# LANGUAGE DEVELOPMENT IN THE DIGITAL AGE

EDITED BY : Mila Vulchanova, Giosuè Baggio, Angelo Cangelosi
and Linda Smith

frontiers Research Topics

## 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 DEVELOPMENT IN THE DIGITAL AGE

Topic Editors:
**Mila Vulchanova,** Norwegian University of Science and Technology, Norway
**Giosuè Baggio**, Norwegian University of Science and Technology, Norway
**Angelo Cangelosi,** Plymouth University, United Kingdom
**Linda Smith,** Indiana University at Bloomington, United States

The digital age is changing our children's lives and childhood dramatically. New technologies transform the way people interact with each other, the way stories are shared and distributed, and the way reality is presented and perceived. Parents experience that toddlers can handle tablets and apps with a level of sophistication the children's grandparents can only envy. The question of how the ecology of the child affects the acquisition of competencies and skills has been approached from different angles in different disciplines. In linguistics, psychology and neuroscience, the central question addressed concerns the specific role of exposure to language. Two influential types of theory have been proposed. On one view the capacity to learn language is hard-wired in the human brain: linguistic input is merely a trigger for language to develop. On an alternative view, language acquisition depends on the linguistic environment of the child, and specifically on language input provided through child-adult communication and interaction. The latter view further specifies that factors in situated interaction are crucial for language learning to take place. In the fields of information technology, artificial intelligence and robotics a current theme is to create robots that develop, as children do, and to establish how embodiment and interaction support language learning in these machines. In the field of human-machine interaction, research is investigating whether using a physical robot, rather than a virtual agent or a computer-based video, has a positive effect on language development.

The Research Topic will address the following issues:

- What are the methodological challenges faced by research on language acquisition in the digital age?
- How should traditional theories and models of language acquisition be revised to account for the multimodal and multichannel nature of language learning in the digital age?
- How should existing and future technologies be developed and transformed so as to be most beneficial for child language learning and cognition?
- Can new technologies be tailored to support child growth, and most importantly, can they be designed in order to enhance specifically vulnerable children's language learning environment and opportunities?
- What kind of learning mechanisms are involved?
- How can artificial intelligence and robotics technologies, as robot tutors, support language development?

These questions and issues can only be addressed by means of an interdisciplinary approach that aims at developing new methods of data collection and analysis in cross-sectional and longitudinal perspectives.

We welcome contributions addressing these questions from an interdisciplinary perspective both theoretically and empirically.

**Citation:** Vulchanova, M., Baggio, G., Cangelosi, A., Smith, L., eds. (2017). Language Development in the Digital Age. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-313-9

# Table of Contents

# Editorial: Language Development in the Digital Age

Mila Vulchanova[1]*, Giosuè Baggio[1], Angelo Cangelosi[2] and Linda Smith[3]

[1] Language Acquisition and Language Processing Lab, Department of Language and Literature, Norwegian University of Science and Technology, Trondheim, Norway, [2] School of Computing and Mathematics, Plymouth University, Plymouth, United Kingdom, [3] Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN, United States

**Editorial on the Research Topic**

**Language Development in the Digital Age**

## INTRODUCTION

The digital age is changing our children's lives and childhood dramatically. New technologies transform the way people interact with each other, the way stories are shared and distributed, and the way reality is presented and perceived. Parents experience that toddlers can handle tablets and apps with a level of sophistication the children's grandparents can only envy. In Great Britain, a recent survey of preschoolers shows that a rising number of toddlers are now put to bed with a tablet instead of a bedtime story. In the USA, a telephone survey of 1,009 parents of children aged 2–24 months (Zimmerman et al., 2007a) documents that by 3 months of age, about 40% of children regularly watched television, DVDs or videos, while by 24 months the proportion rose to 90%. Moreover, with the advance and exponential use of social media, children see their parents constantly interacting with mobile devices, instead of with people around them. Still, research in the US indicates that assistive social robots seem to have a favorable effect on children's language development (Westlund et al.).

