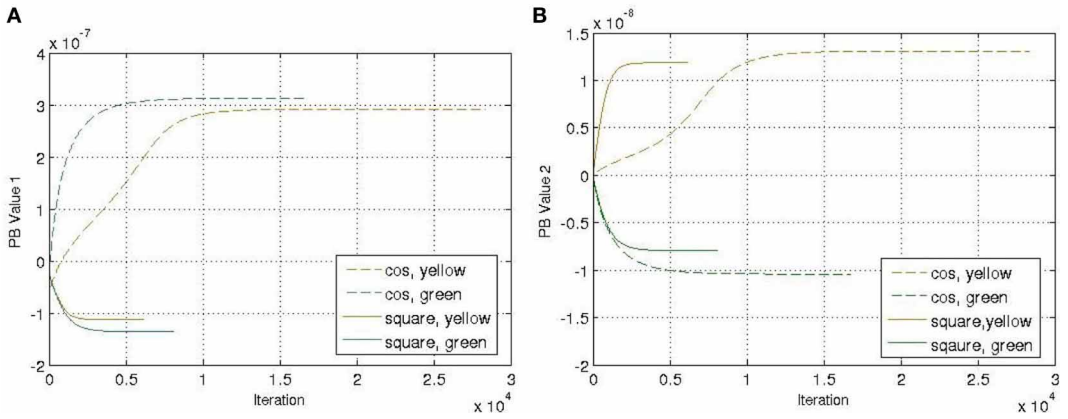


FIGURE 5 | Update of the PB values while executing the recognition mode. (A) PB value 1. **(B)** PB value 2.

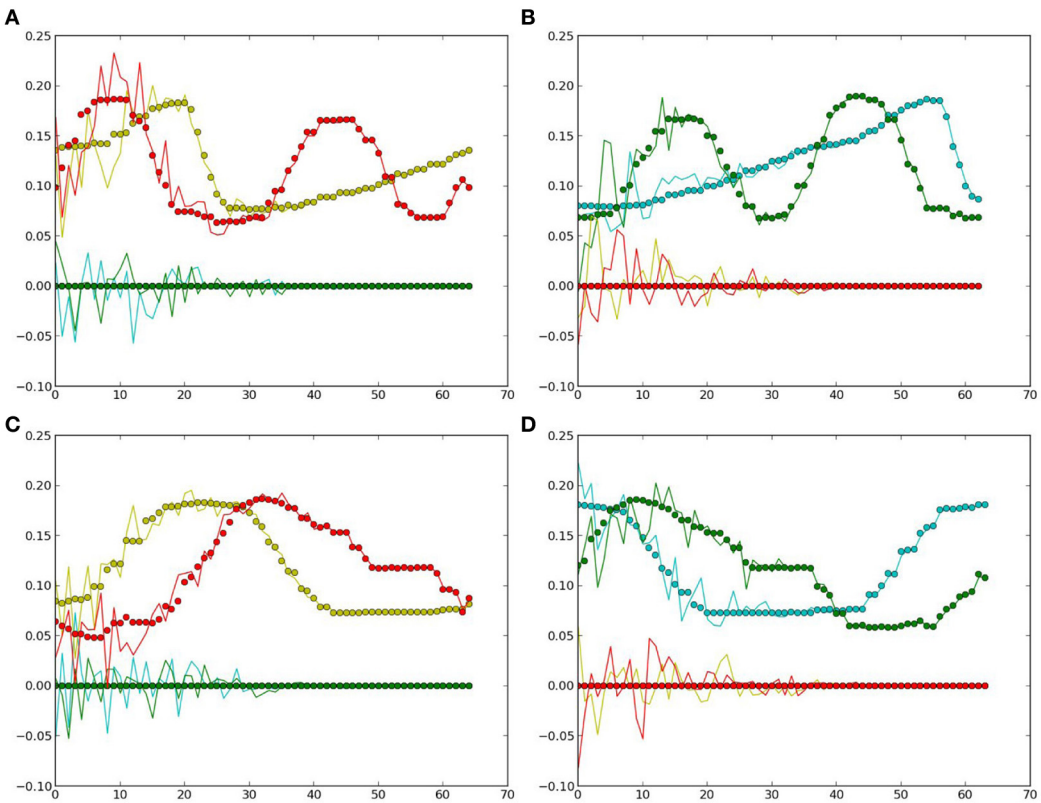


FIGURE 6 | Generated values: the dots denote the true values for comparison, curves show the estimated ones. Yellow and red colors represent the values of the two neurons in the first layer (yellow), the colors green and clan represent those in the second layer (green). **(A)** Cosine curve, yellow ball. **(B)** Cosine curve, green ball. **(C)** Square curve, yellow ball. **(D)** Square curve, green ball.

3.4. GENERALIZATION IN RECOGNITION

To testify whether our new computational model has the generalization ability as Cuijpers et al. (2009) proposed, we recorded another set of sequences of a circle trajectory. The trajectory is defined as:

$$x = 12 \tag{19}$$

$$y = 4 \cdot \sin(2t) + 0.04 \tag{20}$$

$$z = 4 \cdot \cos(2t) + 0.10 \tag{21}$$

The yellow and green balls were still used. We ran the recognition experiment again with the weight previously trained. The update of the PB units were shown in **Figure 7**. Comparing to **Figure 4**,

we can observe that the positive and negative signs of PB values are similar as the square trajectory. This is probably because the visual perception of circle and square movements have more similarities than those between circle and cosine movements.

3.5. PB REPRESENTATION WITH DIFFERENT SPEEDS

We further generated 20 trajectories with the same data functions (Equations 13–18) but with a slower sampling time. In other words, the movement of the balls seemed to be faster with robot's observation. The final PB values after training were shown in **Figure 8**.

It can be seen that generally the PB values were smaller comparing to **Figure 4**, which was probably because there was less error being propagated during training. Moreover, the corresponding PB values corresponding to colors (green and yellow) and movements (cosine and square) were interchanged within the same PB unit (i.e., along the same axis) due to the difference of random initial parameters of the network. But the PB unit along with the dorsal stream still encoded color information, while the PB unit along with ventral stream encoded movement information. The network was still able to show properties of spatio-temporal sequences data in the PB units' representation.

4. DISCUSSION

4.1. NEURAL DYNAMICS

An advancement of the HP-RNN model is that it can learn and encode the "what" and "where" information separately in two streams (more specifically, in two hidden layers). Both streams are connected through horizontal products, which means fewer connections than full multiplication (as the conventional bilinear model) (Zhong et al., 2012a). In this paper, we further augmented the HP-RNN with the PB units. One set of units, connecting to one visual stream, reflects the dynamics of sequences in the other stream. This is an interesting result since it shows the neural dynamics in the hybrid combination of the RNNPB units and the horizontal product model. Taking the dorsal-like hidden layer for example, the error of the attached PB units is

$$\delta_{n_2}^{PB} = \sum_l \delta_l^d \cdot f'(\rho_{n_2}^v) \cdot \bar{w}_{ln_2}^d \tag{22}$$

$$= \sum_l \left[\sum_k (s_k^b - s_k^o) g'(s_k^o) u_{kl}^d f'(s_l^v) \right] f'(\rho_{n_2}^v) \cdot \bar{w}_{ln_2}^d \tag{23}$$

where $g'(\cdot)$ and $f'(\cdot)$ are the derivatives of the linear and sigmoid transfer functions. Since we have the linear output, according to the definition of the horizontal product, the equation becomes,

$$\delta_{n_2}^{PB} = \sum_l \left[\sum_k (s_k^b - s_k^o) \odot x_k^v \cdot u_{kl}^d f'(s_l^d) \right] f'(\rho_{n_2}^v) \cdot \bar{w}_{ln_2}^d \tag{24}$$

The update of the internal values of the PB units becomes

$$\rho_{n_2}^v(e+1) = \rho_i^v(e) + \gamma \sum_{t=T-a}^{T} \delta_{n_2}^{PB}(t) \tag{25}$$

$$= \rho_i^v(e) + \gamma \sum_{t=T-a}^{T} \left\{ \sum_l \left[\sum_k (s_k^b(t) - s_k^o(t)) \odot x_k^v(t) \cdot u_{kl}^d(t) f'(s_l^d) \right] f'(\rho_{n_2}^v(t)) \cdot \bar{w}_{ln_2}^d(t) \right\} \tag{26}$$

where the $x_k^v(t)$ term refers to the contribution of the weighted summation from the ventral-like layer at time t. Note that the term $f'(\rho_{n_2}^v(t))$ is actually constant within one epoch and it is only updated after each epoch with a relatively small updating rate. Therefore, from the experimental perspective, given the same

Table 2 | Prediction error.

Error of outputs	Unit 1	Unit 2	Unit 3	Unit 4
Cosine, yellow	2.28×10^{-4}	8.09×10^{-5}	7.29×10^{-4}	8.63×10^{-4}
Cosine, green	8.34×10^{-4}	7.04×10^{-4}	1.50×10^{-4}	2.01×10^{-4}
Square, yellow	3.91×10^{-4}	9.64×10^{-5}	1.74×10^{-3}	3.23×10^{-4}
Square, green	1.40×10^{-3}	3.27×10^{-4}	3.54×10^{-4}	2.60×10^{-4}

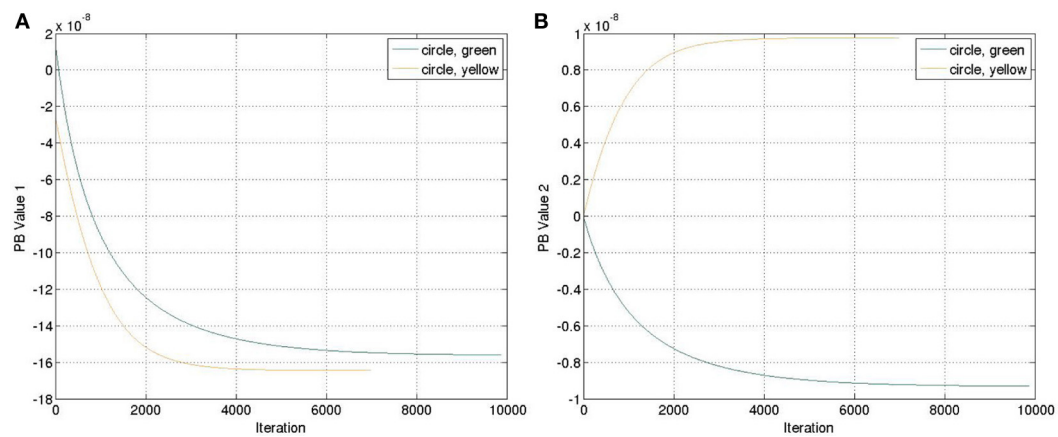


FIGURE 7 | Update of the PB values while executing the recognition mode with an untrained feature (circle). (A) PB value 1. **(B)** PB value 2.

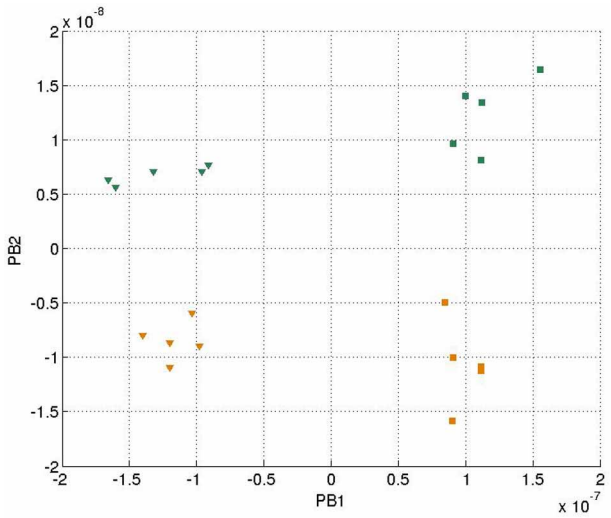


FIGURE 8 | Values of two sets of PB units in the two streams after training with faster speed. The representation of the markers is the same as **Figure 4**.

object movement but different object features, the difference of the PB values mostly reflects the dynamic changes in the hidden layer of the ventral stream. The same holds for the PB units attached to the ventral-like layer. This brief analysis shows the PB units for one modularity in RNNPB networks with horizontal product connections, effectively accumulating the non-linear dynamics of other modularities.

4.2. CONCEPTUALIZATION IN VISUAL PERCEPTION

The visual conceptualization and perception are intertwined processes. As experiments from Schyns and Oliva (1999) show, when the visual observation is not clear, the brain automatically extrapolates the visual percept and updates the categorization labels on various levels according to what has been gained from the visual field. On the other hand, this conceptualization also affects the immediate visual perception in a top–down predictive manner. For instance, the identity conceptualization of a human face predictively spreads conceptualizations in other levels (e.g., face emotion). This top–down process propagates from object identity to other local conceptualizations, such as object affordance, motion, edge detection and other processes at the early stages of visual processing. This can be tested by classic illusions, such as "the goblet illusion," where perception depends largely on top–down knowledge derived from past experiences rather than direct observation. This kind of illusion may be explained by the error in the first few time steps of the prediction experiment of our model. Therefore, our model to some extent also demonstrates the integrated process between the conceptualization and the spatio-temporal visual perception. This top–down predictive perception may also arouse other visual based predictive behaviors such as object permanence.

Particularly, the PB units act as a high-level conceptualization representation, which is continuously updated with the partial sensory information perceived in a short-time scale. The prediction process of the RNNPB is assisted by the conceptualized PB units of visual perception, which is identical to the integration conceptualization and (predictive) visual perception. This is the reason why PB units were not processed as a binary representation, as Ogata et al. (2007) did for human-robot-interaction; the original values of PB units are more accurate in generating the prediction of the next time-step and performing generalization tasks. As we mentioned, this model is merely a proof-of-concept model that bridges conceptualized visual streams and sensorimotor prediction. For more complex tasks, besides expanding of the network size as we mentioned, more complex networks that are capable of extracting and predicting higher-level spatio-temporal structures (e.g., predictive recurrent networks owning large learning capacity by Tani and colleagues: Yamashita and Tani, 2008; Murata et al., 2013) can be also applied. It should be interesting to further investigate the functional modularity representation of these network models when they are interconnected with horizontal product too.

Furthermore, the neuroscience basis that supports this paper, in the context of the mirror neuron system based on object-oriented-actions (grasping), can be stated as the "data-driven" models such as MNS (Oztop and Arbib, 2002) and MNS2 (Bonaiuto et al., 2007; Bonaiuto and Arbib, 2010), although the main hypothesis in our model is not taken from the mirror neuron system theory. In the MNS review paper by Oztop et al. (2006), the action generation mode of the RNNPB model was considered to be excessive as there has no evidence yet to show that the mirror neuron system participates in action generation. However, in our model the generation mode has a key role of conceptualized PB units in the sensorimotor integration of object interaction. Nevertheless, the similar network architecture (RNNPB) used in modeling mirror neurons (Tani et al., 2004) and our pre-symbolic sensorimotor integration models may imply a close relationship between language (pre-symbolic) development, object-oriented actions, and the mirror neuron theory.

5. CONCLUSION

In this paper a recurrent network architecture integrating the RNNPB model and the horizontal product model has been presented, which sheds light on the feasibility of linking the conceptualization of ventral/dorsal visual streams, the emergence of pre-symbol communication, and the predictive sensorimotor system.

Based on the horizontal product model, here the information in the dorsal and ventral streams is separately encoded in two network streams and the predictions of both streams are brought together via the horizontal product while the PB units act as a conceptualization of both streams. These PB units allow for storing multiple sensory sequences. After training, the network is able to recognize the pre-learned conceptualized information and to predict the up-coming visual perception. The network also shows robustness and generalization abilities. Therefore, our approach offers preliminary concepts for a similar development of conceptualized language in pre-symbolic communication and further in infants' sensorimotor-stage learning.

ACKNOWLEDGMENTS

The authors thank Sven Magg, Cornelius Weber, Katja Kösters as well as reviewers (Matthew Schlesinger, Stefano Nolfi) for improvement of the paper, Erik Strahl for technical support and NAO assistance in Hamburg and Torbjorn Dahl for the generous allowance to use the NAOs in Plymouth.

FUNDING

This research has been partly supported by the EU projects RobotDoC under grand agreement 235065 ROBOT-DOC, KSERA under grant agreement n° 2010-248085, POETICON++ under grant agreement 288382, ALIZ-E under grant agreement 248116 and UK EPSRC project BABEL.

REFERENCES

Bassano, D. (2000). Early development of nouns and verbs in french: exploring the interface between lexicon and grammar. *J. Child Lang.* 27, 521–559. doi: 10.1017/S0305000900004396

Bergmann, U., and von der Malsburg, C. (2011). Self-organization of topographic bilinear networks for invariant recognition. *Neural Comput.* 23, 2770–2797. doi: 10.1162/NECO_a_00195

Bloom, L., Tinker, E., and Margulis, C. (1993). The words children learn: evidence against a noun bias in early vocabularies. *Cogn. Dev.* 8, 431–450. doi: 10.1016/S0885-2014(05)80003-6

Bonaiuto, J., and Arbib, M. (2010). Extending the mirror neuron system model, II: what did I just do? A new role for mirror neurons. *Biol. Cybern.* 102, 341–359. doi: 10.1007/s00422-010-0371-0

Bonaiuto, J., Rosta, E., and Arbib, M. (2007). Extending the mirror neuron system model, I. *Biol. Cybern.* 96, 9–38. doi: 10.1007/s00422-006-0110-8

Bonniec, P. (1985). From visual-motor anticipation to conceptualization: reaction to solid and hollow objects and knowledge of the function of containment. *Infant Behav. Dev.* 8, 413–424. doi: 10.1016/0163-6383(85)90005-0

Bowerman, M., and Levinson, S. (2001). *Language Acquisition and Conceptual Development*, Vol. 3. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620669

Chemla, E., Mintz, T., Bernal, S., and Christophe, A. (2009). Categorizing words using "frequent frames": what cross-linguistic analyses reveal about distributional acquisition strategies. *Dev. Sci.* 12, 396–406. doi: 10.1111/j.1467-7687.2009.00825.x

Corbetta, D., and Snapp-Childs, W. (2009). Seeing and touching: the role of sensory-motor experience on the development of infant reaching. *Infant Behav. Dev.* 32, 44–58. doi: 10.1016/j.infbeh.2008.10.004

Cuijpers, R., Stuijt, F., and Sprinkhuizen-Kuyper, I. (2009). "Generalisation of action sequences in RNNPB networks with mirror properties," in *Proceedings of the European Symposium on Neural Networks (ESANN)* (Bruges, Belgium).

Damasio, A. R., and Tranel, D. (1993). Nouns and verbs are retrieved with differently distributed neural systems. *Proc. Natl. Acad. Sci. U.S.A.* 90, 4957–4960. doi: 10.1073/pnas.90.11.4957

Fields, C. (2011). Trajectory recognition as the basis for object individuation: a functional model of object file instantiation and object-token encoding. *Front. Psychol.* 2:49. doi: 10.3389/fpsyg.2011.00049

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Comput.* 3, 194–200. doi: 10.1162/neco.1991.3.2.194

Gelman, R., and Spelke, E. (1981). "The development of thoughts about animate and inanimate objects: implications for research on social cognition," in *Social Cognitive Development: Frontiers and Possible Futures*, eds J. H. Flavell and L. Ross (Cambridge: Cambridge University Press), 43–66.

Gentner, D. (1982). "Why nouns are learned before verbs: linguistic relativity versus natural partitioning," in *Language Development*, Vol. 2, ed S. Kuczaj (Hillsdale, NJ: Lawrence Eribaum), 301–334.

Gibson, E. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annu. Rev. Psychol.* 39, 1–42. doi: 10.1146/annurev.ps.39.020188.000245

Gruber, T., and Olsen, G. (1994). An ontology for engineering mathematics. *Knowl. Represent. Reason.* 94, 258–269.

Habel, C., and Tappe, H. (1999). "Processes of segmentation and linearization in describing events," in *Representations and Processes in Language Production*, eds C. von Stutterheim and R. Klabunde (Wiesbaden: Deutscher Universitaetsverlag), 43–66.

Hollerbach, J. M. (1982). Computers, brains and the control of movement. *Trends Neurosci.* 5, 189–192. doi: 10.1016/0166-2236(82)90111-4

Jacobs, R., Jordan, M., and Barto, A. (1991). Task decomposition through competition in a modular connectionist architecture: the what and where vision tasks. *Cogn. Sci.* 15, 219–250. doi: 10.1207/s15516709cog1502_2

Kawato, M., Kuroda, T., Imamizu, H., Nakano, E., Miyauchi, S., and Yoshioka, T. (2003). Internal forward models in the cerebellum: fMRI study on grip force and load force coupling. *Prog. Brain Res.* 142, 171–188. doi: 10.1016/S0079-6123(03)42013-X

Kersten, A. W. (1998). An examination of the distinction between nouns and verbs: associations with two different kinds of motion. *Mem. Cogn.* 26, 1214–1232. doi: 10.3758/BF03201196

Kleesiek, J., Badde, S., Wermter, S., and Engel, A. K. (2013). "Action-driven perception for a humanoid," in *Agents and Artificial Intelligence*, eds J. Filipe and A. Fred (Berlin: Springer), 83–99. doi: 10.1007/978-3-642-36907-0_6

LeCun, Y., Bottou, L., Orr, G. B., and Müller, K. (1998). "Efficient backprop," in *Neural Networks: Tricks of the Trade*, eds G. B. Orr and K.-R. Müller (Berlin: Springer), 9–50. doi: 10.1007/3-540-49430-8_2

Mandler, J. M. (1992). The foundations of conceptual thought in infancy. *Cogn. Dev.* 7, 273–285. doi: 10.1016/0885-2014(92)90016-K

Mandler, J. M. (1999). "Preverbal representation and language," in *Language and Space*, eds P. Bloom, M. A. Peterson, L. Nadel, M. F. Garrett (Cambridge, MA: The MIT Press), 365.

Mareschal, D., and Johnson, M. H. (2003). The what and where of object representations in infancy. *Cognition* 88, 259–276. doi: 10.1016/S0010-0277(03)00039-8

Mareschal, D., Plunkett, K., and Harris, P. (1999). A computational and neuropsychological account of object-oriented behaviours in infancy. *Dev. Sci.* 2, 306–317. doi: 10.1111/1467-7687.00076

Memisevic, R., and Hinton, G. (2007). "Unsupervised learning of image transformations," in *2007 IEEE Conference on Computer Vision and Pattern Recognition* (Minneapolis, MN), 1–8. doi: 10.1109/CVPR.2007.383036

Murata, S., Namikawa, J., Arie, H., Sugano, S., and Tani, J. (2013). Learning to reproduce fluctuating time series by inferring their time-dependent stochastic properties: application in robot learning via tutoring. *IEEE Trans. Auton. Ment. Dev.* 5, 298–310. doi: 10.1109/TAMD.2013.2258019

Newman, C., Atkinson, J., and Braddick, O. (2001). The development of reaching and looking preferences in infants to objects of different sizes. *Dev. Psychol.* 37, 561. doi: 10.1037/0012-1649.37.4.561

Ogata, T., Matsumoto, S., Tani, J., Komatani, K., and Okuno, H. (2007). "Human-robot cooperation using quasi-symbols generated by RNNPB model," in *2007 IEEE International Conference on Robotics and Automation* (Roma), 2156–2161. doi: 10.1109/ROBOT.2007.363640

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

Oztop, E., and Arbib, M. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biol. Cybernet.* 87, 116–140. doi: 10.1007/s00422-002-0318-1

Oztop, E., Kawato, M., and Arbib, M. (2006). Mirror neurons and imitation: a computationally guided review. *Neural Netw.* 19, 254–271. doi: 10.1016/j.neunet.2006.02.002

Rocha, N., Silva, F., and Tudella, E. (2006). The impact of object size and rigidity on infant reaching. *Infant Behav. Dev.* 29, 251–261. doi: 10.1016/j.infbeh.2005.12.007

Romberg, A., and Saffran, J. (2010). Statistical learning and language acquisition. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 906–914. doi: 10.1002/wcs.78

Rosch, E. (1973). Natural categories. *Cogn. Psychol.* 4, 328–350. doi: 10.1016/0010-0285(73)90017-0

Ruff, H. A. (1984). Infants' manipulative exploration of objects: effects of age and object characteristics. *Dev. Psychol.* 20, 9. doi: 10.1037/0012-1649.20.1.9

Schyns, P. G., and Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition* 69, 243–265. doi: 10.1016/S0010-0277(98)00069-9

Sloutsky, V. M., and Robinson, C. W. (2008). The role of words and sounds in infants' visual processing: from overshadowing to attentional tuning. *Cogn. Sci.* 32, 342–365. doi: 10.1080/03640210701863495

Smith, L., Jones, S., Landau, B., Gershkoff-Stowe, L., and Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychol. Sci.* 13, 13–19. doi: 10.1111/1467-9280.00403

Streri, A., Pownall, T., and Kingerlee, S. (1993). *Seeing, Reaching, Touching: The Relations Between Vision and Touch in Infancy*. Oxford: Harvester Wheatsheaf.

Tani, J., Ito, M., and Sugita, Y. (2004). Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB. *Neural Netw.* 17, 1273–1289. doi: 10.1016/j.neunet.2004.05.007

Tardif, T. (1996). Nouns are not always learned before verbs: evidence from mandarin speakers' early vocabularies. *Dev. Psychol.* 32, 492. doi: 10.1037/0012-1649.32.3.492

Tomasello, M., and Farrar, M. J. (1986). Object permanence and relational words: a lexical training study. *J. Child Lang.* 13, 495–505. doi: 10.1017/S030500090000684X

Von Hofsten, C. (1982). Eye–hand coordination in the newborn. *Dev. Psychol.* 18, 450. doi: 10.1037/0012-1649.18.3.450

von Stutterheim, C., and Nuse, R. (2003). Processes of conceptualization in language production: language-specific perspectives and event construal. *Linguistics* 41, 851–882. doi: 10.1515/ling.2003.028

Wiskott, L., and Sejnowski, T. (2002). Slow feature analysis: unsupervised learning of invariances. *Neural Comput.* 14, 715–770. doi: 10.1162/089976602317318938

Wolpert, D., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882. doi: 10.1126/science.7569931

Yamashita, Y., and Tani, J. (2008). Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment. *PLoS Comput. Biol.* 4:e1000220. doi: 10.1371/journal.pcbi.1000220

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Lang. Learn. Dev.* 4, 32–62. doi: 10.1080/15475440701739353

Zhong, J., Weber, C., and Wermter, S. (2012a). "Learning features and predictive transformation encoding based on a horizontal product model," in *Artificial Neural Networks and Machine Learning, ICANN*, eds A. E. P. Villa, W. Duch, P. Érdi, F. Masulli, and G. Palm (Berlin: Springer), 539–546.

Zhong, J., Weber, C., and Wermter, S. (2012b). A predictive network architecture for a robust and smooth robot docking behavior. *Paladyn* 3, 172–180. doi: 10.2478/s13230-013-0106-8

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 July 2013; accepted: 14 January 2014; published online: 04 February 2014. Citation: Zhong J, Cangelosi A and Wermter S (2014) Toward a self-organizing pre-symbolic neural model representing sensorimotor primitives. Front. Behav. Neurosci. ***8****:22. doi: 10.3389/fnbeh.2014.00022*

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 25 March 2014
doi: 10.3389/fnbeh.2014.00097

Temporal relation between top-down and bottom-up processing in lexical tone perception

Lan Shuai*[1]* and Tao Gong*[2]*

[1] *Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA*
[2] *Department of Linguistics, University of Hong Kong, Hong Kong, China*

Edited by:
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA

Reviewed by:
Ryan Giuliano, University of Oregon, USA
I-Fan Su, The University of Hong Kong, Hong Kong
Wentao Gu, Nanjing Normal University, China

****Correspondence:***
Lan Shuai, Department of Electrical and Computer Engineering, Johns Hopkins University, North Charles Street 3400, Baltimore, MD 21218, USA
e-mail: lshuai@jhu.edu;
Tao Gong, Department of Linguistics, University of Hong Kong, Pokfulam Road, Hong Kong, China
e-mail: gtojty@gmail.com

Speech perception entails both top-down processing that relies primarily on language experience and bottom-up processing that depends mainly on instant auditory input. Previous models of speech perception often claim that bottom-up processing occurs in an early time window, whereas top-down processing takes place in a late time window after stimulus onset. In this paper, we evaluated the temporal relation of both types of processing in lexical tone perception. We conducted a series of event-related potential (ERP) experiments that recruited Mandarin participants and adopted three experimental paradigms, namely dichotic listening, lexical decision with phonological priming, and semantic violation. By systematically analyzing the lateralization patterns of the early and late ERP components that are observed in these experiments, we discovered that: auditory processing of pitch variations in tones, as a bottom-up effect, elicited greater right hemisphere activation; in contrast, linguistic processing of lexical tones, as a top-down effect, elicited greater left hemisphere activation. We also found that both types of processing co-occurred in both the early (around 200 ms) and late (around 300–500 ms) time windows, which supported a parallel model of lexical tone perception. Unlike the previous view that language processing is special and performed by dedicated neural circuitry, our study have elucidated that language processing can be decomposed into general cognitive functions (e.g., sensory and memory) and share neural resources with these functions.

Keywords: lexical tone, ERP, lateralization, serial model, parallel model

INTRODUCTION

Perception in general comprises two types of processing, bottom-up (or data-based) processing and top-down (or knowledge-based) processing, which are based, respectively, on incoming data and prior knowledge (Goldstein, 2009). For speech perception, the TRACE model (McClelland and Elman, 1986) claims that both types of processing are necessary, and the auditory sentence processing model (Friederici, 2002) proposes that the cognitive processes involved in speech perception proceed in a series of steps. Following these models, bottom-up processing such as acoustic processing of incoming signal and generalization of speech features happens first, whereas top-down processing such as recognition based on knowledge of phonemes, semantics, or syntax takes effect at a later stage of perception. These models, as well as other theories or models of speech perception (e.g., Liberman and Mattingly, 1985; Fowler, 1986; Stevens, 2002; Diehl et al., 2004; Hickok and Poeppel, 2007), are based primarily on evidence from non-tonal languages. Two issues remain to be explored: (a) whether the processing of tonal languages, which take up about 60–70% of world languages (Yip, 2002), follows the same cognitive processes; and (b) what is the role of lexical tone perception in a general model of speech perception. In addition, considering that the lexical tone attached to a syllable is carried mainly by the vowel nucleus of the syllable, the temporal dimension of cognitive processes underlying lexical tone perception is of special interest.

In this paper, we discussed the temporal relationship between bottom-up and top-down processing in lexical tone perception, with the purpose of not only examining the underlying mechanisms of lexical tone perception but also shedding valuable light on the general models of speech perception concerning tonal languages. Lexical tone is a primary use of pitch variations to distinguish lexical meanings (Wang, 1967). Noting that pitch perception belongs to the general auditory perception that is also shared by other animals (Hulse et al., 1984; Izumi, 2001; Yin et al., 2010) and word semantics are acquired primarily through language learning, lexical tone perception also entails both bottom-up and top-down processing. In our study, we defined *bottom-up processing* as auditory processing and feature extraction of incoming acoustic signals, which referred specifically to pitch contour perception. By contrast, we defined *top-down processing* as recognition and comprehension of incoming signals according to language knowledge, which referred specifically to influence of language experience on recognizing and comprehending a certain syllable in a tonal language. The recognition of a pitch contour as a certain tonal category was also ascribed to top-down processing.

Ample of available studies on lexical tone perception focused on the lateralization patterns of lexical tone processing (e.g., Van Lancker and Fromkin, 1973; Baudoin-Chial, 1986; Hsieh et al., 2001; Wang et al., 2001; Gandour et al., 2002, 2004; Tervaniemia and Hugdahl, 2003; Luo et al., 2006; Zatorre and Gandour, 2008; Li et al., 2010; Krishnan et al., 2011; Jia et al., 2013), and reported

mixed results even under the same experimental paradigms. For example, by employing the dichotic listening (DL) paradigm and materials from Mandarin, Baudoin-Chial (1986) reported no hemisphere advantages of lexical tone perception, but Wang et al. (2001) found a left hemisphere advantage. Using fMRI, Gandour and colleagues compared lexical tone processing with intonation or vowel processing. The study of lexical tone and intonation (Gandour et al., 2003b) revealed a left hemisphere advantage in frontal lobe, whereas the study of lexical tone and segments (Li et al., 2010) discovered a right hemisphere advantage in fronto-parietal area for the perception of tones. A right lateralization of lexical tone perception was also reported in an ERP (Luo et al., 2006) and a DL experiment (Jia et al., 2013).

The inconsistent laterality effects could be due to different experimental conditions in these studies. For example, in the DL experiments reporting a left hemisphere advantage (Van Lancker and Fromkin, 1973; Wang et al., 2001), tonal language speakers participated into more difficult tasks than non-tonal language speakers, and the heavier load of these tasks (e.g., hearing trials at a faster pace) might enhance the left hemisphere advantage in tonal language speakers. By contrast, there were no hemisphere advantages in the study that had no task differences between tonal and non-tonal language speakers (Baudoin-Chial, 1986). In addition, in the DL tasks that involved meaningless syllables and hums, which could direct participants' attention toward pitch contours only, a right hemisphere advantage was shown (Jia et al., 2013). More importantly, whether language-related tasks are involved is the primary noticeable difference between studies showing an explicit right hemisphere advantage (e.g., Luo et al., 2006) and those reporting a left hemisphere advantage of lexical tone perception (Van Lancker and Fromkin, 1973; Hsieh et al., 2001; Wang et al., 2001; Gandour et al., 2002, 2004). For example, Luo et al. (2006) conducted a passive listening task in which participants were engaged in a silent movie, whereas the other studies carried out explicit language tasks such as lexical tone identification. Accordingly, the right lateralization reported in Luo et al. (2006)'s study could be attributed to the pure bottom-up effect without top-down influence, whereas the other studies did not address the underlying mechanisms of lexical tone perception. This could lead to the inconsistent results between these studies. These mixed results also reflect a multifaceted perspective on lexical tone processing and hemispheric lateralization. As stated in Zatorre and Gandour (2008)'s review, in tonal processing, "it appears that a more complete account will emerge from consideration of general sensory-motor and cognitive processes in addition to those associated with linguistic knowledge."

To our knowledge, among the available studies, there was only one work (Luo et al., 2006) that discussed these two types of processing in lexical tone perception and a few that examined the cognitive processes involved for lexical tone perception (Ye and Connie, 1999; Schirmer et al., 2005; Liu et al., 2006; Tsang et al., 2010). In Luo et al. (2006)'s study, a serial model of lexical tone processing was proposed, which suggested that bottom-up processing (i.e., pitch perception) took effect in an early time window around 200 ms and top-down processing (i.e., semantic comprehension) happened in a late time window around 300–500 ms. The first half of this model was based on their experimental results that phonemes with slow- (lexical tone) and fast-changing (stop-consonant) acoustic properties inducted, respectively, right and left lateralization patterns of the MMN (Mismatch Negativity) component. The second half was proposed to address the confliction between their results and the previous literature that showed a general left hemisphere advantage of lexical tone perception. They proposed that during the late stage a left lateralization should be shown in the semantics-associated late ERP component, N400 (Kutas and Hillyard, 1980).

This serial model associated the right hemisphere advantage with bottom-up processing of lexical tones, and the left hemisphere advantage with top-down processing. In terms of lexical tone perception, there exists ample evidence in support of such association between the two types of processing and the two types of hemisphere advantage. For example, in studies of language experience and prosody, Gandour et al. (2004) dissociated linguistic processing in the left hemisphere and acoustic processing in the right hemisphere, by locating a left lateralization in certain brain regions in tonal language speakers and a right lateralization in non-tonal language speakers during speech prosody processing. Pitch processing has a right hemisphere advantage, as shown in behavior experiments such as DL (Sidtis, 1981), PET studies (Zatorre and Belin, 2001), and later fMRI studies (Boemio et al., 2005; Jamison et al., 2006); for review, see Zatorre et al. (2002). By contrast, compared to non-tonal language speakers, tonal language speakers have greater left hemisphere activities during lexical tone perception (Gandour et al., 1998, 2004; Hsieh et al., 2001; Wang et al., 2004), and multiple brain regions in the left hemisphere were believed to be the primary source of N400 (Lau et al., 2008). Noting these, we also adopted the lateralization pattern in our study to investigate top-down and bottom-up processing of lexical tones.

In addition, in Luo et al. (2006)'s study, there was insufficient direct evidence to manifest the top-down effect at the late stage of processing, because this study only explored acoustic factor without involving explicit language-related tasks or any linguistic factor. Therefore, it is hard to comprehensively evaluate Luo et al. (2006)'s serial model. Considering these, in order to make sure that language knowledge (top-down) would take effect, we adopted a number of explicit language-related tasks, including DL, lexical decision with phonological priming, and semantic violation. Meanwhile, we manipulated both the acoustic (requiring bottom-up processing) and semantic (requiring top-down processing) factors in the experimental design and analyzed the ERP components at both the early (around 200 ms) and late (around 300–500 ms) processing stages to explore the temporal relationship of the two types of processing during lexical tone perception.

Our experimental results showed that both bottom-up (acoustic) processing and top-down (semantic) processing exist in both the processing state around 200 ms and that around 300–500 ms, which inspired a parallel model of top-down and bottom-up processing in lexical tone perception. In the rest of the paper, we described the two ERP components traced in our experiments of Mandarin lexical tone perception (section ERP Components Reflecting Bottom-up and Top-down Processing), reported these experiments and their findings (section ERP

Experiments of Lexical Tone Perception), discussed the lateralization patterns of the ERP components shown in these experiments and the derived parallel model of lexical tone perception (section General Discussions), connected language processing with general cognitive functions (section Language Processing and General Cognitive Functions), and finally, concluded the paper (section Conclusion).

ERP COMPONENTS REFLECTING BOTTOM-UP AND TOP-DOWN PROCESSING

We examine two ERP components in our experiments, namely auditory P2 and auditory N400, which occur, respectively, in the early and late time windows after stimulus onset.

Auditory P2 is the second positive going ERP component. It usually has a central topographic distribution, and peaks in the early time window around 200 ms (Luck, 2005). The lateralization of P2 is subject to both acoustic properties and tasks (e.g., categorizing emotional words, Schapkin et al., 2000). The corresponding MEG component is P2m or M200. Previous research reported a general left lateralization of P2m in doing language-related tasks (e.g., perceiving consonants and vowels, Liebenthal et al., 2010), but acoustic properties of incoming signals also affect the lateralization of P2m (e.g., the voice onset time of consonants, Ackermann et al., 1999).

As a negative going potential, the auditory N400 appears in the late time window (around 250–550 ms) when the target sound stimulus is incongruent with the context (Kutas and Federmeier, 2011). The semantic violation paradigm can elicit N400 (Kutas and Hillyard, 1980). The phonological priming experiment can also elicit N400, when comparing the control condition with the priming condition (Praamstra and Stegeman, 1993; Dumay et al., 2001). The auditory N400 usually has a more frontal topological distribution than the visual N400 (Holcomb and Anderson, 1993; Kutas and Federmeier, 2011). In young population, the auditory N400 tends to have a frontal distribution (Curran et al., 1993; Tachibana et al., 2002). The source of N400 is believed to lie in the frontal and temporal brain areas (Maess et al., 2006; Lau et al., 2008), starting from 250 ms in the posterior half of the left superior temporal gyrus, migrating forward and ventrally to the left temporal lobe by 365 ms, and then moving to the right anterior temporal lobe and both frontal lobes after 370 ms (Kutas and Federmeier, 2011).

ERP EXPERIMENTS OF LEXICAL TONE PERCEPTION

We designed three ERP experiments to explore the temporal relation of bottom-up and top-down processing in lexical tone perception. These experiments recruited Mandarin participants and traced the above two ERP components in three tasks, respectively, at the syllable, word, and sentence levels, which cover aspects of acoustics and phonetics, phonology, and semantics processing. According to the serial models (e.g., Friederici, 2002), these types of processing could be reflected by different ERP components shown at the early and late stages. However, a parallel model would predict a co-existence of these types of processing at both the early and late stages of lexical tone perception.

These experiments were designed primarily for the following two reasons. First, we were interested in clarifying whether top-down effects could happen at the early stage of a "lower-level" processing. To this purpose, we designed Experiment 1 using the DL task. Apart from the bottom-up effect on phoneme identification, we introduced a semantics factor to see whether a top-down effect inducted by this factor could exist in the early stage of perception and whether such effect could be reflected by the early ERP components (e.g., P2).