Existing theories of language acquisition emphasize the role of language input and the child's interaction with the environment as crucial to language development. From this perspective, we need to ask: What are the consequences of this new digital reality for children's acquisition of the most fundamental of all human skills: language and communication? Are new theories needed that can help us understand how children acquire language? Do the new digital environment and the new ways of interaction change the way languages are learned, or the quality of language acquisition? Is the use of new media beneficial or harmful to children's language and cognitive development? Can new technologies be tailored to support child growth and, most importantly, can they be designed to enhance language learning in vulnerable children?

These questions and issues can only be addressed by means of an interdisciplinary approach that aims at developing new methods of data collection and analysis in a longitudinal perspective. This type of research is however not yet documented.

## Past and Current Research

The question of how the ecology of the child affects the acquisition of competencies and skills has been approached from different perspectives in different disciplines. In linguistics, the central question addressed concerns the specific role of exposure to language. Two influential types of theory have been proposed. One view is that the capacity to learn language is hard-wired in the

human brain (Chomsky, 1965; Pinker, 1994); linguistic input is merely a trigger for language to develop. From an alternative view, language acquisition depends on the linguistic environment of the child, and specifically on language input provided through child-adult communication and interaction (Tomasello, 2003). The latter view further specifies that factors in interaction are crucial for language learning to take place. Such views are aligned with overarching theories of human development in cognitive science and psychology. These theories (known as embodied and situated cognition theories) hold that knowledge is acquired by humans through rich physical and social interaction with their environment (Barsalou, 2008). This interaction leaves multiple traces provided by a number of modalities (auditory, visual, haptic etc.) and helps consolidate knowledge in the brain by strengthening the neural networks that support learning and the use of knowledge. Exactly how input received from multiple, and multi-sensory in nature sources, interacts in both knowledge acquisition and use is, however, still poorly understood.

A current theme in the fields of information technology, artificial intelligence and robotics is to create robots that develop, as children do, and to establish how embodiment and interaction support language learning in these machines. These artificial models will eventually inform us about child development and vice versa (Cangelosi and Schlesinger, 2015, forthcoming). In the field of human-machine interaction, research is investigating whether using a physical robot, rather than a virtual agent or a computer-based video, has a positive effect on language development. Kennedy et al. (2015), for example, investigate how toy-like robots, such as, the Aldebaran Nao, are used in the classroom instead of, or together with, digital tools such as tablets, to show how a richer embodied technology method further improves language learning. Vogt envisage that, in the digital age, social robots will increasingly be used for educational purposes, such as, second language tutoring. They propose a number of design features to develop a child-friendly social robot that can effectively support children in second language learning, and discuss the technical challenges for developing such tutors.

In education research, the main question is the extent to which the use of tablets can facilitate learning to read and write, and how this type of learning compares to traditional learning. In this context, Guerra and Mellado observe that implementing information and communication technologies for educational contexts that have robust and long-lasting effects on student learning outcomes is still a challenge. They further suggest that any such system must be theoretically motivated and designed to tackle specific cognitive skills (e.g., inference making) supporting a given cognitive task (e.g., reading comprehension), and must be able to identify and adapt to the user's profile. Furthermore, a field that combines the concerns of education and digital technology is newly emerging, where one of the questions is how games should be designed to facilitate learning. Zhang et al. provide a review of the educational application of Massive Multiple Online Role-Playing Games (MMORPGs) based on relevant macroscopic and microscopic studies, showing that gamers' overall language proficiency or some specific language skills can be enhanced by real-time online interaction with peers and game narratives or instructions embedded in the MMORPGs. Mechanisms underlying the educational assistant role of MMORPGs in second language learning are discussed from both behavioral and neural perspectives, highlighting the role of attentional bias. Child-media interaction has also been approached in psychology, raising the issue of how new technologies change behavior and interaction, including values and communication patterns.

A recurrent problem in most recent research, however, is that the topic has been approached from a single disciplinary perspective, and often with a single theory in mind. Accounts are piecemeal and explain only one phenomenon at a time. Despite considerable advances in the past 20 years, we miss a holistic model of language development that also integrates the impact of digital technology on its outcomes. Such a model must take into account the weighting of all factors involved. One major challenge is the nature and amount of data that need to be collected and analyzed to build such a model. These data are, in their nature, multi-modal, complex, and dense. It then becomes mandatory to develop new analytic methods and to integrate the complex data needed in order to answer the following three fundamental questions:

- How should traditional theories and models of language acquisition be revised to account for the multimodal and multichannel nature of language learning in the digital age?
- How should existing and future technologies be developed and transformed so as to be most beneficial for child language learning and cognition?
- Can new technologies be tailored to support child growth, and most importantly, can they be designed in order to enhance specifically vulnerable children's language learning environment and opportunities?