Second, we were interested in identifying bottom-up effects at the late stage of a "high-level" processing. To this purpose, we designed Experiment 2 using an auditory lexical decision task, which entailed a top-down semantic processing and a bottom-up processing induced by various types of phonological primes. We also designed Experiment 3 using a semantic violation task. In this task, semantic integration could be reflected by the late ERP component (e.g., N400). Meanwhile, phonemes bearing different acoustic properties could also induce the bottom-up acoustic processing at this stage.

Experiment 1 involved a DL task, which is a widely-adopted paradigm in behavioral and ERP studies examining the lateralization in the auditory modality. For example, Eichele et al. (2005) adopted a DL task using stop consonants as stimuli, and discovered that the latency of the ERP waveforms in the left hemisphere were shorter than those in the right hemisphere, thus reflecting a quicker response of the left hemisphere in perceiving stop consonants. Wioland et al. (1999) explored pitch perception in a DL task, and found that the ERP waveforms had higher amplitudes when the tone change happened in the left ear than in the right ear, thus indicating that the right hemisphere had prevalence in pitch discrimination. In our experiment, we adopted the DL paradigm to explore tone lateralization, and used the amplitude of auditory P2 as a temporal indicator, rather than ear advantages as in previous contradictory behavioral responses (Van Lancker and Fromkin, 1973; Baudoin-Chial, 1986), to reflect hemispheric specialization. We compared the lateralization patterns under tones and stop consonants in both words and non-words. In terms of acoustic properties, the stop consonants have fast-changing properties, whereas the lexical tones in Mandarin have slow-changing properties.

We expected an increase in the activity of the hemisphere for a certain processing, when there was a heavier load of information in the corresponding hemisphere. For example, in dichotic trials containing two different lexical tones, there would be a relatively greater right hemisphere advantage (equivalent to a less left hemisphere advantage) than dichotic trials containing two different stop-consonants but the same lexical tones. Similarly, dichotic trials containing words should generate a greater left hemisphere activity than dichotic trials containing non-words. In line with previous literature (Wioland et al., 1999; Luo et al., 2006), we examined the ERP waveforms in the C3 and C4 electrode groups.

Experiment 2 involved a lexical decision task with phonological priming. Priming refers to the phenomenon of acceleration in response after repetition. An early study of child language acquisition (Bonte and Blomert, 2004) adopted such a task. It used Dutch words and non-words as testing materials, and discovered different N400 reduction patterns in different language groups. Our experiment adopted a similar design, but used consonants and tones, as well as Chinese words and non-words as testing

materials. In Experiment 2, consonant or tone primes appeared before target words, and we examined the auditory P2 and auditory N400 under the tone or consonant priming paradigm.

Other than the enhancement effect of DL as in Experiment 1, we expected that there would be a reduction of ERP components (smaller amplitude) due to the priming effect, and that semantic violation would induce a reduction of ERP amplitudes (as shown by the smaller amplitude in the positive component and greater amplitude in the negative component). These reductions could be greater in the corresponding hemisphere related to a certain processing. For example, the reduction caused by lexical tone priming should have a greater right hemisphere advantage (equivalent to a less left hemisphere advantage) compared to that of consonants, whereas the reduction caused by non-words should be greater in the left hemisphere compared to that of words. Considering the topographic distributions of auditory P2 and N400 (Curran et al., 1993; Tachibana et al., 2002; Luck, 2005) as well as the auditory brain regions involved in the tone priming tasks (Wong et al., 2008), In Experiment 2, we examined the ERP waveforms in the posterior (P3, P4) and frontal (F3, F4) electrode groups.

Experiment 3 involved a semantic violation task in sentences. N400 has been one of the most widely-explored ERP components in such studies, and we adopted the semantic violation paradigm to explore whether acoustic property affected the lateralization of auditory N400 occurring in the late time window during sentence comprehension, which was a high-level linguistic task. The violation was induced by changing either the stop consonant or the tone of the target syllable in a sentence. We expected a greater right lateralization of the N400 induced by lexical tone violation compared to consonant violation. Here we set the central (C3, C4) electrode groups as the regions of interest.

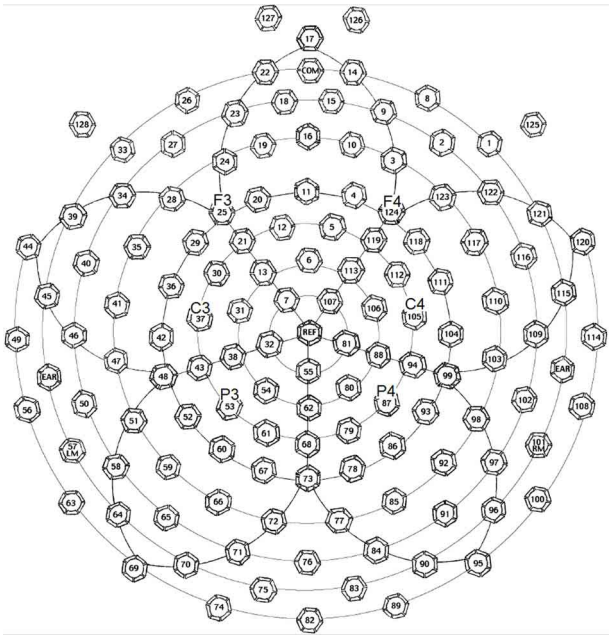


FIGURE 1 | Electrode positions of the EGI 128-channel Geodesic Sensor Net, in which the key electrodes (F3, F4, C3, C4, P3, and P4) for forming electrode groups to trace interested ERP waveforms and components are marked. This figure is available at: ftp.egi.com/pub/documentation/placards/gsn200_128_map.pdf.

PARTICIPANTS AND SETTINGS

All these experiments were approved by the College Research Ethics Committee (CREC) of Hong Kong. Thirty-two university students (16 females, 16 males) volunteered for these experiments (age range: 19–29, mean = 27, *SD* = 4.2). In Experiment 1, data from all participants were analyzed. In Experiment 2, the data of one participant were excluded due to excessive eye movements, thus leaving 31 participants (age range: 19–29, mean = 25, *SD* = 2.3). In Experiment 3, the data of three participants were excluded, thus leaving 29 participants (age range: 19–29, mean = 26, *SD* = 2.8).

All these participants were native Mandarin speakers with no musical training. They had normal hearing (below 25 dBHL) in both ears and less than 10 dBHL differences at 125, 250, 500, 750, and 1000 Hz between the two ears, according to the PTA (pure tone analysis) test. They were all right-handed according to the Edinburgh handedness test (Oldfield, 1971), and reported no history of head damage or mental illness. They signed informed consent forms before each of these experiments, and got compensation at a rate of 50 HKD per hour after completing these experiments.

These experiments were conducted on three separate days. They were conducted in a dimly lit, quiet room. During experiment, participants were seated comfortably in front of a computer monitor, and the sound stimuli were presented via ER-3A air-conducting insert earphones, which diminished the environmental noise by 20–30 dB. Sound pressure level, measured by a sound level meter, was set to 75 dBSPL during experiment. The sound materials in these experiments were recorded from a female, native Mandarin speaker. The recording was conducted in a sound-proof booth using Shure SM10A microphone and Sony PCM-2700A audio recorder. The adjustments on recorded sound materials were implemented by the PSOLA (pitch-synchronous overlap add) algorithm in Praat (Boersma and Weenink, 2013), the experimental procedures were implemented using E-Prime (Psychology Software Tools, Pittsburgh, PA), and the statistical analyses were conducted using the SPSS software (version 18.0, SPSS Inc. Chicago, IL).

EEG DATA RECORDING AND ERP PROCESSING

The EEG (electroencephalography) data were collected by a 128-channel EEG system with Geodesic Sensor Net (EGI Inc., Eugene, OR, USA) (see **Figure 1**). The impedances of all electrodes were kept below 50 kΩ at the beginning of the recording. In all the three experiments, participants were encouraged to avoid blinking or moving their body parts at certain points. Eye blinks and movements were monitored through electrodes located above and below each eye and outside of the outer canthi. The original reference point was the vertex. The ERPs were re-referenced to the averages of all 129 scalp channels in data processing (average reference). During recording, signals were sampled at 250 Hz with a 0.01–100 Hz band-pass filter.

During offline ERP processing, the recorded continuous data were filtered by a 40 Hz low-pass filter and segmented from −100

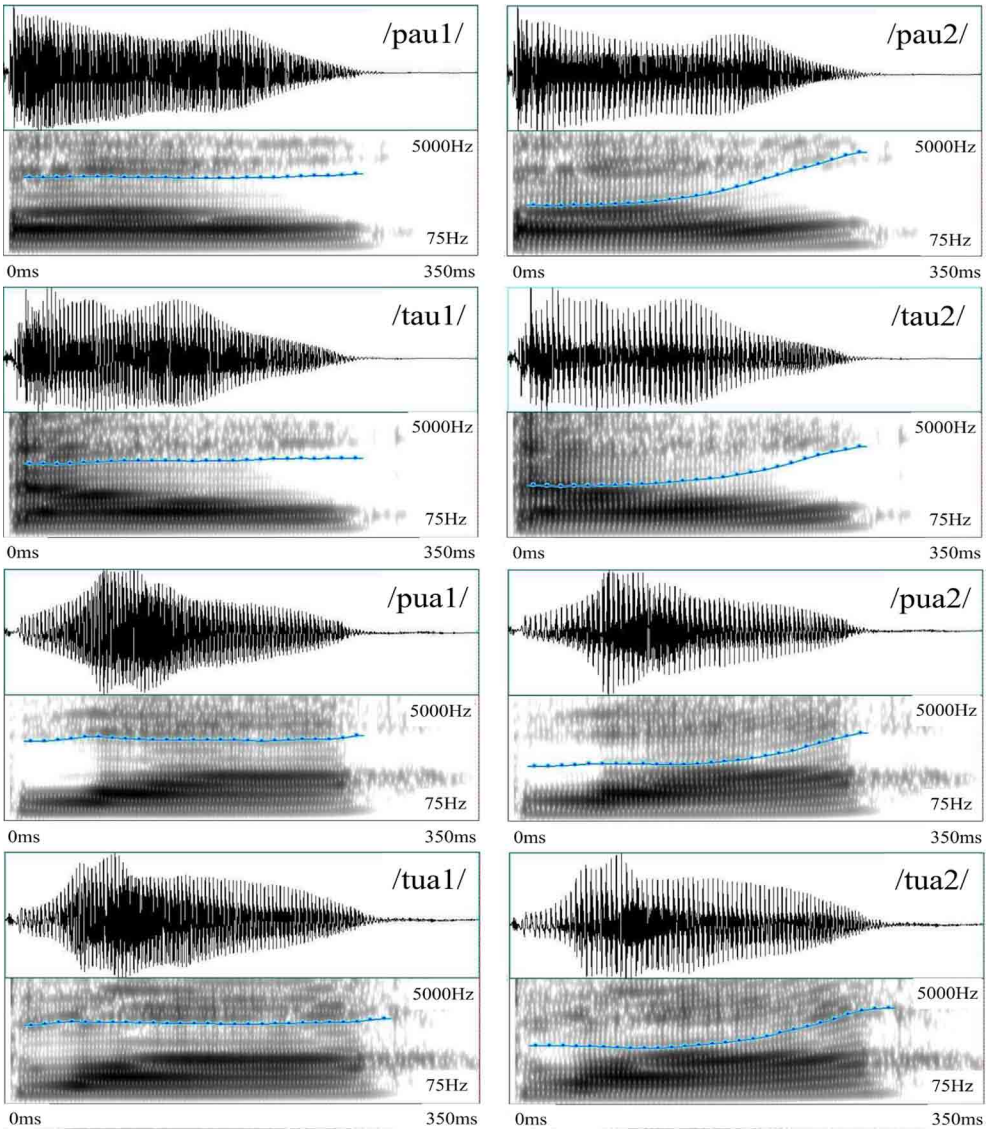


FIGURE 2 | Sound waveforms and spectrograms of the eight Mandarin syllables in Experiment 1. The pitch contours of the syllables having level tones have stable level portions throughout the duration, whereas those of the syllables having rising tones start with a level portion (about half of the duration), followed by a rising portion (about half of the duration). The x-axis represents time (0–350 ms) and the y-axis represents frequency (75–5000 Hz) in spectrograms. The blue curve represents pitch contour at a different scale (50–500 Hz) superimposed on the spectrograms.

to 900 ms by referring to the stimulus onset. The segments having either an amplitude change exceeding 100 μV in the vertical eye channels and all electrodes, or a voltage fluctuation exceeding 50 μV in the horizontal eye channels were excluded from analyses. In each experiment, at least a half number of total trials were preserved for analysis in each condition and for every participant. The baseline correction was conducted from −100 to 0 ms.

In the following sections, we reported the materials, procedures, and results of these three experiments.

EXPERIMENT 1: MANDARIN TONE DICHOTIC LISTENING TASK

Materials

The recorded stimuli included Mandarin real- and pseudo-syllables, which are formed by two stop consonants (/p/ and /t/ in the IPA notation), two diphthongs (/au/ and /ua/ in the IPA notations), and two Mandarin tones (tone 1, the high level tone; and tone 2, the high rising tone). Eight syllables were constructed using these phonemes and tonemes, among which four were real-syllables, having corresponding Chinese characters, whereas the other four were pseudo-syllables, made of valid consonants and diphthongs but having no corresponding Chinese characters. All these stimuli were cut to 350 ms based on intensity profile. **Figure 2** shows their waveforms and spectrograms, among which the pitch contours are also marked.

Procedure

In the DL task, participants simultaneously heard two distinct syllables, respectively, in their left and right ears, and were asked to

Table 1 | Experimental conditions in Experiment 1, each containing two syllables and four DL trials.

Conditions	Word	Non-word
Consonant	/pau1/ (包,"bag"), /tau1/(刀,"knife")	/tua2/,/pua2/
Tone	/pau2/ (雹,"hail"),/pau1/(包 ,"bag")	/pua1/,/pua2/

Words have corresponding Chinese characters (shown in brackets, together with their meanings). The pronunciations are annotated with the IPA characters, the numbers in which denote Mandarin tones.

respond, according to the Chinese character "左/右" (left/right) shown on the screen, both the consonant and the tone of the corresponding side of the auditory input, by pressing the corresponding keys on the respond pad.

We adopted a two-by-two design, with word and non-word as two levels of the lexicality factor, and stop consonant and tone as two levels of the acoustic contrast factor. The eight syllables formed four experimental conditions, including the word, consonant condition; word, tone condition; non-word, consonant condition; and non-word, tone condition, each containing two syllables (see examples in **Table 1**). In the two word conditions, the words were formed by meaningful real-syllables in Chinese; in the two non-word conditions, the non-words were formed by pseudo-syllables having no meanings in Chinese. In the two consonant conditions, the two syllables had the same diphthong and tone, but different initial consonants; in the two tone conditions, however, the two syllables had the same initial consonant and diphthong, but different tones. To balance the two syllables, respectively, played to the left and right ears of participants and the two directions in participants' responses (left or right), each of these four conditions corresponded to four DL trials. In total, there were 16 DL trials.

In each trial, a fixation first appeared on the center of the screen and remained there. After 400 ms, the two syllables in a DL trial were simultaneously played to the left and the right ears of participants, respectively. Participants were encouraged not to blink or move their body parts during the appearance of the fixation. After 1000 ms, the fixation on the screen was replaced by the Chinese character that indicated left or right, and accordingly, participants reported the consonant and the tone of the syllable heard by their corresponding ears. The purpose of letting participants hear the stimuli before seeing the indication (left or right) was to avoid inducing prior bias in their attention. The indication stayed on the screen for 2000 ms, during which participants gave their responses. The presentation sequence of the stimuli was randomized, and the order of choices between the two consonants and between the two tones on the response box was counter-balanced across participants.

Participants first went through a practice session (16 trials) to familiarize the experimental paradigm. In the experimental session, a total of 256 trials were presented to participants, each lasting around 5 s. The experiment consisted of four blocks, each having 64 trials that lasted about 5 min. In each block, the 16 DL trials randomly repeat four times. Participants could take a 2-min break after each block, and the whole experiment lasted approximately 30 min.

Data analysis and results

As for the behavioral data, the overall rate of response was 95.7%. A Three-Way repeated-measures ANOVA of rates of correct response, with lexicality (word vs. non-word), acoustic contrast (consonant vs. tone), and hemisphere (left vs. right) as three factors, revealed a significant three-way interaction [$F_{(1,\ 31)} = 5.981$, $p < 0.024$, $\eta_p^2 = 0.239$]. In addition, the *post-hoc* analysis revealed a significant left hemisphere (right ear) advantage in the non-word, consonant condition [$t_{(19)} = -2.280$, $p < 0.034$]. Since participants needed to respond to both the consonant and the tone of the syllable in one ear, we did not analyze the reaction time.

As for the ERP data, considering the central distribution of auditory P2 (Luck, 2005) and previous literature (Luo et al., 2006), we averaged the data recorded by the four homolog pairs of adjacent central electrodes including C3 and C4 [electrodes 37 (C3), 38, 42, 43 in the left hemisphere, and 105 (C4), 88, 104, 94 in the right hemisphere, according to the EGI system] for analysis. Since the P2 peak appeared between 180 and 200 ms, we averaged the amplitude of P2 within this time range. A Three-Way repeated-measures ANOVA of P2 amplitudes, with lexicality, acoustic contrast, and hemisphere as three factors, revealed two significant interactions, one between acoustic contrast and hemisphere [$F_{(1,\ 31)} = 7.744$, $p < 0.0091$, $\eta_p^2 = 0.200$] and the other between lexicality and hemisphere [$F_{(1,\ 31)} = 12.687$, $p < 0.0012$, $\eta_p^2 = 0.290$], and two main effects, hemisphere [$F_{(1,\ 31)} = 14.393$, $p < 0.0006$, $\eta_p^2 = 0.317$] and acoustic contrast [$F_{(1,\ 31)} = 22.024$, < 0.0001, $\eta_p^2 = 0.415$].

Figure 3 shows the average ERP waveforms of the C3 and C4 electrode groups, **Figure 4** shows the topographies of the ERP component contrasts, and **Figure 5** shows the average P2 amplitudes between 180 and 200 ms in different conditions.

The greater left lateralization of P2 shown by comparing the word conditions with the non-word conditions reflected top-down processing in the early time window around 200 ms. This indicates the involvement of language experience in the process. In order to differentiate words from non-words, participants needed prior language knowledge, which casted as a top-down effect, regardless of whether this effect belonged to word form recognition (Friederici, 2002) or semantic processing.

The greater left lateralization of P2 shown by comparing the consonant conditions with the tone conditions also reflected bottom-up processing in the same time window. The difference between these conditions was the speed of changes in acoustic cues. In line with previous results (Jamison et al., 2006), we found a greater left lateralization in perceiving fast changing acoustic cues (formant transition in stop consonants) and a less left lateralization in perceiving relatively-slow changing acoustic cues (pitch changes in tones). We expected that the less left lateralization of tone processing compared to consonant processing was due to the greater right lateralization of tone processing compared to consonant and rhyme in certain brain regions in Li et al.'s work 2010.

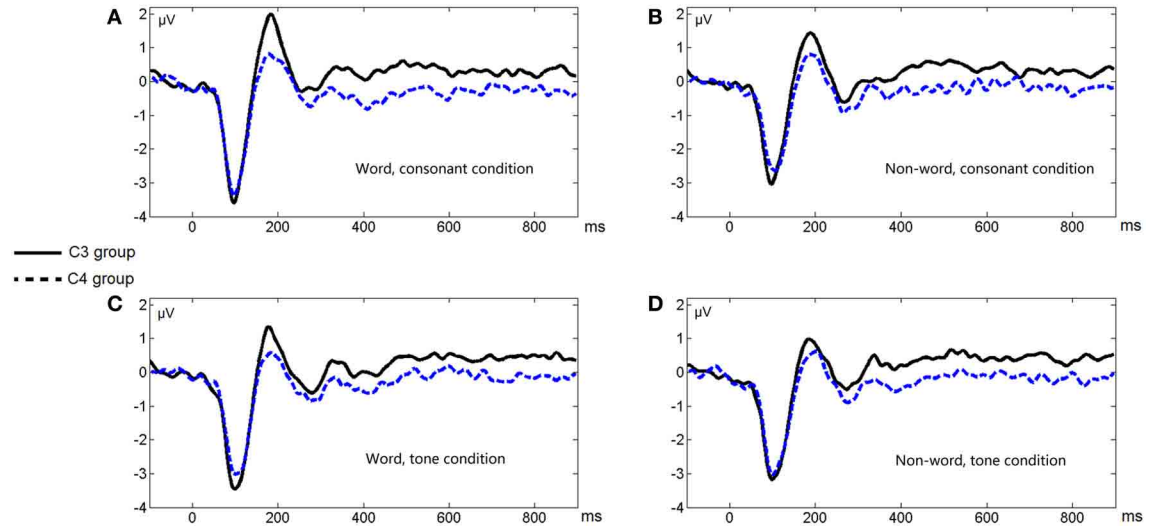


FIGURE 3 | Average ERP waveforms of the C3 and C4 electrode groups under the four conditions of Experiment 1. (A) Word, consonant condition; **(B)** Non-word, consonant condition; **(C)** Word, tone condition; **(D)** Non-word, tone condition.