# FIRST LANGUAGE DEVELOPMENT

## Early Research on the Mass Media and Language Development

Interest in the impact of the mass media on language development started as early as the late 70-ies. One of the questions that was asked was "Does the language of the mass media contribute a "new" language compared to traditional forms of communication (e.g., books or oral language)?" It was suggested that the new mass media (film, radio, TV) offer "new" languages whose grammar was yet unknown (McLuhan, 1964; Willie, 1979), and, as such, were potentially qualitatively different form oral human-to-human communication. One specific aspect where this difference was particularly salient is the multimodal nature of media, such as television and film. It has been observed that the vehicles of messages in these media involve the marriage of two languages with completely different characteristics (auditory/oral & visual/pictures) (Willie, 1979).

Some results from this early research indicate that there are certain behavioral consequences. For instance, TV-viewing appears to lead to less reading, yet subject to individual variation (Himmelweit et al., 1960). Furthermore, TV-viewing leads to less listening to the radio, and, in particular, with more adverse effect for "brighter" children (greater loss). In contrast, a study on the popular children's programme Sesame Street found a positive

effect of TV viewing on language development, however, only in combination with adult intervention (Winn, 1977). Other research suggests that TV viewing overall has a negative effect on the development of children's attention and cognition and the American Academy of Pediatrics has recommended that children below 2 years of age not watch any television (Anderson and Pempek, 2005).

A valid question if we should expect any impact of mass media on language development is the extent to which the content provided through the media is comprehensible. How much of what children view on TV do they understand? Studies have shown that comprehension tends to increase with age with only 20% understanding among 4-year olds. Also, since this kind of input is mediated through both modalities, the visual and the auditory, advance in language development ought to depend both on the child's non-verbal (visual cognition) and verbal cognitive status at point of exposure. As evidenced by the papers in the current volume, tailoring the features of the technology used to the individual level of cognitive and language skills of the learner is a major prerequisite for successful outcomes. Moreover, as argued by Acerbi, one needs to understand how cultural transmission processes (e.g., transmission biases), of which language learning is arguably one instance, function in the new context of digital media.

When comparing the effects of TV and radio exposure, there is a crucial difference between language experience that requires no reciprocal participation (radio, TV) in contrast to active exchange with another person. Furthermore, TV-images do not go through a complex symbolic transformation; the mind does not decode or manipulate information, as with other types of oral or written language input.

Later research has focused on the extent to which first language acquisition from exposure exclusively to the mass media (radio and TV) deviates from typical language acquisition through interaction with care-givers and peers. Several findings suggest that overwhelming exposure to the kind of input from the radio or TV can have adverse effects, especially for very young children (toddlers). Thus, in a longitudinal study, Zimmerman and Christakis (2005) document that early TV exposure in children younger than 3 years of age was associated with deleterious effects on cognitive development, such as reading at age 7, while infant exposure (between 8 and 16 months) to videos/DVDs was associated with a 16.99-point decrement in CDI score (Zimmerman et al., 2007b). Tanimura et al. (2007) studied 18-month old infants ($n = 1,900$) and found that those who engaged in frequent TV-viewing (>4 h per day), even when accompanied by parental talking, had delayed language development/speech production (in terms of meaningful words). An observational study of 14 pairs of children (age range 7–24 months) and parents videotaped while watching television together shows that both the quality and quantity of parental utterances (Child-directed Speech) significantly declined while the TV was on, and especially when the infants were watching. This also led to an increase of frequency of 1-word sentences, quite often only short phrases, such as nouns (names). From a broader perspective, there is evidence that educational programmes targeting infants and toddlers have not achieved their purported learning goals (cf. Hirsh-Pasek et al., 2015 for a review).