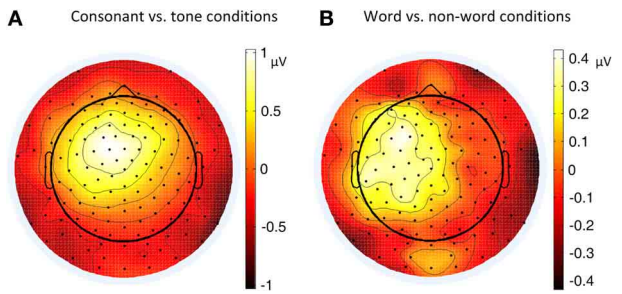


FIGURE 4 | Topographies of P2 (180–200 ms) contrasts in different conditions of Experiment 1. (A) Consonant vs. tone condition; **(B)** Word vs. non-word condition.

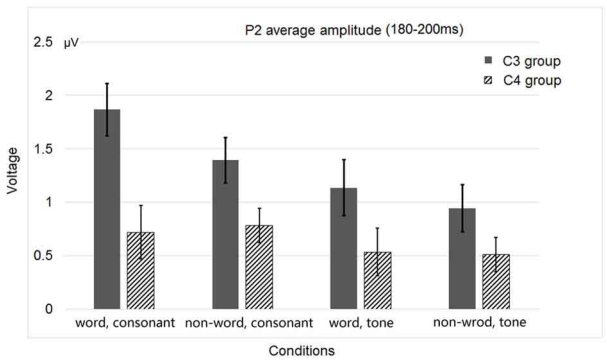


FIGURE 5 | Average P2 amplitudes of the C3 and C4 electrode groups in Experiment 1.

EXPERIMENT 2: LEXICAL DECISION TASK WITH PHONOLOGICAL PRIMING

Materials

The recorded stimuli included monosyllabic words as primes and disyllabic words as targets. These words could be real- or non-words in Chinese. The mean duration of the primes was 383.06 ms (range: 251–591 ms, $SD = 49.17$), and that of the targets 619.58 ms (range: 510–751 ms, $SD = 52.05$). There was no significant differences of either the prime duration [$F_{(5,\ 354)} = 1.098$, $p = 0.3609$, $\eta_p^2 = 0.015$] or the target duration [$F_{(5,\ 354)} = 1.244$, $p = 0.2878$, $\eta_p^2 = 0.017$] between conditions. The onset asynchrony between the primes and the targets was fixed at 1000 ms. The sound intensity of the primes was set to 55 dB, and that of the targets 75 dB. The purpose of presenting primes at a lower intensity was to maximize the priming effect (Lau and Passingham, 2007).

Procedure

In the lexical decision task, participants were asked to judge whether the heard disyllabic words (targets) were words or non-words. They were instructed to ignore the monosyllabic words (primes) played before the disyllabic words and focus on the latter.

Similar to Experiment 1, we adopted a two-by-two design, with word and non-word as the two levels of the lexicality factor, and stop consonant and tone as the two levels of the priming condition factor. The materials in **Table 2** formed six experimental conditions. In the two consonant conditions, the syllable in the prime shared the initial consonant with the first syllable of the target; in the two tone conditions, the syllable in the prime shared the tone with the first syllable of the target; and in the two control conditions, the syllable in the prime and the first syllable of the target shared no phonemes or tonemes.

In each trial, participants first heard the prime. After 600 ms, a fixation appeared on the center of the screen and remained there. After another 400 ms, participants heard the target. After another 1000 ms, the fixation disappeared, and participants had 3000 ms to give their response, by pressing one of the two keys marked by "yes" and "no" in the keyboard. Participants were encouraged not

Table 2 | Example materials and experimental conditions in Experiment 2.

Conditions	Prime	Target
Word, consonant priming	/tɕʰyɛ2/ (瘸)	/tɕʰi4//tʂʰɤ1/ (汽车, "vehicle")
Non-word, consonant priming	/tɕiəŋ4/ (静)	/tɕiɑʊ1//pən3/ (*交本)
Word, tone priming	/xʊŋ1/ (烘)	/ɕiɑŋ1//kɑŋ3/ (香港, "Hong Kong")
Non-word,tone priming	/tɕʰyn2/ (群)	/kuɔ2// mən4/ (*国闷)
Word, control	/fa1/ (发)	/ɕyɛ2//tʂɤ3/ (学者, "scholar")
Non-word,control	/lun2/ (轮)	/jiəŋ3/fəŋ4/ (*影缝)

*The pronunciations are annotated with the IPA characters, the numbers in which denote Mandarin tones. As for the primes, Chinese characters are shown in brackets. As for the targets, the Chinese words are shown in brackets, together with their meanings. Each non-word includes two real-syllables (shown in brackets), but their combination does not form a meaningful disyllabic word in Chinese (marked by *).*

to blink or move their body parts during the appearance of the fixation, and not to respond until the fixation disappeared. One half of the participants responded to the "yes" key with their left index finger and the "no" key with their right index finger. The other half did the reverse. The left or right response order was randomly assigned to participants.

There were in total 360 trials, with 60 trials in each of the six conditions. Each trial lasted around 5 s. Participants first went through a practice session (36 trials) to familiarize the experimental paradigm. The experiment consisted of six blocks, each having 60 trials and lasting about 5 min. Trials were arranged in a random order. Participants could take a 2-min break after each block, and the whole experiment lasted approximately 40 min.

Data analysis and results

As for the behavioral data, the response correctness was 96.0%. The average reaction time was 962.72 ms ($SD = 159.99$). Outliers greater than three times of standard deviation from mean were replaced with mean value in each participant. A marginal significant priming effect in the word, consonant priming condition was observed [$t_{(30)} = -1.993$, $p < 0.055$], while the tone priming conditions showed interference effects [as for the word, tone priming condition, $t_{(30)} = 1.822$, $p = 0.078$; as for the non-word, tone priming condition, $t_{(30)} = 2.524$, $p < 0.017$].

As for the ERP data, considering both of the P2 topography (Luck, 2005) and the brain regions for tone priming (Wong et al., 2008), we averaged four homolog pairs of adjacent posterior electrodes including P3 and P4 [electrodes 53 (P3), 61, 54, 38 in the left hemisphere, and 87 (P4), 79, 80, 88 in the right hemisphere, according to the EGI system] for analysis. We calculated the priming effects of tones or consonants by subtracting the ERP waveforms in the word or non-word control conditions from those in the word or non-word experimental conditions. Similar to Experiment 1, based on a Three-Way repeated-measures ANOVA, with lexicality (word vs. non-word), priming condition (consonant vs. tone), and hemisphere (left vs. right) as three factors, we found that the average P2 amplitude in the early time window (200–220 ms, P2 peak values were within this time range) showed two significant interactions, one between lexicality and hemisphere [$F_{(1, 30)} = 5.618$, $p < 0.0244$, $\eta_p^2 = 0.158$], and the other between priming condition and hemisphere [$F_{(1, 30)} = 8.515$, $p < 0.0066$, $\eta_p^2 = 0.221$], and a main effect of lexicality [$F_{(1, 30)} = 8.242$, $p < 0.0074$, $\eta_p^2 = 0.216$]. Similar to Experiment 1, these results indicated that both semantics and acoustic properties affected lateralization.

Apart from P2, we conducted another analysis of the ERP waveform in the late time window (500–550 ms) based on four homolog pairs of adjacent frontal electrodes including F3 and F4 [electrodes 25 (F3), 28, 29, 35 in the left hemisphere, and 124 (F4), 123, 118, 117 in the right hemisphere, according to the EGI system]. Rather than deriving auditory N400 by contrasting non-word and word conditions, we analyzed these conditions separately in order to preserve the lexicality factor and make factors in statistic analysis consistent with the previous one based on P2, though the interested time windows in these two analyses were different. The data for statistical analysis here were all from priming conditions without subtracting control conditions. By examining the same three factors as in the previous analysis, this analysis showed three main effects, priming condition [$F_{(1, 30)} = 6.564$, $p < 0.0157$, $\eta_p^2 = 0.180$], lexicality [$F_{(1, 30)} = 7.892$, $p < 0.0087$, $\eta_p^2 = 0.208$], and hemisphere [$F_{(1, 30)} = 9.193$, $p < 0.0050$, $\eta_p^2 = 0.235$]. Priming condition interact significantly with lexicality [$F_{(1, 30)} = 5.636$, $p < 0.0242$, $\eta_p^2 = 0.158$]. More importantly, there was a significant interactions between lexicality and hemisphere [$F_{(1, 30)} = 10.729$, $p < 0.0027$, $\eta_p^2 = 0.263$].

Figure 6 shows the average ERP waveforms of the P3 and P4 electrode groups, and **Figure 7** shows those of the F3 and F4 electrode groups. **Figure 8** shows the topographies of the contrasts of the P2 component around 200 ms (200–220 ms), and **Figure 9** shows those of the late ERP component around 500 ms (500–550 ms). **Figure 10** shows the average amplitudes of P2, and **Figure 11** shows those of the late component in different conditions.

The lateralization pattern of P2 could be interpreted as follows. The greater the priming effect, the lower the amplitude of P2, due to the repetition effect. Since there was no main effect of priming condition, the priming effects of consonants and tones were not much different from each other. However, the left and right hemispheres showed different trends of these priming effects, due to the significant interaction between hemisphere and priming condition. By examining the amplitude difference between the priming effects of consonants and tones (see **Figure 8A**) and those of words and non-words (see **Figure 8B**), we found a stronger priming effect of consonants than tones was shown in the left hemisphere compared to the right hemisphere around centro-parietal region, and a greater left hemisphere advantage in processing words compared to non-words. These showed that the left hemisphere responded significantly differently in the consonant priming and tone priming conditions, as well as word and non-word conditions. In line with Experiment 1, these results illustrated that both bottom-up and top-down processing took place around 200 ms. The left hemisphere responded to the fast changing acoustic cues greater than the slow changing acoustic cues, and it also responded to word semantics greater than non-words that had no meanings.

By comparing the word and non-word conditions, we found that the amplitudes of frontal region at the late component (around 500 ms) were lower in the non-word conditions compared to the word conditions (consistent with the main effect of lexicality), which was right lateralized (consistent with the

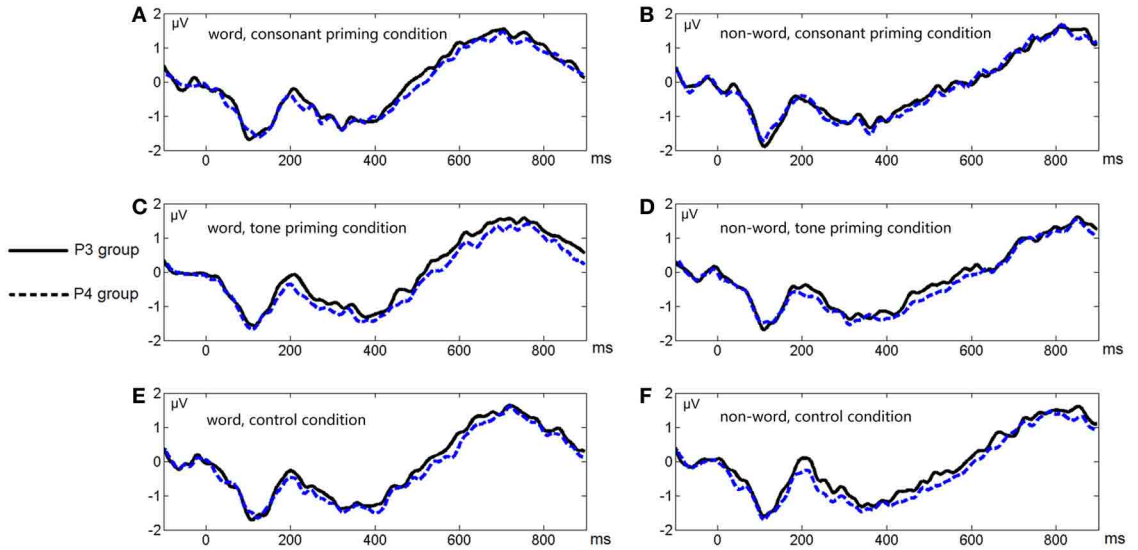


FIGURE 6 | Average ERP waveforms of the P3 and P4 electrode groups under the six conditions of Experiment 2. (A) Word, consonant priming condition; **(B)** Non-word, consonant priming condition; **(C)** Word, tone priming condition; **(D)** Non-word, tone priming condition; **(E)** Word, control condition; **(F)** Non-word, control condition.

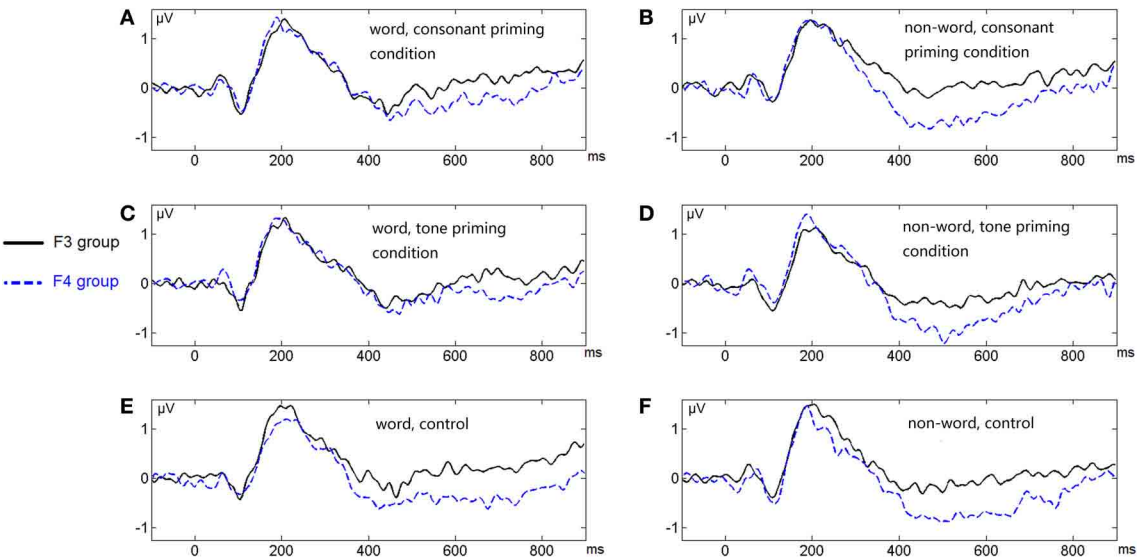


FIGURE 7 | Average ERP waveforms of the F3 and F4 electrode groups under the six conditions of Experiment 2. (A) Word, consonant priming condition; **(B)** Non-word, consonant priming condition; **(C)** Word, tone priming condition; **(D)** Non-word, tone priming condition; **(E)** Word, control condition; **(F)** Non-word, control condition.

interaction between lexicality and hemisphere) (see **Figure 9**). Such a right lateralization was also shown in Experiment 3 when comparing the tone-induced N400 with the consonant-induced N400.

EXPERIMENT 3: SEMANTIC VIOLATION IN SENTENCES

Materials

The recorded stimuli included a number of Chinese sentences. Each sentence consisted of 11 syllables, and the last two were always a verb and its object. Semantic violation was induced by changing the tone or consonant of the last syllable of a sentence. The average duration of these sentences from the onset of the first syllable to the stop of the last one was 3206.1 ms (SD = 156.1). There was no significant difference of these durations between different conditions [$F_{(2,\ 177)} = 0.697$, $p = 0.4996$, $\eta_p^2 = 0.008$]. The intensity of these sentences was adjusted to 75 dB. **Table 3** shows examples of such sentences.

Procedure

In the sentence comprehension task, participants were asked to judge whether the last syllable in a sentence was consistent with

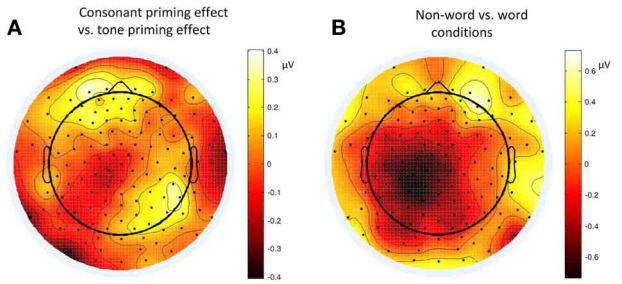


FIGURE 8 | Topographies of P2 (200–220 ms) contrasts in different conditions of Experiment 2. (A) Consonant priming effect vs. tone priming effect; **(B)** Non-word vs. word condition.

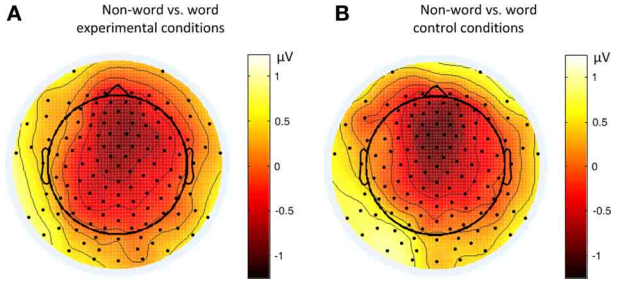


FIGURE 9 | Topographies of the contrasts of the late ERP component (500–550 ms) in different conditions of Experiment 2. (A) Non-word vs. word experimental conditions; **(B)** Non-word vs. word control conditions.

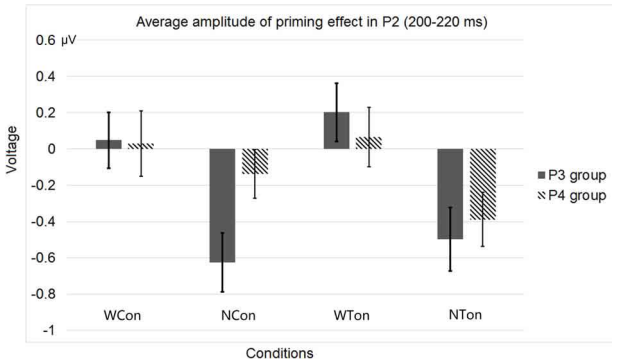


FIGURE 10 | Average P2 amplitudes within 200 and 220 ms in the P3 and P4 electrode groups in Experiment 2. "WCon" denotes the comparison between the word, consonant priming condition and the word, control condition, and "NCon" between the non-word, consonant priming condition and the non-word, control condition, "WTon" between the word, tone priming condition and the word, control condition, "NTon" between the non-word, tone priming condition and the non-word, control condition.

the context or not. Since the violation appeared toward the end of the sentence, participants were encouraged not to blink or move their body parts toward the end of the sentences.

There were three experimental conditions (see **Table 3**): in the control condition, there was no semantic violation; in the consonant violation condition, the violation was induced by changing the initial of the last syllable of the sentence; and in the tone

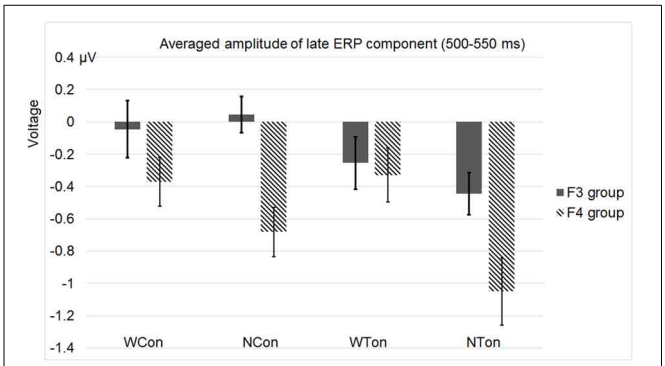


FIGURE 11 | Average amplitudes of the late ERP component around 500–550 ms in the F3 and F4 electrode groups in Experiment 2.

Table 3 | Example sentences and experimental conditions in Experiment 3.

Conditions	Example sentences
Control	早晨一起来，他就开始梳头 /ʂu1//tʰoʊ2/。 ("Once getting up in the morning, he combs his hair.")
Consonant violation	下午没有课，男生都去踢油[球] /tʰi1//joʊ2/ [/tɕʰiəu2/]。 ("With no class in the afternoon, all boys go playing oil [football].")
Tone violation	拿到学位后，他马上回过 [国] /xuəi2//kuɔ4/ [/kuɔ2/]。 ("After obtaining the degree, he immediately returns to his pass [motherland].")

The last syllable of each sentence is the target word, and all the previous ones are the context. In each condition, the Chinese transcription of the sentence and its meaning are shown. The pronunciations of the last two syllables are annotated with the IPA characters, the numbers in which denote Mandarin tones. In the violation conditions, the last syllable induces violation, and the syllable after change is still a real-syllable in Chinese. For comparison, the syllables within the square brackets are consistent with the context.

violation condition, the violation was induced by changing the tone of the last syllable of the sentence.

In each trial, a fixation first appeared on the screen and remained there. After 400 ms, one of the sentences was presented to participants. The fixation disappeared at 1000 ms after the onset of the last syllable in the sentence, and then, participants had 2000 ms for response, by pressing one of the two keys marked by "yes" and "no" in the keyboard. One half of the participants pressed the "yes" key with their left index finger and the "no" key with their right index finger. The other half did the reverse. The left or right response order was randomly assigned to participants.

There were in total 180 testing sentences, with 60 sentences in each condition. We also added ten filler sentences, each having the same length as the testing sentences, no semantic violation, and a free structure. The purpose of incorporating filler sentences was to make the yes and no responses have equal chances. The average length of each trial was 6606.1 ms ($SD = 156.1$). Participants first went through a practice session (30 trials) to familiarize the experimental paradigm. The experiment consisted of six blocks, each containing 40 sentences. These sentences included ten randomly chosen sentences from each of the three conditions, and ten filler

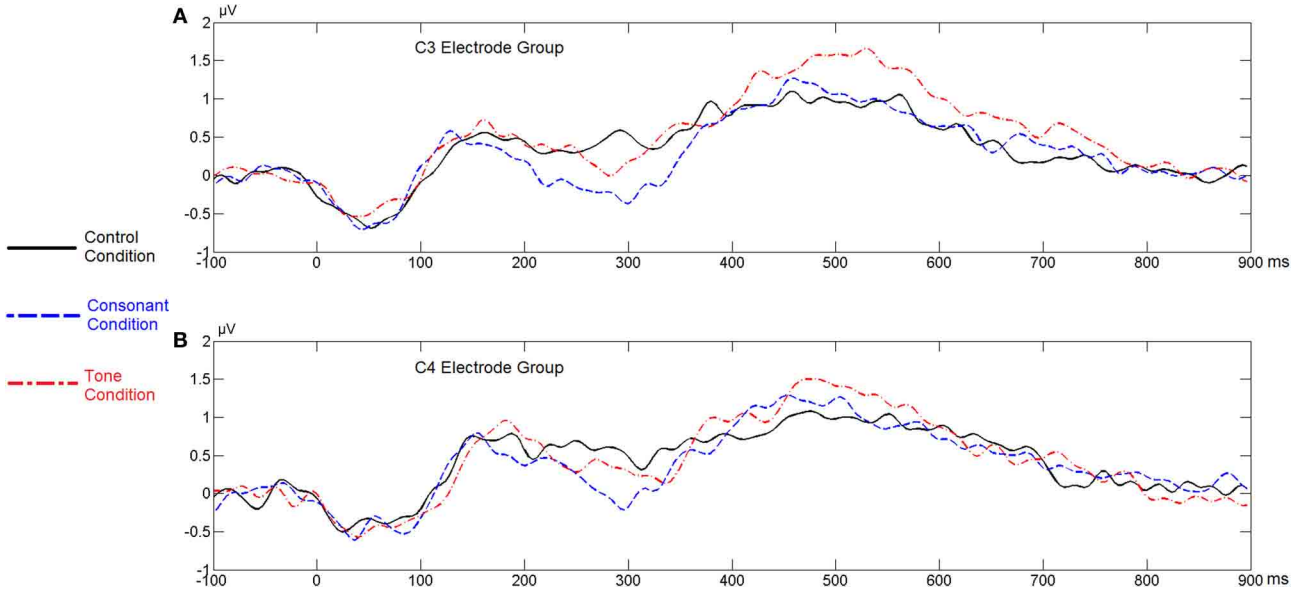


FIGURE 12 | Average ERP waveforms of the F3 and F4 electrode groups in Experiment 3. (A) C3 electrode group; **(B)** C4 electrode group.

sentences. The order of these sentences was randomized. Each block lasted about 4 min. Participants could take a 2-min break between blocks. The whole experiment lasted about 30 min.

Data analysis and results

As for the behavioral data, the response correctness was 94.3%. The averaged reaction time was 854.13 ms (SD = 186.61). Outliers greater than three times of standard deviation from mean were replaced by the mean value in each participant.

As for the ERP data, we referred to the data recorded by the four homolog pairs of adjacent electrodes including C3 and C4 [electrodes 37 (C3), 38, 42, 43 in the left hemisphere, and 105 (C4), 88, 104, 94 in the right hemisphere, according to the EGI system] for analysis. A Two-Way repeated-measures ANOVA, with violation type (consonant vs. tone) and hemisphere (left vs. right) as two factors, revealed a main effect of violation type [$F_{(1,\ 28)} = 9.622$, $p < 0.0044$, $\eta_p^2 = 0.256$], and a significant interaction between hemisphere and violation type [$F_{(1,\ 28)} = 9.573$, $p < 0.0044$, $\eta_p^2 = 0.255$]. A *post-hoc* T-test revealed a significance of the right lateralized N400 [$t_{(28)} = 2.164$, $p < 0.0391$] in the tone violation condition, and no significant lateralization in the consonant violation condition. Similar to Experiment 2, this analysis considered the control conditions, by subtracting the ERP waveforms in them from those in the experimental conditions.

Figure 12 shows the average ERP waveforms of the C3 and C4 electrode groups, **Figure 13** shows the topographies and differences of auditory N400, and **Figure 14** shows the average amplitudes of N400 (300–350 ms) in different conditions.

The significant interaction between hemisphere and violation type reflected bottom-up processing. Noticeably, there was a right lateralization in the difference between the N400 inducted by tone violation and that induced by consonant violation, which supported that bottom-up (acoustic) processing also existed in the late stage of perception.

GENERAL DISCUSSIONS

To sum up, in Experiment 1 and Experiment 2, we discovered both top-down (semantic) and bottom-up (acoustic) processing in the early time window around 200 ms. In the late stage around 300–500 ms, we only found the top-down effect in Experiment 2, probably because that the phonological primes were presented too early and the bottom-up effect could not last long enough. However, in Experiment 3, the bottom-up effect was reflected by N400 in the late stage. As indicated by the late component in Experiment 2 and Experiment 3, we suggested that both top-down and bottom-up processing existed at the late stage. The N400 component had a shorter latency in Experiment 3 than that in Experiment 2 because of the context effect, and the topography of the earlier N400 in Experiment 3 had a more central distribution compared to the late frontal N400 in Experiment 2, which is consistent to the description of N400 in time and spatial domains (Kutas and Federmeier, 2011).

RELATION BETWEEN TOP-DOWN AND BOTTOM-UP PROCESSING DURING LEXICAL TONE PERCEPTION

During lexical tone perception, the prior knowledge formed by language experience helps match a large variety of pitch contours onto clear tonal categories, and the semantic representation requires combining tonal categories with carrying syllables. Therefore, the prior knowledge of tonal categories and the lexical semantics of linking tonal categories with carrying syllables become the primary top-down factors during lexical tone perception. Since semantic and categorical information of phonemes is processed dominantly in the left hemisphere (McDermotta et al., 2003; Liebentral et al., 2005), a general left lateralization pattern during lexical tone perception reflects top-down processing. Similarly, the primary acoustic cue of lexical tone is pitch variation, and processing of pitch variations is bottom-up. Since it is widely accepted that the right hemisphere is dominant for pitch

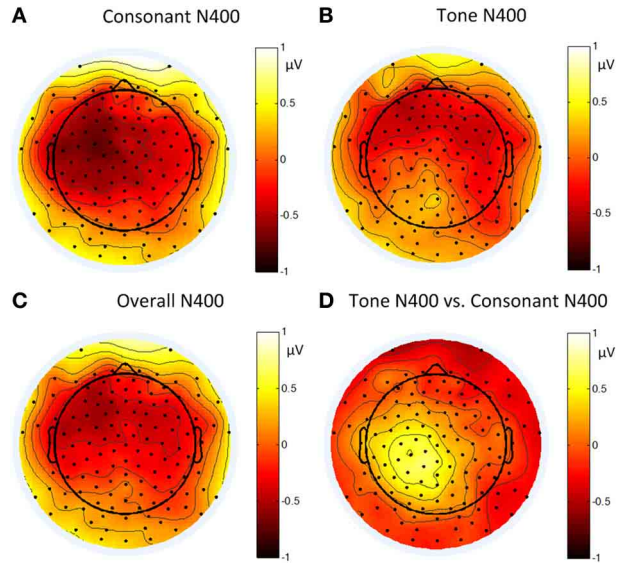


FIGURE 13 | Topographies of: (A) the N400 induced by consonant violation; (B) the N400 induced by tone violation; (C) the N400 induced by summing up consonant and tone violation; and (D) the difference between the N400 in (B) and (A) in Experiment 3.

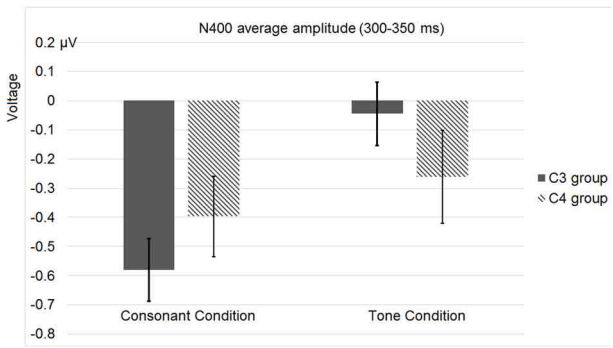


FIGURE 14 | Average N400 (300–350 ms) amplitudes of the C3 and C4 electrode groups in Experiment 3. "Consonant" and "tone" denote the consonant and tone violation conditions.

processing (Sidtis, 1981; Tenke et al., 1993), a relative right lateralization pattern during lexical tone perception also reflects bottom-up processing.

By tracing the auditory ERP components in the early and late time windows, our experiments explore the general and relative lateralization patterns in conditions with or without lexical semantics and slow or fast changing acoustic cues. Though involving distinct tasks, these experiments reveal two consistent lateralization patterns of early (P2) and late (N400) ERP components: (a) manipulation of linguistic information in words modulates lateralization: meaningful words tend to generate a greater left lateralization; and (b) manipulation of physical property of auditory input modulates lateralization: faster changing cues generate a greater left lateralization. These two patterns reflect top-down (lexical semantic) and bottom-up (acoustic phonetic) processing, respectively. Since lexical tone perception concerns both acoustic properties and lexical semantics, the observed lateralization patterns are never a simple dichotomy of purely left or right lateralization, as observed in the previous studies focusing only on one aspect of lexical tone perception. The modulation effects on lateralization at both the early and late time windows suggest that both top-down and bottom-up processing exist at different stages of perception, which support a parallel model of top-down and bottom-up processing.

THREE-STAGE, PARALLEL LEXICAL TONE PROCESSING MODEL

Previous explorations revealed that lexical tone differed from segmental cues (Ye and Connie, 1999; Lee, 2007), and that lateralization of lexical tone processing differed from that of segment processing (Li et al., 2010). Neuroimaging studies of lexical tone processing also revealed that separate brain regions were involved in perceiving lexical tones compared to segments (Gandour et al., 2003a). Apart from perception, differences between tone and segment processing were also found in lexical tone production (Liu et al., 2009). However, all these explorations did not disentangle language experience as top-down factors and acoustic cues as bottom-up factors. Treating processing of pitch information and semantic role as a cohort processing makes it difficult to figure out the cognitive processes during tone processing (Zatorre and Gandour, 2008), since lexical tone processing involves both acoustic and linguistic factors.

In our study, we regard pitch processing as bottom-up processing in lexical tone perception, since it concerns acoustic cues, and semantic processing as top-down processing, since it involves language experience. In this way, we separate these two cognitive functions called for tone perception. By exploring the temporal relation between bottom-up and top-down processing in the time windows around 200 ms and around 300–500 ms after stimulus onset, we confirm that both types of processing participate in tone perception during these early and late time periods.

Apart from these findings, there was also evidence showing a greater left lateralization of contour tones than level tones as well as a general left lateralization of Cantonese lexical tone perception in the N1 component around 100 ms after stimulus onset (Ho, 2010; Shuai et al., in press). The result of Cantonese perception reflected a bottom-up acoustic effect at around 100 ms, whereas the general left lateralization was consistent with the lateralization of top-down effect as observed in our experiments of Mandarin tone perception.

Based on the findings in our experiments and those previous studies, we propose a detailed, three-stage, parallel model of lexical tone processing. The three stages are defined based on the occurrences of different ERP components in our and previous experiments (e.g., N1, around 100 ms after stimulus onset; P2, around 100–300 ms; and N400, after 300 ms; among these components, N1 and P2 belong to the early stage and N400 belongs to the late stage).

At the first stage (before and around 100 ms after stimulus onset, as in Ho, 2010 and Shuai et al., in press), syllable initials are processed to provide the basic structure of the syllable. At this stage, if the syllable starts with a vowel or a sonorant consonant,

pitch information is available; if it starts with a voiceless consonant, there is no pitch information. In either case, tonal category is not formed yet, since the recognition of pitch variation or pattern, as slow-varying acoustic cues, requires a time window longer than 100 ms. At this stage, top-down linguistic processing also occurs, no matter whether there is contextual information before the syllable initial. With contextual information, top-down processing would become stronger, though it may play distinct roles from bottom-up processing at this stage.

At the second stage (100–300 ms after stimulus onset, as in our experimental conditions), predictions about pitch patterns and following segmental information are made, based on the information gathered at the first stage and prior language experience. At this stage, semantic information at a gross level is also activated via top-down processing, which is initiated based on the information gathered at the first stage. Since lexical tone is not fully recognized, bottom-up processing is still ongoing. According to previous literature (Kaan et al., 2007, 2008), pitch variation in the middle proportion of a syllable is the most important for recognizing contour tones for native tonal language speakers. At this stage, listeners keep integrating the incoming pitch information with the information gathered at the first stage in order to recognize the tone and the whole syllable. In this sense, both top-down prediction of tonal categories and bottom-up generalization of tonal categories are taking places at this stage.

At the third stage (after 300 ms, as in our experimental conditions), top-down processing becomes more prominent, helping listeners recognize the tone, the whole syllable, and its meaning, based on prior or previous language experience and the information gathered at the first two stages. Detailed semantic information is recognized at this stage. However, bottom-up processing keeps taking effect, helping to confirm the recognized tone and syllable.

Compatible with our findings in the series of experiments involving various levels of language processing, this parallel model of lexical tone perception can shed important lights on general speech perception models in many aspects.

First, this parallel model, as a cognitive model, proposes that top-down processing is available in both the early and the later stages of lexical tone perception, especially when contextual information is available.

Second, this parallel model refutes the claim that semantic processing (top-down) occurs always after acoustic processing (bottom-up), as in Friederici's general auditory processing model and Luo et al.'s serial model. The influence of previous language experience always exists during speech perception, especially in the case of auditory sentence processing. There are various types of cues and ample information that can serve as context for perceiving incoming syllables, and the neural system in humans always makes predictions. Even in the case of single syllable perception, if the task is linguistic relevant, top-down processing based on language experience is inevitable. As shown in our experiences, the greater left lateralization of P2 and N400 under the semantic conditions explicitly reflects such language-relevant, top-down processing.

Third, the bilateral lexical tone processing also complements to the neuroimaging models of speech perception. For example, in the dorsal-ventral pathway hypothesis of speech perception (Hickok and Poeppel, 2007), phoneme perception was regarded as involving only the left hemisphere, since such generalization did not involve tonal languages. Considering that 60–70% of world languages are tonal languages (Yip, 2002), a speech perception model leaving out tones is incomplete.

TOP-DOWN PROCESSING AT THE PREATTENTIVE STAGE

Humans often predict incoming signals based on experience. Therefore, top-down processing could accompany the whole processes of speech perception. In addition, information generated by bottom-up processing is also used to match the prediction coming from top-down processing. In this sense, top-down processing is a pre-determined process, making the relevant hemisphere or brain regions get prepared for the forthcoming task. When stimuli come in, they will evoke responses from the corresponding hemisphere or brain regions, and these responses may adjust or even alter the degrees of lateralization. A similar effect is shown in the attention or memory modulation of lateralization in DL tasks (Hugdahl, 2005; Saetrevik and Hugdahl, 2007). For example, by asking participants to attend to stimuli in either left or right ears (Hugdahl, 2005), the degree of lateralization is adjusted in favor of the side attended.

Even though long-term language experience keeps affecting automatic processing at the preattentive stage, it is generally hard to observe online top-down processing at this stage. It is only until recently that the automatic top-down processing has gained researchers' attention (Kherif et al., 2011; Wager et al., 2013). Considering that the automatic preattentive processing discovered via the MMN paradigm can be induced by either acoustic properties or long-term language experience, the MMN component is able to reflect not only bottom-up processing, but also top-down processing relevant for semantics and syntax at the preattentive stage (Pulvermüller, 2001; Pulvermüller et al., 2001a,b; Pulvermüller and Shtyrov, 2006; Penolazzi et al., 2007; Shtyrov and Pulvermüller, 2007; Gu et al., 2012). Top-down processing was also found as early as around 200 ms after stimulus onset with attention (Bonte et al., 2006). However, there was an MMN study (Luo et al., 2006) of tone perception only found a general right lateralization of tone perception. Two factors may cause this. First, many of these experiments only concern acoustic processing, i.e., bottom-up processing at the preattentive stage (e.g., Luo et al., 2006). Second, without recruiting linguistic factors, top-down processing that is dominant primarily in the left hemisphere and in the case of semantics processing would have the least influence at the preattentive stage (e.g., Xi et al., 2010).

In our study, we consider both semantic roles that require linguistic top-down processing and pitch variations that require acoustic bottom-up processing in active, language-relevant tasks, which distinguish our experiments from those MMN experiments. Via a series of tasks that involve different levels of language processing, we explicitly address top-down processing at both the early and late stages of lexical tone perception, and gather consistent evidence of the co-occurrence of top-down and bottom-up processing at those stages.

LANGUAGE PROCESSING AND GENERAL COGNITIVE FUNCTIONS

Separating lexical tone perception into bottom-up and top-down processing not only decomposes this language-specific function into general cognitive functions like sensory or memory, but also reveals that speech processing share similar mechanisms with other cognitive functions. For example, there are bottom-up attention (automatic attention shift to an unexpected event, without requiring any sort of executive processing nor involving any active engagement beforehand) and task-related top-down attention (Connor et al., 2004; Buschman and Miller, 2007; Pinto et al., 2013), both of which take part in information processing.

On the one hand, although previous work on speech perception focuses mainly on the left hemisphere, there are ample findings arguing against the existence of a centralized "core" in the left hemisphere dedicated exclusively to language processing. For example, the language function of intonation shows a right hemisphere advantage (Gandour et al., 2003b). Following a decompositional view, such right hemisphere advantage can be ascribed to consistent right hemisphere advantages of general cognitive components, including perception of slow-varying cues and emotions. Although lexical tone perception is special in the sense that it involves advantageous components in both the left hemisphere (semantic processing) and the right hemisphere (pitch processing), we can apply the same view to it. Similarly, this decompositional view can also be extended to aspects of semantics and syntax. Rather than arguing that there is no brain region that is specific for language processing, what the decompositional view emphasizes is that language must be supported by many general functions and share or recruit similar computational resources as those general functions.

On the other hand, it is not uncommon to conceptualize complex cognitive functions like language as a combination of general functions in terms of cognitive models and neural circuitry (Dehaene and Cohen, 2007; Hurley, 2007; Anderson, 2010). Take the example of attention, there is a heat debate on whether our attention is drawn voluntarily by top-down, task-dependent factor or involuntarily by bottom-up, saliency factor (Theeuwes, 1991; Buschman and Miller, 2007). Similarly, language processing also involves domain-general functions (Yip, 2002; Hurford, 2007; Fitch, 2010; Arbib, 2012). Although examining top-down and bottom-up mechanisms in cognitive functions is already prevailing, such a separation has not been commonly practiced in previous language processing literature.

Noting these, unlike the previous research that puts too much emphasis on discovering domain-specific cognitive or neural mechanisms for language processing, we advocate that decomposing language functions into more basic components (Fitch, 2010) and locating the neural networks that systematically marshal these functions (Sporns, 2011) can lead to rigorous views about the essential commonalities between language and other cognitive functions.

CONCLUSION

In this paper, we reported three ERP experiments that collectively illustrated that both bottom-up processing and top-down processing during lexical tone perception co-occurred in both the early (around 200 ms) and late (around 300–500 ms) time windows of processing. Based on these findings, we proposed a parallel lexical tone processing model that entailed both types of processing throughout various processing stages. This experimental study discussed not only the temporal relation between bottom-up and top-down processing during tone perception, but also the similarities between language processing and other cognitive functions, the latter of which pointed out an important direction in future research of language processing and general cognition.

ACKNOWLEDGMENTS

This work is supported in part by the Seed Fund for Basic Research of the University of Hong Kong. The preliminary results in this paper were first presented in the 3rd International Symposium on Tonal Aspects of Languages (TAL 2012). We thank William S.-Y. Wang for his generous support on this work. We are also grateful to the three anonymous reviewers for their useful comments on this work.

REFERENCES

Ackermann, H., Lutzenberger, W., and Hertrich, I. (1999). Hemispheric lateralization of the neural encoding of temporal speech features: a whole-head magnetencephalography study. *Brain Res. Cogn. Brain Res.* 7, 511–518. doi: 10.1016/S0926-6410(98)00054-8

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–313 doi: 10.1017/S0140525X10000853

Arbib, M. (2012). *How the Brain Got Language: The Mirror System Hypothesis*. Oxford: Oxford University Press.

Baudoin-Chial, S. (1986). Hemispheric lateralization of modern standard Chinese tone processing. *J. Neurolinguist.* 2, 189–199. doi: 10.1016/S0911-6044(86)80012-4

Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* 8, 389–395. doi: 10.1038/nn1409

Boersma, P., and Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.51. Available online at: http://www.praat.org

Bonte, M., and Blomert, L. (2004). Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. *Clin. Neurophysiol.* 115, 409–423. doi: 10.1016/S1388-2457(03)00361-4

Bonte, M., Parviainen, T., Hytonen, K., and Salmelin, R. (2006). Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cereb. Cortex* 16, 115–123. doi: 10.1093/cercor/bhi091

Buschman, T. J., and Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860–1862. doi: 10.1126/science.1138071

Connor, C. E., Egeth, H. E., and Yantis, S. (2004). Visual attention: bottom-up versus top-down. *Curr. Biol.* 14, R850–R852. doi: 10.1016/j.cub.2004.09.041

Curran, T., Tucker, D. M., Kutas, M., and Posner, M. I. (1993). Topography of the N400: brain electrical activity reflecting semantics expectancy. *Electroencephalogr. Clin. Neurophysiol.* 88, 188–209. doi: 10.1016/0168-5597(93)90004-9

Dehaene, S., and Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron* 56, 384–398. doi: 10.1016/j.neuron.2007.10.004

Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). Speech perception. *Annu. Rev. Psychol.* 55, 149–179. doi: 10.1146/annurev.psych.55.090902.142028

Dumay, N., Benraiss, A., Barriol, B., Colin, C., Radeau, M., and Besson, M. (2001). Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. *J. Cogn. Neurosci.* 13, 121–143. doi: 10.1162/089892901564117

Eichele, T., Nordby, H., Rimol, L. M., and Hugdahl, K. (2005). Asymmetry of evoked potential latency to speech sounds predicts the ear advantage in dichotic listening. *Cogn. Brain Res.* 23, 405–412. doi: 10.1016/j.cogbrainres.2005.02.017

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *J. Phonetics* 14, 3–28.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/S1364-6613(00)01839-8

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., et al. (2003b). Temporal integration of speech prosody is shaped by language experience: an fMRI study. *Brain Lang.* 84, 318–336. doi: 10.1016/S0093-934X(02)00505-9

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage* 23, 344–357. doi: 10.1016/j.neuroimage.2004.06.004

Gandour, J., Wong, D., and Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *Neuroreport* 9, 2115–2119. doi: 10.1097/00001756-199806220-00038

Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y., et al. (2002). A crosslinguistic fMRI study of spectral and temporal cues underlying phonological processing. *J. Cogn. Neurosci.* 14, 1076–1087. doi: 10.1162/089892902320474526

Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X., et al. (2003a). Neural correlates of segmental and tonal information in speech perception. *Hum. Brain Mapp.* 20, 185–200. doi: 10.1002/hbm.10137

Goldstein, B. E. (2009). *Sensation and Perception, 8th Edn.* Pacific Grove, CA: Wadsworth Inc.

Gu, F., Li, F., Wang, X., Hou, Q., Huang, Y., and Chen, L. (2012). Memory traces for tonal language words revealed by auditory event-related potentials. *Psychophysiology* 49, 1353–1360. doi: 10.1111/j.1469-8986.2012.01447.x

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113

Ho, J. P.-K. (2010). *An ERP Study on the Effect of Tone Features on Lexical Tone Lateralization in Cantonese*. Master Thesis, The Chinese University of Hong Kong.

Holcomb, P. J., and Anderson, J. E. (1993). Cross-modal semantic priming: a time-course analysis using event-related brain potentials. *Lang. Cogn. Proc.* 8, 379–411. doi: 10.1080/01690969308407583

Hsieh, L., Gandour, J., Wong, D., and Hutchins, G. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang.* 76, 227–252. doi: 10.1006/brln.2000.2382

Hugdahl, K. (2005). Symmetry and asymmetry in the human brain. *Eur. Rev.* 13, 119–133. doi: 10.1017/S1062798705000700

Hulse, S. H., Cynx, J., and Humpal, J. (1984). Absolute and relative pitch discrimination in serial pitch perception by birds. *J. Exp. Psychol. Gen.* 113, 38–54. doi: 10.1037/0096-3445.113.1.38

Hurford, J. R. (2007). *The Origin of Meaning*. Oxford: Oxford University Press.

Hurley, S. (2007). The shared circuits model: how control, mirroring and simulation can enable imitation, deliberation, and mindreading. *Behav. Brain Sci.* 31, 1–22. doi: 10.1017/S0140525X07003123

Izumi, A. (2001). Relative pitch perception in Japanese monkeys (Macaca fuscata). *J. Comp. Psychol.* 115, 127–131. doi: 10.1037/0735-7036.115.2.127

Jamison, H. L., Watkins, K. E., Bishop, D. V., and Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb. Cortex* 16, 1266–1275. doi: 10.1093/cercor/bhj068

Jia, S. Tsang, Y.-K., Huang, J., and Chen, H.-C. (2013). Right hemisphere advantage in processing Cantonese level and contour tones: evidence from dichotic listening. *Neurosci. Lett.* 556, 135–139. doi: 10.1016/j.neulet.2013.10.014

Kaan, E., Barkley, C. M., Bao, M., and Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. *BMC Neurosci.* 9:53. doi: 10.1186/1471-2202-9-53

Kaan, E., Wayland, R., Bao, M., and Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: an ERP study. *Brain Res.* 1148, 113–122. doi: 10.1016/j.brainres.2007.02.019

Kherif, F., Josse, G., and Price, C. J. (2011). Automatic top-down processing explains common left occipito-temporal responses to visual words and objects. *Cereb. Cortex* 21, 103–114. doi: 10.1093/cercor/bhq063

Krishnan, A., Gandour, J., Ananthakrishnan, S., Bidelman, G., and Smalt, C. (2011). Functional ear (a)symmetry in brainstem neural activity relevant to encoding of voice pitch: a precursor for hemispheric specialization? *Brain Lang.* 119, 226–231. doi: 10.1016/j.bandl.2011.05.001

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–208. doi: 10.1126/science.7350657

Lau, E. F., Phillips, C., and Peoppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Lau, H. C., and Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *J. Neurosci.* 27, 5805–5811. doi: 10.1523/JNEUROSCI.4335-06.2007

Lee, C.-Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Lang. Speech* 50, 101–123. doi: 10.1177/00238309070500010501

Li, X., Gandour, J. T., Talavage, T., Wong, D., Hoffa, A., Lowe, M., et al. (2010). Hemispheric asymmetries in phonological processing of tones vs. segmental units. *Neuroreport* 21, 690–694. doi: 10.1097/WNR.0b013e32833b0a10

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36. doi: 10.1016/0010-0277(85)90021-6

Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., and Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cereb. Cortex* 20, 2958–2970. doi: 10.1093/cercor/bhq045

Liebentral, E., Binder, J. R., Spitzer, S. M., Possing, E. T., and Medler, D. A. (2005). Neural substrates of phonemic perception. *Cereb. Cortex* 15, 1621–1631. doi: 10.1093/cercor/bhi040

Liu, L., Deng, X., Peng, D., Cao, F., Ding, G., Jin, Z., et al. (2009). Modality- and task-specific brain regions involved in Chinese lexical processing. *J. Cogn. Neurosci.* 21, 1473–1487. doi: 10.1162/jocn.2009.21141

Liu, L., Peng, D., Ding, G., Jin, Z., Zhang, L., Li, K., et al. (2006). Dissociation in the neural basis underlying Chinese tone and vowel production. *Neuroimage* 29, 515–523. doi: 10.1016/j.neuroimage.2005.07.046

Luck, S. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

Luo, H., Ni, J.-T., Li, Z.-H., Li, X.-O., Zhang, D.-R., Zeng, F.-G., et al. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19558–19563. doi: 10.1073/pnas.0607065104

Maess, B., Herrmann, C. S., Hahne, A., Nakamura, A., and Friederici, A. D. (2006). Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing. *Brain Res.* 1096, 163–172. doi: 10.1016/j.brainres.2006.04.037

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0

McDermotta, K. B., Petersen, S. E., Watson, J. M., and Ojemanna, J. G. (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia* 41, 293–303. doi: 10.1016/S0028-3932(02)00162-8

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Penolazzi, B., Hauk, O., and Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related potentials. *Biol. Psychol.* 74, 374–388. doi: 10.1016/j.biopsycho.2006.09.008

Pinto, Y., van der Leji, A. R., Sligte, I. G., Lamme, V. A. F., and Scholte, S. H. (2013). Bottom-up and top-down attention are independent. *J. Vis.* 13:16. doi: 10.1167/13.3.16

Praamstra, P., and Stegeman, D. F. (1993). Phonological effects on the auditory N400 event-related brain potential. *Cogn. Brain Res.* 1, 73–86. doi: 10.1016/0926-6410(93)90013-U

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends Cogn. Sci.* 5, 517–524. doi: 10.1016/S1364-6613(00)01803-9

Pulvermüller, F., Assadollahi, R., and Elbert, T. (2001a). Neuromagnetic evidence for early semantic access in word recognition. *Eur. J. Neurosci.* 13, 201–205. doi: 10.1046/j.0953-816X.2000.01380.x

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., et al. (2001b). Memory traces for words as revealed by the mismatch negativity (MMN). *Neuroimage* 14, 607–616. doi: 10.1006/nimg.2001.0864

Pulvermüller, F., and Shtyrov, Y. (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Prog. Neurobiol.* 79, 49–71. doi: 10.1016/j.pneurobio.2006.04.004

Saetrevik, B., and Hugdahl, K. (2007). Priming inhibits the right ear advantage in dichotic listening: implications for auditory laterality. *Neuropsychologia* 45, 282–287. doi: 10.1016/j.neuropsychologia.2006.07.005

Schapkin, S. A., Gusev, A. N., and Kuhl, J. (2000). Categorization of unilaterally presented emotional words: an ERP analysis. *Acta Neurobiol. Exp.* 60, 17–28. Available online at: http://www.ane.pl/pdf/6003.pdf

Schirmer, A., Tang, S. L., Penney, T. B., Gunter, C. T., and Chen, H. C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *J. Cogn. Neurosci.* 17, 1–12. doi: 10.1162/0898929052880057

Shtyrov, Y., and Pulvermüller, F. (2007). Early MEG activation dynamics in the left temporal and inferior frontal cortex reflect semantic context integration. *J. Cogn. Neurosci.* 19, 1633–1642. doi: 10.1162/jocn.2007.19.10.1633

Shuai, L., Gong, T., Ho, J. P.-K., and Wang, W. S.-Y. (in press). "Hemispheric lateralization of perceiving Cantonese contour and level tones: an ERP study," in *Studies on Tonal Aspect of Languages* (*Journal of Chinese Linguistics Monograph Series, No. 25*), ed W. Gu (Hong Kong: Journal of Chinese Linguistics).

Sidtis, J. J. (1981). The complex tone test: implications for the assessment of auditory laterality effects. *Neuropsychologia* 19, 103–112. doi: 10.1016/0028-3932(81)90050-6

Sporns, O. (2011). *The Networks of the Brain*. Cambridge, MA: MIT Press.

Stevens, K. N. (2002). Toward a model of lexical access based on acoustic landmarks and distinctive features. *J. Acoust. Soc. Am.* 111, 1872–1891. doi: 10.1121/1.1458026

Tachibana, H., Minamoto, H., Takeda, M., and Sugita, M. (2002). Topographical distribution of the auditory N400 component during lexical decision and recognition memory tasks. *Int. Congr. Ser.* 1232, 197–202. doi: 10.1016/S0531-5131(01)00818-4

Tenke, C. E., Bruder, G. E., Towey, J. P., Leite, P., and Sidtis, J. J. (1993). Correspondence between brain ERP and behavioral asymmetries in a dichotic complex tone test. *Psychophysiology* 30, 62–70. doi: 10.1111/j.1469-8986.1993.tb03205.x

Tervaniemia, M., and Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Res. Rev.* 43, 231–246. doi: 10.1016/j.brainresrev.2003.08.004

Theeuwes, J. (1991). Exogenous and endogenous control of attention—the effect of visual onsets and offsets. *Percept. Psychophys.* 49, 83–90. doi: 10.3758/BF03211619

Tsang, Y.-K., Jia, S., Huang, J., and Chen, H.-C. (2010). ERP correlates of pre-attentive processing of Cantonese lexical tones: the effects of pitch contour and pitch height. *Neurosci. Lett.* 487, 268–272. doi: 10.1016/j.neulet.2010.10.035

Van Lancker, D., and Fromkin, V. A. (1973). Hemispheric specialization for pitch and "tone": evidence from Thai. *J. Phonetics* 1, 101–109.

Wager, E. E., Peterson, M. A., Folstein, J. R., and Scalf, P. E. (2013). Automatic top-down processes mediate selective attention. *J. Vis.* 13:137. doi: 10.1167/13.9.137

Wang, W. S.-Y. (1967). Phonological features of tone. *Int. J. Am. Linguist.* 33, 93–105. doi: 10.1086/464946

Wang, Y., Behne, D., Jongman, A., and Sereno, J. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Appl. Psycholinguist.* 25, 449–466. doi: 10.1017/S0142716404001213

Wang, Y., Jongman, A., and Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain Lang.* 78, 332–348. doi: 10.1006/brln.2001.2474

Wioland, N., Rudolf, G., Metz-Lutz, M. N., Mutschler, V., and Marescaux, C. (1999). Cerebral correlates of hemispheric lateralization during a pitch discrimination task: an ERP study in dichotic situation. *Clin. Neurophysiol.* 110, 516–523. doi: 10.1016/S1388-2457(98)00051-0

Wong, P. C. M., Warrior, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., et al. (2008). Volume of left Heschl's gyrus and linguistic pitch learning. *Cereb. Cortex* 18, 828–836. doi: 10.1093/cercor/bhm115

Xi, J., Zhang, L., Shu, H., Zhang, Y., and Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity (MMN). *Neuroscience* 170, 223–231. doi: 10.1016/j.neuroscience.2010.06.077

Ye, Y., and Connie, C. M. (1999). Processing spoken Chinese: the role of tone information. *Lang. Cogn. Proc.* 14, 609–630. doi: 10.1080/016909699386202

Yin, P., Fritz, J. B., and Shamma, S. A. (2010). Do ferrets perceive relative pitch? *J. Acoust. Soc. Am.* 127, 1673–1680. doi: 10.1121/1.3290988

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139164559.006

Zatorre, R., and Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1087–1104. doi: 10.1098/rstb.2007.2161

Zatorre, R. J., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953. doi: 10.1093/cercor/11.10.946

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46. doi: 10.1016/S1364-6613(00)01816-7

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 July 2013; paper pending published: 11 December 2013; accepted: 09 March 2014; published online: 25 March 2014.

Citation: Shuai L and Gong T (2014) Temporal relation between top-down and bottom-up processing in lexical tone perception. Front. Behav. Neurosci. ***8****:97. doi: 10.3389/fnbeh.2014.00097*

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 28 January 2014
doi: 10.3389/fnbeh.2014.00016

Left-right compatibility in the processing of trading verbs

Carmelo M. Vicario[1,2,3]* and Raffaella I. Rumiati[2]

[1] School of Psychology, The University of Queensland, Brisbane, QLD, Australia
[2] Cognitive Neuroscience Sector, SISSA, Trieste, Italy
[3] School of Psychology, Bangor University, Bangor, UK

Edited by:
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA

Reviewed by:
Daniel Casasanto, University of Chicago, USA
Feng Kong, Beijing Normal University, China

****Correspondence:***
Carmelo M. Vicario, School of Psychology, Bangor University, Brigantia Building, Bangor, Gwynedd, Wales LL57 2AS, UK
e-mail: carmelo.vicario@uniroma1.it

The research investigating the nature of cognitive processes involved in the representation of economical outcomes is growing. Within this research, the mental accounting model proposes that individuals may well use cognitive operations to organize, evaluate, and keep track of their financial activities (Thaler, 1999). Here we wanted to test this hypothesis by asking to a group of participants to detect a syntax mistake of verbs indicating incoming and going out activities related to economical profit (trading verbs), swapping (swapping verbs) and thinking (thinking verbs). We reported a left-right compatibility for trading verbs (i.e., participants were faster with their right hand while detecting verb referring to a monetary gain with respect to a monetary loss; and faster with their left hand while detecting a monetary loss with respect to a monetary gain). However, this pattern of result was not reported while detecting swapping verbs. Results are discussed taking into account the mental accounting theory as well as to the spatial mapping of valence hypothesis.

Keywords: language, economics, SNARC effect, mental accounting theory, spatial valence hypothesis

INTRODUCTION

The interest in the nature of cognitive processes involved in the representation of economical outcomes has been growing in recent years (see e.g., Wu et al., 2012 for a recent review). Several studies in cognitive sciences and financial economics propose the inextricable interdependence between rationality and emotion (Grossberg and Gutowski, 1987; Damasio, 1994; Elster, 1998; Loewenstein, 2000; Harvey et al., 2010) in influencing human economical choices and behaviors (see also Glimcher and Rustichini, 2004). For instance, Loewenstein (2000) highlights the impact of immediate emotions, as well as wide range of visceral factors associated with them, in determining systematic behaviors that could also be amenable in a formal model. Moreover, a crucial role is played by the activation of reward-related brain areas, such as the striatum (Fehr and Camerer, 2007).

Nevertheless, other factors, beyond those mentioned above, might play a role in processing and representing economic outcomes. The mental accounting theory (Thaler, 1980), a model developed in the field of behavioral economy, proposes some intriguing suggestions in this direction. This model attempts to describe the process whereby people code, categorize and evaluate economic outcomes. According to this model, individuals use cognitive operations to organize, evaluate, and keep track of their financial activities (Thaler, 1999). Since this model holds that accounting operations are engaged in evaluating economical outcomes, one could expect a specific role of the mathematical brain processes in representing financial meanings.

A way to test this hypothesis is provided by the study of language. In fact, one could argue that the same cognitive processes active while manipulating quantity might be involved in processing financial words. Thus, following this suggestion, linguistic items such as verbs referring to monetary *gain* and/or *loss* could be conceptualized in terms of mental shifts toward higher or lower quantities or as two mental accounting operations such as *addition* and *subtraction*. This proposal originates from the evidence that Western populations are endowed with a left-to-right Mental Number Line (MNL) for representing quantities, from lower to higher, respectively (Dehaene et al., 1990). Moreover, Knops et al. (2009) have shown that during no-symbolic addition, subjects preferentially selected numbers at the upper right location, whereas during no-symbolic subtraction, they were biased toward the upper left location.

In consideration of these findings one could expect a similar left-to-right spatial encoding for the representation of linguistic terms which refer to monetary gain and loss. Accordingly, one can hypothesize that the same cognitive mechanisms underlying the representation of quantity are covertly engaged when people read verbs associable to a monetary gain or loss. Given the direct relation, in the cognitive system, between quantity (low vs. high), arithmetic operations (addition vs. subtraction) and spatial coordinates (left vs. right), one could expect to detect faster reaction times (RTs) in using the right hand while processing verbs related to a monetary *gain* with respect to verbs related to a monetary *loss* (namely trading verbs). On the other hand, one could expect faster RTs in using the left hand while processing verbs associated to a monetary *loss* with respect to verbs associated to a monetary *gain*. Participants were also performed a second block of stimuli (namely swapping verbs) which refer to verbs describing incoming and coming out outcomes perceived as an exchange. In fact, the main difference

between trading and the swapping verbs is that only trading verbs explicitly suggest the meaning of "economical profit", although a monetary outcome can be associated to both categories (e.g., money loss vs. money donation). We use of swapping verbs to create an incoming vs. going out condition in absence of high economical relevance (compared to the trading category). In this way, we could have more elements to understand whether the origin of the hypothesized left-right encoding is linked to the economical relevance of the linguistic term rather than to the incoming vs. going out meaning covertly suggested by all these verbs.

MATERIALS AND METHODS

PARTICIPANTS

Twenty-two right-handed graduate students (10 men, 12 women, mean age: 26 ± 7.03 years) recruited from the University of Trieste, participated in the studies after providing verbal informed consent. The experiment was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. Participants received a payment of 10 Euros for having taken part in this study.

PROCEDURE AND INSTRUMENTS

Using their left and right index fingers, participants were required to establish, as soon as possible and in two consecutive sessions (counterbalanced design), whether 108 verbal stimuli contained (or not) a syntax mistake. The task was identical for both trading and swapping verbs.

Trading verbs block

Fifty-four of 108 items were spelt correctly; of these, 18 (6 verbs × 3 trials) indicated a monetary gain and 18 (6 verbs × 3 trials) indicated a monetary loss. Moreover this block included 18 *thinking* verbs (6 verbs × 3 trials) as control items (see **Table 1** for the complete list).

Swapping verbs block

Fifty-four of 108 items were spelt correctly; of these, 18 (6 verbs × 3 trials) indicated a receiving action and 18 (6 verbs × 3 trials) indicated a giving action. Even in this block were included 18 *thinking* verbs (6 verbs × 3 trials) as control items (see **Table 1** for the complete list).

All verbs were presented in first person and in the simple present tense. Each trial was preceded by an alerting sentence (ready) lasting 500 ms and followed by fixation cross lasting 500 ms. The within subjects variable was the responding finger. Incorrect responses (Trading verbs: 2.82%; Swapping verbs: 3.05%) were not considered in the analysis.

STATISTICAL ANALYSIS

The RTs performance in detecting stimuli written correctly was analyzed using ANOVA for repeated measures, with VERB (trading vs. swapping), MANUAL RESPONSE (left vs. right) and MEANING (incoming, thinking and going out) as main factors. Trading and swapping verbs were presented in separated blocks. *Post-hoc* comparisons were performed using the Duncan *post-hoc* test. For all statistical analyses, a p-value of 0.05 was considered to be significant. Data analysis was performed using Statistica software, version 8.0, StatSoft, Inc., Tulsa, USA. We also performed a permutation analysis where we relabelled and shuffled verbs across conditions. This analysis was conducted to have an approximation of what could have happen if Swapping and Trading verbs were randomly assigned. In this case, the analysis was conducted by using Matlab software, R 2013 A version. The number of permutation selected for this procedure was 1.000; p-value of 0.05 was considered to be significant.

RESULTS

In order to evaluate the grade of familiarity of participants with the proposed verbs, they were asked to use a five point rating scale to have a subjective measure of their level of experience/familiarity. Therefore, the higher the reported score the higher the subjective experience/familiarity with a verb.

TRADING VERBS

We detected a significant difference ($F_{(2,40)} = 35.00$, $p < 0.001$). *Post-hoc* comparisons revealed that both Incoming vs. Going out verbs significantly differed from the control category (*Thinking*: $M = 4.753 \pm 0.250$ vs. *Incoming*: $M = 3.531 \pm 0.922$ SD, $p < 0.001$; *Thinking*: $M = 4.789 \pm 0.234$ vs. *Going out*: $M = 3.515 \pm 0.783$ SD, $p < 0.001$), while no difference was observed between them ($p = 0.100$).

Table 1 | This table reports the complete list of items used for the three verb categories.

Trading	**Gain**	Incassare	Riscuotere	Ricavare	Guadagnare	Intascare	Arricchire
	verbs	(To cash)	(To cash)	(To derive)	(To gain)	(To rake in)	(To enrich)
	Loss	Pagare	Risarcire	Indennizzare	Perdere	Saldare	Impoverire
	verbs	(To pay)	(To compensate)	(To indemnify)	(To lose)	(To pay)	(To impoverish)
Swapping	**Receiving**	Ricevere	Ereditare	Accettare	Accogliere	Rilevare	Acquisire
	verbs	(to receive)	(to inherit)	(to accept)	(to welcome)	(to take)	(to acquire)
	Giving	Regalare	Donare	Offrire	Devolvere	Consegnare	Porgere
	verbs	(to give)	(to donate)	(to offer)	(to devolute)	(to deliver)	(to hand)
Thinking Verbs		Ritenere	Credere	Supporre	Pensare	Immaginare	Sperare
		(to consider)	(to believe)	(to suppose)	(to think)	(to imagine)	(to hope)

The corsive items refer to the verbs (written in Italian) originally used in the task. In the parenthesis it is proposed the English translation.

SWAPPING VERBS

We detected a significant difference ($F_{(2,40)} = 67.50$, $p < 0.001$). *Post-hoc* comparison showed that both types of *Swapping Verbs* significantly differed from the control category (*Thinking:* $M = 4.734 \pm 0.260$ vs. *Incoming*: $M = 3.174 \pm 0.749$, $p < 0.001$; *Thinking:* $M = 4.734 \pm 0.260$ vs. *Going out:* $M = 3.795$, $p < 0.001$). A significant difference was found also between swapping verbs ($p < 0.001$) showing that going out verbs were perceived as more familiar than incoming verbs.

The analysis of RTs was conducted by excluding 2 participants from the original sample: one participant was excluded because his low performance accuracy (i.e., <80%); the other participant was excluded because he did not complete the experiment (i.e., the swapping block). According to our research hypothesis, we detected a significant result for the VERB * MEANING * MANUAL RESPONSE interaction factor ($F_{(2,38)} = 5.24$, $p = 0.009$). *Post-hoc* comparison shows a double dissociation in RTs for verbs of the trading category. In particular, participants were significantly faster in detecting going out verbs ($M = 998.7 \pm 59.69$) with respect to incoming verbs ($M = 1063.6 \pm 62.94$) while using their left hand ($p = 0.013$); and incoming verbs ($M = 1056.9 \pm 67.02$) with respect to going out verbs ($M = 1109.9 \pm 75.74$) while using their right hand ($p = 0.042$). On the other hand, *post-hoc* comparison concerning verbs of the swapping category only showed faster RTs in detecting going out items, with respect to incoming items. This was significant for both left ($p = 0.007$) and right ($p = 0.001$) hand responses (see **Figure 1** for details).

This pattern of results was confirmed by the permutation analysis in almost all cases. In particular we show significant RTs difference for the trading category by comparing going out with respect to incoming verbs for the left ($p = 0.024$) and the right ($p = 0.005$) hand; We also detected a significant RTs difference for the swapping category by comparing going out with respect to incoming verbs for the right hand ($p < 0.001$). However, this difference is not significant for the left hand ($p = 0.1727$).

Ngram viewer by Google was also used to provide an idea about the frequency of use of these verbs in the Italian language. In particular, we focused on the temporal interval between the 1998 and the 2008 (i.e., the most recent temporal range available with Google Ngram viewer).

First, we performed a *t*-test analysis by comparing the Ngram viewer output (i.e., the amount of citations) provided for the 12 Trading verbs ($M = 7921698,8$) with that of the 12 Swapping verbs ($M = 12908273,3$). Results did not report a significant difference ($t = -0.91$, $p = 0.367$). We also performed two repeated measures ANOVA in which we compared the Google Ngram viewer output of Incoming, Going-out and Thinking verbs of both Trading and Swapping categories. The analysis for the Trading category documented a significant differences ($F_{(2,10)} = 5.99$, $p = 0.01$). In particular we found that Thinking verbs ($M = 38387857,5$) were more frequently cited than Incoming ($M = 3805269,5$), ($p = 0.009$) and Going out ($M = 12038128,3$) verbs ($p = 0.030$). No difference was reported by comparing them to each other ($p = 0.448$). On the other hand, we did not detect a significant difference for the swapping block, although the trend ($F_{(2,10)} = 3,67$, $p = 0.06$). **Figure 2** shows a plot of the three verb categories.

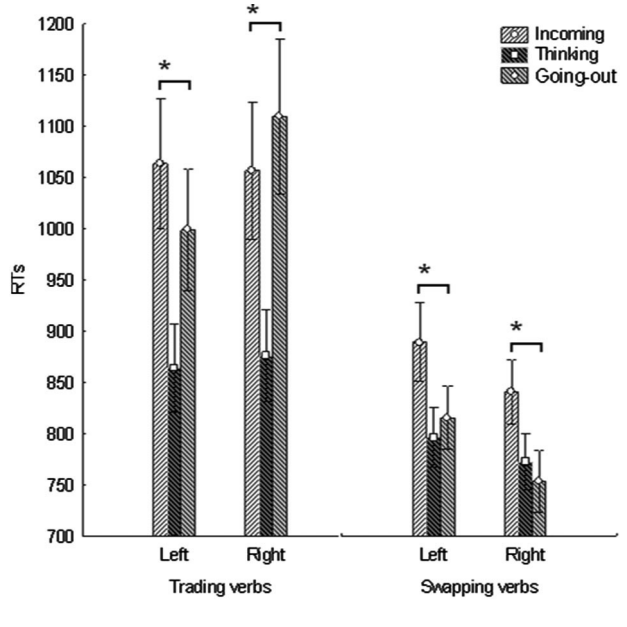


FIGURE 1 | The interplay between manual response and the comprehension of trading and swapping verbs. Reaction times for detecting the correctly spelt items of both *trading* and *swapping* categories. Vertical bars indicate one Standard Error. The "*" indicates significant *post-hoc* comparison differences.

DISCUSSION

The purpose of this study was to address the question suggested by the mental accounting theory, that is, whether people use cognitive operations for processing financial activities (Thaler, 1999). This hypothesis was addressed by studying the participants' performance in a linguistic task.

Recently, Baroni et al. (2013) have documented a left-right compatibility by using financial words (i.e., *monopoly*, *salary*, *discount*) with faster left hand RTs while detecting words indicating loss concepts (i.e., unemployment) and, *vice versa*, faster right had RTs while detecting words indicating gain concepts (i.e., salary). However, the effect on RTs was only found when participants were required to explicitly discriminate between gain and loss, while there was failure in detecting this effect when they were required to discriminate between economic and no economic terms. On the other hand, in a further experiment, in which participants were asked to arbitrary allocate financial words along a line, the authors documented a left space preference in a spontaneous allocation of "loss" words and, *vice versa*, a right space preference in the spontaneous allocation of "gain" words. This last experiment suggests a left to right encoding for gain and loss meaning even in implicit tasks. Our study differs from the research conducted by Baroni et al. (2013) not only with respect to the adopted procedure (i.e., participants were asked to identify a syntax mistake), but also with respect to the verbal material [i.e., Baroni et al. (2013), used financial words while we used verbs which referred to incoming vs. going out outcomes, with high (i.e., trading) vs. low (i.e., swapping) economical relevance].

FIGURE 2 | Google Ngram viewer outputs for the three verb categories. From top to down this figure show the output % provided by google Ngram viewer for Trading, Swapping and Thinking verbs from 1998 to 2008.

According to the initial prediction we found a left-to-right spatial compatibility for trading verbs. In particular, participants were significantly faster in detecting verbs indicating a monetary loss (i.e., going out) with respect to verbs indicating a monetary gain (i.e., incoming) while using their left hand; on the other hand, participants were faster in detecting verbs indicating a monetary gain with respect to verbs indicating a monetary loss while using their right hand. However, we did not detect a left-to-right

spatial compatibility for swapping verbs. This suggests that the incoming vs. going out meaning implicitly associated to both verb categories might be not relevant, *per se*, in explaining the left-right compatibility found for the trading category. In fact, if this was the case, a left-right compatibility would have been detected also for the swapping category.

In the light of this argument, one could argue at least two alternative suggestions to explain the current result.

One possibility is that the left-right compatibility reported for trading verbs might reflect the higher economical relevance of this linguistic category, compared to the swapping category. As already discussed in the introduction, a dense literature (Dehaene et al., 1990; Loetscher et al., 2008; Vicario, 2012; Holmes and Lourenco, 2013; Shaki and Fischer, 2013) has repeatedly demonstrated a left-to-right mapping for low and high numbers, which is reflected in the so called SNARC effect. Accordingly, the current results can be explained assuming that when Western participants read verbs (with high economical relevance) associated to a monetary *loss*, their cognitive activation moves "leftward" as when detecting small numbers and/or performing arithmetical subtraction; *vice versa*, reading verbs (with high economical relevance) associated to a monetary *gain* activates a mental rightward shift as when detecting high numbers and while performing arithmetic addition (Knops et al., 2009). According to this interpretation our data can be intended as a support to the mental accounting theory (Thaler, 1999) stating that people use cognitive operations for processing financial activities. In fact, the current result suggests that linguistic terms referring to economics are spatially mapped similarly to numbers. This implies the suggestion that the left-right compatibility reported for trading verbs might reflect a SNARC-like effect for this linguistic material as well as for magnitude processing (Vicario and Martino, 2010).

Several arguments can be provided in support of this hypothesis. First, a common Intraparietal Sulcus (IPS) activation has been observed when participants performed calculation, linguistic and saccadic movement tasks (Sereno et al., 2001). In fact, this area has been identified by these authors as the neural correlate of the mental accounting and linguistic competence interplay; Second, learning difficulties in mathematics (i.e., developmental dyscalculia) frequently co-occur with impairments in reading (i.e., developmental dyslexia). This co-morbidity could be related to the malfunctioning of the left angular gyrus, a brain area that has been found to be affected in patients with Gerstmann (1940) who show not only acalculia but also left–right disorientation; Third, patients with cortico-basal degeneration (CBD) can show a severe difficulty in understanding small numbers as well as quantifier terms (McMillan et al., 2006). They also provided a further support to this view by performing a neuroimaging study investigating quantifier comprehension in healthy adults (McMillan et al., 2005). Semantic theorists (e.g., Szymanik and Zajenkowski, 2010) make a general distinction between first-order quantifiers, which identify a number state (e.g., "some" or "at least 3") and higher-order quantifiers, which are those not expressible in first-order logic (e.g., "most" or "every other"). McMillan et al. (2005) reported that first-order and higher-order quantifiers both recruit right inferior parietal cortex, suggesting that a number processing component contributes to quantifier comprehension. In fact, parietal activation was also widely reported in subjects asked to perform a simple number processing (Cohen et al., 2000; Kazui et al., 2000; Pinel et al., 2001; Simon et al., 2002) or arithmetic task (Menon et al., 2000; Knops et al., 2009; Krueger et al., 2011).

An alternative, no less important, interpretation to the current results might refer to the *body-specificity hypothesis* (Casasanto, 2009, 2011; Kominsky and Casasanto, 2013; Kong, 2013), stating that people conceptualize bad and good in terms of left-right spatial encoding, according to their handedness. For example, Casasanto (2009) showed that right-handers tend to associate rightward space with positive ideas and leftward space with negative ideas (this pattern was reversed in left-handers). In fact, while trading verbs might be conceptualized as endowed of an emotionally positive (i.e., incoming) and negative (i.e., going out) meanings, all swapping verbs might be interpreted positively (e.g., "donation" is easily interpreted as "positive", although it indicates a going out outcome).

The results reported for the swapping category provide some support to the Space-valence hypothesis. In fact, the going out verbs of the swapping category such as "to donate" (which were the most positive) show the largest advantage for the right hand, and the moderately positive incoming verbs like "to receive" show an advantage in the same direction. Moreover, the permutation test did not confirm a significant difference comparing going out with respect to incoming swapping verbs when using the left hand. This might be explained with the fact that all our participants were right handed. In fact, the spatial mapping of valence hypothesis predicts a performance advantage with the dominant hand. Therefore, according to this view, one could argue that the left-right compatibility reported in our study might be ruled by the "value" (i.e., positive vs. negative) of the presented verbs. The spatial mapping of valence hypothesis might represent a valid interpretation for explaining the current results. However, this study does not provide definitive evidence in support of this interpretation since we did not test left-handed people.

Worthy of some discussion is the difference in the familiarity ranking score provided by our participants for the three verb categories. In fact, going out swapping verbs were perceived as more familiar than incoming swapping verbs. This difference in familiarity scores for the swapping category might have played some role in the detection of these verbs, although we don't believe that this factor might explain the absence of a left/right compatibility for this linguistic category.

Our study bears some important limitations that might be addressed in future works. First, we did not collect any ranking about the economical relevance subjectively associated to the verbs presented in this research. Second, trading and swapping verbs were administered in separated blocks. This might represent an issue since verbs within blocks might have interacted such as cueing. In fact, thinking verbs, which were used as a control condition, were detected faster in the swapping block than in the trading block. However, the permutation analysis suggests that the reported effects for the trading category are not related to the blocked design. Finally, we did not test RTs performance in left handed participants, this because our research

goal was testing the existence of a SNARC like-effect for linguistic items associated to the economical profit category (i.e., trading verbs).

Further investigations including brain imaging and non-invasive brain stimulation methods, but also left-handed participants are needed to clarify whether the current results underlie the linguistic representation of economical outcomes.

ACKNOWLEDGMENTS

We would like to thank Miss Anica Newman for her help in the checking of English grammar and Dr. Andrea Pavan for the assistance with the permutation analysis.

REFERENCES

Baroni, G., Bolzani, R., Di Pellegrino, G., Moretto, G., and Nicoletti, R. (2013). La rappresentazione spaziale di parole economiche. *G. Ital. Psicol.* a. XL. 1, 231–239. doi: 10.1421/73993

Casasanto, D. (2009). Embodiment of abstract concepts: good and bad in right- and left-handers. *J. Exp. Psychol. Gen.* 138, 351–367. doi: 10.1037/a0015854

Casasanto, D. (2011). Different bodies, different minds: the body-specificity of language and thought. *Curr. Dir. Psychol. Sci.* 20, 378–383. doi: 10.1177/0963721411422058

Cohen, L., Dehaene, S., Chochon, F., Lehericy, S., and Naccache, L. (2000). Language and calculation within the parietal lobe: a combined cognitive, anatomical and fMRI study. *Neuropsychologia* 38, 1426–1440. doi: 10.1016/s0028-3932(00)00038-5

Damasio, A. R. (1994). *Descartes' Error Emotion, Reason and the Human Brain*. New York: Avon Books.

Dehaene, S., Dupoux, E., and Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *J. Exp. Hum. Percept. Perform.* 16, 626–641. doi: 10.1037/0096-1523.16.3.626

Elster, J. (1998). Emotions and economic theory. *J. Econom. Literat.* 36, 47–74.

Fehr, E., and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* 11, 419–427. doi: 10.1016/j.tics.2007.09.002

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia - local diagnostic value. *Arch. Neurol. Psychiatry* 44, 398–408. doi: 10.1001/archneurpsyc.1940.02280080158009

Glimcher, P. W., and Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. *Science* 306, 447–452. doi: 10.1126/science.1102566

Grossberg, S., and Gutowski, W. (1987). Neural dynamics of decision making under risk: affective balance and cognitive emotional interactions. *Psych. Rev.* 94, 300–318. doi: 10.1037//0033-295x.94.3.300

Harvey, A. H., Kirk, U., Denfield, G. H., and Montague, P. R. (2010). Monetary favors and their influence on neural responses and revealed preference. *J. Neurosci.* 30, 9597–9602. doi: 10.1523/JNEUROSCI.1086-10.2010

Holmes, K. J., and Lourenco, S. F. (2013). When numbers get heavy: is the mental number line exclusively numerical? *PLoS One* 8:e58381. doi: 10.1371/journal.pone.0058381

Kazui, H., Kitagaki, H., and Mori, E. (2000). Cortical activation during retrieval of arithmetical facts and actual calculation: a functional magnetic resonance imaging study. *Psych Clin. Neurosci.* 54, 479–485. doi: 10.1046/j.1440-1819.2000.00739.x

Knops, A., Viarouge, A., and Dehaene, S. (2009). Dynamic representations underlying symbolic and nonsymbolic calculation: evidence from the operational momentum effect. *Attent. Percept. Psychophys.* 71, 803–821. doi: 10.3758/app.71.4.803

Kominsky, J., and Casasanto, D. (2013). Specific to whose body? Perspective taking and the spatial mapping of valence. *Front. Psy.* 4:266. doi: 10.3389/fpsyg.2013.00266

Kong, F. (2013). Space-valence associations depend on handedness: evidence from a bimanual output task. *Psychol. Res.* 77, 773–779. doi: 10.1007/s00426-012-0471-7

Krueger, F., Landgraf, S., van der Meer, E., Deshpande, G., and Hu, X. (2011). Effective connectivity of the multiplication network: a functional MRI and multivariate granger causality mapping study. *Hum. Brain Mapp.* 32, 1419–1431. doi: 10.1002/hbm.21119

Loetscher, T., Schwarz, U., Schubiger, M., and Brugger, P. (2008). Head turns bias the brain' s internal random generator. *Curr. Biol.* 18, R60–R62. doi: 10.1016/j.cub.2007.11.015

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *Am. Econ. Rev.* 90, 426–432. doi: 10.1257/aer.90.2.426

McMillan, C. T., Clark, R., Moore, P., De Vita, C., and Grossman, M. (2005). Neural basis for generalized quantifier comprehension. *Neuropsychologia* 43, 1729–1737. doi: 10.1016/j.neuropsychologia.2005.02.012

McMillan, C. T., Clark, R., Moore, P., and Grossman, M. (2006). Quantifier comprehension in corticobasal degeneration. *Brain Cogn.* 62, 250–260. doi: 10.1016/j.bandc.2006.06.005

Menon, V., Rivera, S. M., White, C. D., Glover, G. H., and Reiss, A. L. (2000). Dissociating prefrontal and parietal cortex activation during arithmetic processing. *Neuroimage* 12, 357–365. doi: 10.1006/nimg.2000.0613

Pinel, P., Dehaene, S., Riviere, D., and Le Bihan, D. (2001). Modulation of parietal activation of semantic distance in a number comparison task. *Neuroimage* 14, 1013–1026. doi: 10.1006/nimg.2001.0913

Sereno, M. I., Pitzalis, S., and Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science* 294, 1350–1354. doi: 10.1126/science.1063695

Shaki, S., and Fischer, M. H. (2013). Your neighbors define your value: a study of spatial bias in number comparison. *Acta. Psychol. (Amst.)* 142, 308–313. doi: 10.1016/j.actpsy.2013.01.004

Simon, O., Mangin, J. F., Cohen, L., Le Bihan, D., and Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron* 33, 475–487. doi: 10.1016/s0896-6273(02)00575-5

Szymanik, J., and Zajenkowski, M. (2010). Comprehension of simple quantifiers: empirical evaluation of a computational model. *Cogn. Sci.* 34, 521–532. doi: 10.1111/j.1551-6709.2009.01078.x

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *J. Econ. Behav. Organ.* 1, 39–60. doi: 10.1016/0167-2681(80)90051-7

Thaler, R. H. (1999). Mental accounting matters. *J. Behav. Decis. Mak.* 12, 183–206. doi: 10.1002/(sici)1099-0771(199909)12:3<183::aid-bdm318>3.0.co;2-f

Vicario, C. M., and Martino, D. (2010). The neurophysiology of magnitude: one example of extraction analogies. *Cogn. Neurosci.* 1, 144–145. doi: 10.1080/17588921003763969

Vicario, C. M. (2012). Perceiving numbers affects the internal random movements generator. *Sci. World J.* 2012:347068. doi: 10.1100/2012/347068

Wu, C. C., Sacchet, M. D., and Knutson, B. (2012). Toward an affective neuroscience account of financial risk taking. *Front. Neurosci.* 6:159. doi: 10.3389/fnins.2012.00159

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 September 2013; paper pending published: 29 October 2013; accepted: 10 January 2014; published online: 28 January 2014.

*Citation: Vicario CM and Rumiati RI (2014) Left-right compatibility in the processing of trading verbs. Front. Behav. Neurosci. **8**:16. doi: 10.3389/fnbeh.2014.00016*

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

frontiers in
BEHAVIORAL NEUROSCIENCE

OPINION ARTICLE
published: 06 June 2013
doi: 10.3389/fnbeh.2013.00057

Why open-access publication should be nonprofit—a view from the field of theoretical language science

Martin Haspelmath*

Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
**Correspondence: haspelmath@eva.mpg.de*

Edited by:
Leonid Perlovsky, Harvard University, USA

Many of my fellow theoretical linguistics researchers have not noticed the momentous changes in the world of science publication yet. When confronted with the idea that publication costs should be covered by author fees ("author processing charges," or APCs), they often react with disbelief and indignation.

But the signs of inefficiency of the old subscription-based system are just as clear in my field as elsewhere, so I see no reasonable alternative to Gold open access (i.e., freely accessible electronic publications on the publisher's website). Green open access is inefficient because of the duplication of efforts, and subscription is inefficient because it is very difficult to predict for an institution to what extent its members will want to use a journal or book. Moreover, the subscription-based model is even worse for scholars with low budgets: While a low-budget scholar can at least read the richer scholars' works on the APC-based open access model, not even that is possible on the traditional model, and usually one can publish in prestigious places only if one knows the relevant literature.

But is APC-based publication of scientific results by profit-oriented companies (such as Macmillan Publishers, which owns Nature Publishing Group, the partner of Frontiers) a good alternative to subscription? Clearly, the old author-pays model removes a major inefficiency of the subscription-based system, because the authors know that they want to publish, whereas the subscribers only suspect that they want to use the publications. According to Stuart Shieber, an open-access expert and theoretical linguist at Harvard University, subscription-based publication can lead to market dysfunction (unreasonably high publication prices) because science journals are not competitive goods: If you subscribe to one science journal, this doesn't mean that you don't need another one (see Shieber, 2013). But from the author's perspective, Shieber says, they are competitive goods: You just need to publish in one journal, and you can choose the cheapest one.

Shieber's article is very sophisticated from an economics perspective, but it completely leaves aside a crucial component of scientific publication that I will argue leads to market dysfunction also with the APC-based open-access model: Scientific publications serve both to disseminate research results and to build careers of scientists. The success of a scientist (and of groups of scientists) is routinely measured by the place of publication of the work. When evaluating a scientist, the evaluators not only look at the amount of research output and the amount of citations, but also at the place of publication. Moreover, when deciding what to cite, scientists routinely privilege papers published in more prestigious journals and books published in more prestigious imprints. Thus, to be a successful scientist, one needs to publish in the same places as other successful scientists. Thus, journals and imprints have a significance for science that goes far beyond the purpose of dissemination of research results. The latter can nowadays be achieved much more easily, by archives such as Arxiv.org, or by publishing in one's personal blog, or on Academia.edu. The primary purpose of peer review is actually peer selection: One needs to make a special effort to present one's results in such a way that one's peers recognize their value. It is only in this way that one's research is likely to have an impact on others. Being selected for publication in a particular place (journal or book imprint) means being successful.

One could imagine alternative models of establishing scientific credentials, e.g., by a rating system similar to the one found in online bookshops, but discussing these is beyond the scope of this note. The big advantage of anonymous peer review and selection that I see for my own field is that it gives younger scholars the chance to become more widely visible even without traveling to many conferences. In the following, I assume that peer selection of publications will be the prevalent mode of establishing scientific credentials also in the future.

Now crucially, the association of place of publication with prestige means that the market for APC-based journals does NOT provide for competition after all: I cannot simply submit my paper to a cheaper journal if the cheaper journal has much less prestige and will lead to much fewer citations of the article. I will quite likely submit my paper to the best journal in my subfield even if this means that I will pay higher APCs (as long as my budget still allows it). Publishers will be able to price their journals according to their prestige, not according to their services. But in the 21st century, the prestige of a journal is primarily the result of the work of the scientists who publish in it, who serve as editors and as reviewers, and not the result of the publisher's efforts. If I publish an excellent piece of research in a journal, or if I write a careful review of a submitted manuscript, I thereby enhance the prestige of this journal, and I thereby contribute to making the journal more expensive for future submitters. The publishers will reap the benefits of my excellent and conscientious work, because they can charge more without improving their services. This situation is clearly undesirable for science.

Journal and book publication has become very simple and cheap as a result of technological developments: One just needs typesetting, hosting and web presentation, as well as perhaps some kind of print-on-demand service (for open-access

books). This can be done very easily without major investments, and as a result, journal publication in the less wealthy countries has increased dramatically over the last 20 years. For example, the Brazilian platform Scielo.org hosts over 1000 journals that are freely accessible and do not charge any author fees.

Of course, even nowadays journal and book publication does not come for free, and somebody has to pay for it. But in order to have a functioning market with reasonable prices, one needs real competition. My research institution can replace its cleaning company by another one, or it can buy its computers and printers from different companies if we are dissatisfied with the services and products. But we cannot simply replace journals and imprints, because we use these to build our careers and to measure our success.

A functioning model would be one where the scientists own the journal titles and book imprints, and where they choose typesetters, webdesign companies, and hosting companies that can be easily replaced by others if the prices are not right. Just as basic science itself is not a profit-oriented activity, publication of scientific results would not be a profit-oriented activity. APCs could be charged by the nonprofit organizations of the scientists (universities, scientific libraries, scholarly associations), but these would not increase as a result of excellent and high-impact work being published by the journals and imprints. On the contrary, since universities and scholarly associations derive their prestige in part from their publications, it is to be expected that the best work will be published without any APCs: These nonprofit organizations would benefit from their prestigious journals and imprints, so it would make sense for them to subsidize them in much the same way as they are subsidizing nonprofit-oriented basic research itself.

The alternative model, where APCs are charged by profit-oriented publishers, has another serious drawback: It creates a strong incentive to create journals and book imprints that function like "vanity presses," allowing authors to publish their low-quality work without significant risk of rejection. Vanity presses have long existed in the regular book market, and they have not been a problem because no public money went into them. Of course, everyone should be free to publish their bad novels or low-quality scientific articles if they desire. However, when it comes to scientific publications, the idea is that the APCs are covered by grants for scientific research, i.e., mostly by public money that would otherwise go into science. In the traditional system, grant holders are free to publish the results of their research wherever they want—but there used to be a limited set of possibilities, and scientific vanity publishers hardly existed. But nowadays increasingly, grant agencies are trying to impose the restriction that the publication should be open access—and with the for-profit approach, there is an unlimited set of possibilities. Anyone can easily found a new journal and offer publication for APCs, simply claiming that it is peer-reviewed. For example, I recently heard of two Chinese companies that are publishing a large number of open-access journals, some of them in my field of linguistics: Wuhan-based SCIRP (http://www.scirp.org/, over 250 journals) and Beijing-based MDPI (http://www.mdpi.com/, over 120 journals). The business model here is to start a large number of new journals and to hope that some of them will succeed and bring profit. For example, MDPI's journal *Languages* does not even have an editor yet. This is of course reminiscent of the business model of spam e-mail, and in fact, some observers have warned of the danger of "predatory journals." In particular, Jeffrey Beal noted in a *Nature* column in 2012 that there are hundreds of journals with this business model, and he writes:

> *The competition for author fees among fraudulent publishers is a serious threat to the future of science communication. To compete in a crowded market, legitimate open-access publishers are being forced to promise shorter submission-to-publication times; this weakens the peer-review process, which takes time to do properly. To tackle the problem, scholars must resist the temptation to publish quickly and easily... (Beall, 2012)*

But the problem with Beall's argumentation is that it is difficult to say in what sense the business model of "predatory" publishers is "fraudulent." They are just exploiting a new niche that has been created by the notion that authors should pay for publication by profit-oriented companies. Clearly, given the current system, where not only quality, but also quantity of publication counts, scholars have an incentive to publish "quickly and easily." Moral exhortations to "resist the temptation" will not make this problem go away.

In order to prevent scholars from publishing their work in less than fully respectable venues, science funders will have to set up a new control system that monitors journal publishers and that prevents grant holders from using grant money to publish in these journals. It is difficult to see how this can be done efficiently and without unduly restricting the freedom of scientists. In any event, it will cost money that would be saved if publication costs were carried by the publishers (universities, libraries, scholarly associations), rather than by the authors.

Another argument that Shieber (2013) cites against toll-access publication is that traditional publishers typically use price bundling, so that canceling individual journal subscriptions does not significantly reduce the costs of the libraries. But is this different in the for-profit open-access model? Not at all: Once open-access publication becomes the norm, for-profit publishers will introduce price bundling for APCs: If your institution enters into an agreement with the publisher, you will pay only EUR 500 for publishing your paper instead of the usual EUR 1000. There are already signs that this is happening: In January 2013, De Gruyter and the Max Planck Society came to an agreement about open-access publication of Max Planck books by De Gruyter (see http://www.mpdl.mpg.de/news/pressrel_2013/PM_deGruyter_MPG_de.pdf).

To summarize, the major argument for open access is that toll access is inefficient because there can be no functioning market (Shieber, 2013) and because it is difficult for subscribers to predict their needs. How should open-access publication be funded? One common funding option is by public funds, i.e., publication is funded in the same way in which science is funded. The other major funding option is by for-profit companies, on the basis of APCs. The major argument against

for-profit companies is again that there can be no functioning market: Scientific publications not only serve to disseminate research findings, but they also build scientific prestige and reputation. Thus, they should be owned by scientists and their institutions, not by companies whose main purpose is to make money. If scientific work is published by for-profit companies, they make money from the reputation that is built up by publicly-funded scientific work. This means that scientific work should be published by nonprofit organizations—those very organizations that are engaged in doing science. This is in fact the traditional model of the 19th century, when it was primarily the scholarly societies and academies that published scientific works. It turns out that this is also the best model for the future.

REFERENCES

Beall, J. (2012). Predatory publishers are corrupting open access. *Nature* 489, 179. doi: 10.1038/489179a

Shieber, S. (2013). *Why Open Access is Better for Scholarly Societies.* Available online at: http://blogs.law.harvard.edu/pamphlet/2013/01/29/why-open-access-is-better-for-scholarly-societies/

Received: 08 May 2013; accepted: 16 May 2013; published online: 06 June 2013.

Citation: Haspelmath M (2013) Why open-access publication should be nonprofit—a view from the field of theoretical language science. Front. Behav. Neurosci. 7:57. doi: 10.3389/fnbeh.2013.00057

frontiers in
BEHAVIORAL NEUROSCIENCE

GENERAL COMMENTARY
published: 02 August 2013
doi: 10.3389/fnbeh.2013.00091

The importance of Open Access publishing in the field of Linguistics for spreading scholarly knowledge and preserving languages diversity in the era of the economic financial crisis

Nicola L. Bragazzi*

Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOGMI), Section of Psychiatry, and School of Public Health, Department of Health Sciences (DISSAL), University of Genoa, Genoa, Italy
**Correspondence: robertobragazzi@gmail.com*

Edited by:
Leonid Perlovsky, Harvard University and Air Force Research Laboratory, USA

A commentary on

Why open-access publication should be nonprofit-a view from the field of theoretical language science
by Haspelmath, M. (2013). Front. Behav. Neurosci. 7:57. doi: 10.3389/fnbeh.2013.00057

I read with great interest the article recently published on "Frontiers in Behavioral Neuroscience," written by Haspelmath and concerning the Open Access (OA) with a particular emphasis from the field of theoretical linguistics (Haspelmath, 2013). However, there are some points in which I disagree with him and I would like to discuss these points as well as I would like to put forth further arguments, adding also an Italian perspective to the topic.

First, the author claims that OA has been marginally noticed in the field of theoretical linguistics, at least in his university context. Moreover, he describes the reactions of "disbelief and indignation" of his colleagues when they hear that they should pay a fee for having their manuscripts published. However, I maintain that this picture for OA journals (OAJs) devoted to linguistics is not exactly true: if we search for linguistics-related OAJs on the Directory of OAJs (DOAJ) database, the largest database collecting OAJs, we discover that only 11 journals have a publishing fee (out of 213, that is to say approximately the 5.2%), while for 4 of them there are no available information (about the 1.9%).

OA seems to be a vivid and expanding reality, and in the field of linguistics it is highly supported by initiatives such as the OA Initiative in Linguistics (OALI), coordinated by the German linguist Stephen Müller (Müller, 2012) and by the same Haspelmath.

Also in Italy, as in Germany, toll-based or subscription-based publishing has revealed its inefficiency: costs that have inflated and increased out of proportion (Giunta, 2010). This kind of publishing is no longer sustainable and affordable. OA seems more cost-effective.

Another point I would like to put forward is that OA could provide an exceptional opportunity and a platform for preserving languages diversity. Article Processing Charges (APCs) for OAJs are (completely or partially) waived for researchers coming from underdeveloped or developing countries, even if as we have already seen that are no publishing fees for the majority of linguistics-related OAJs. OA is growing especially in underdeveloped countries and developing nations (Bayry, 2013), since for them it represents a great opportunity to make their voices heard. The service offered to scholars in these countries is undoubtedly of value: they can freely access scientific content and distribute and communicate with other scholars. Researchers from emerging countries could exploit the benefits of OA for sharing information about languages at risk of extinction in a cost-effective way, uploading also audio-resources and creating a great digital archive. UNESCO has already recognized the importance of the Internet and of other new media to fill the linguistic divide, to preserve multiculturalism and to foster "pluralistic, equitable, open and inclusive knowledge societies" (UNESCO).

In conclusion, the importance of OA in the era of a severe financial and economic crisis is acknowledged. The current crisis is imposing cuts of funding, recruitment/hiring and turnover freezes, and some countries such as Greece have lost access to prestigious journals (Trachana, 2013). OA can overcome the hurdles and the barriers to the spread and the dissemination of the knowledge. OA could provide an excellent platform for delivering and sharing scholarly data and results, contributing at the same time to make the cyberspace a more multilingual reality.

REFERENCES

Bayry, J. (2013). Journals: Open-access boom in developing nations. *Nature* 497, 40. doi: 10.1038/497040

Giunta, C. (2010). *Quanto (ci) Costa la Cultura Accademica?* Available online at: http://www.claudiogiunta.it/wp-content/uploads/2010/02/quanto-ci-costa-leditoria-accademica_.pdf (last accessed on 23 Jun 2013).

Haspelmath, M. (2013). Why open-access publication should be nonprofit-a view from the field of theoretical language science. *Front. Behav. Neurosci.* 7:57. doi: 10.3389/fnbeh.2013.00057

Müller, S. A. (2012). Personal note on Open access in linguistics. *J. Lang. Model.* 0, 9–39.

Trachana, V. (2013). Austerity-led brain drain is killing Greek science. *Nature* 496, 271. doi: 10.1038/496271a

UNESCO. *Linguistic Diversity and Multilingualism on Internet*. Available online at: http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/linguistic-diversity-and-multilingualism-on-internet/ (last accessed on 23 Jun 2013).

Received: 23 June 2013; accepted: 09 July 2013; published online: 02 August 2013.

Citation: Bragazzi NL (2013) The importance of Open Access publishing in the field of Linguistics for spreading scholarly knowledge and preserving languages diversity in the era of the economic financial crisis. Front. Behav. Neurosci. ***7:****91. doi: 10.3389/fnbeh.2013.00091*