Given that what children watch is important for subsequent vocabulary development (Anderson, 1998; Linebarger and Walker, 2005), and how children watch (with parent or not) is also relevant (Jordan, 2004; Anderson and Pempek, 2005), such findings are extremely pertinent for current research to follow up on. Moreover, the results from the study by Zimmerman et al. (2007b) reporting a negative correlation between DVD viewing and vocabulary development have been challenged by a recent re-analyses of the data set from that study (Ferguson and Donnellan, 2014). This replication found that effect sizes were negligible between analyses for positive, neutral, and negative effects. Interestingly, infants exposed to no media had lower levels of language development compared to infants with some exposure. Thus, it seems that more variables are necessary to take into account in the equation.

## From TV and Radio to Tablets and Robots

Modern digital technology has attracted the attention of scholars due to its favorable affordances. It allows for multi-sensory interaction and provides rich input in the form of visual, auditory, and haptic stimuli (Belpaeme et al., 2012). A recent study by Allen et al. (2015) exploits the multi-modal nature of the input provided by iPads. The main question addressed in that study is whether iPads might promote symbolic understanding and word learning in children with autism in comparison with age-matched typically developing controls. The hypothesis was that multiple, differently colored exemplars of target referents, as afforded by the iPad technology, might promote phonological pattern-meaning/referent associations, e.g., compared to single exemplars. The study included four conditions, contrasting the use of an iPad vs. a Book, and exposing the children to single vs. multiple exemplars of the target items. Participating children were tested on whether they would associate the word to a 3-D referent (real life object) and whether they would generalize it to another member of the same category, but shown in a different color. The results indicated no differences between the two types of media (iPad or book) in symbolic understanding and level of generalization. They further demonstrate that exposure to multiple exemplars increases the rate of extension from picture to 3-D object.

Other studies have focused on how technology can assist exposure to language through reading. Chang and Breazeal (2011) propose to combine a basic primer book with interactivity in order to support parent-child reading interactions during shared book-reading. The design targets very young children (2–5 years) and offers a variety of features: it enables physical proximity, is visually accessible, responds to touch, is navigable to both child and parent, and encourages vocal expression. One specific aspect deserves mention, the Multisensory Contextual Selections. Thus, speech and touch combine to alter the content, and the reader can change story elements using a combination of touch and speech, encouraging creativity and variation. This design is based on interviews and suggestions thereof with educational experts, designers and researchers and exploits the

interactive affordances of digital technology. From the point of view of child-parent interactions, Kucirkova et al. (2014) suggest that multimedia story sharing resembles interactions similar to those when experiencing a piece of art in terms of its holistic nature. Furthermore, there is some evidence that personalization of digital multimedia formats leads to more spontaneous speech production in children (Kucirkova et al., 2014).

## Second Language Learning

Westlund et al. (2016) investigated the role of social robots in second language learning. The study had two main goals. The first one was to test whether a socially assistive robot could help children learn new words in a foreign language more effectively by personalizing its affective feedback. The second aim was to demonstrate the feasibility of creating and deploying a fully autonomous robotic system at a school for several months. The design included a socially assistive robotic learning companion to support English-speaking children's acquisition of a new language (Spanish). In a two-month microgenetic study, 34 preschool children played an interactive game with a fully autonomous robot and the robot's virtual sidekick, a Toucan shown on a tablet screen. Two aspects of the interaction were personalized to each child: (1) the content of the game (i.e., which words were presented), and (2) the robot's affective responses to the child's emotional state and performance. The results from the study indicate that the children learned new words and affective personalization led to greater positive responses from the children.

Vogt et al. propose a number of features for an L2 robot tutor including ways to develop the robot such that it can act as a peer to motivate the child during second language learning and build trust at the same time, while still being more knowledgeable than the child and scaffolding that knowledge in adult-like manner. The authors suggest that the first impression of the child are crucial for building trust and common ground, thus supporting child-robot interactions in the long term. Other important features relate to the ability to adapt to the language proficiency level of the individual child, respond contingently, both temporally and semantically, provide effective feedback and monitor children's learning progress, as well as establish joint attention, and use meaningful gestures. There are a number of technical challenges associated with such an optimal design, such as, automatic speech recognition (ASR) for children, reliable object recognition to facilitate semantic contingency and establishing joint attention, and developing human-like gestures with a robot that does not have the same morphology as humans. The paper presents an experiment which investigates how children respond to different forms of feedback from such a robot.

## CHILD-ROBOT INTERACTION

While we still lack in-depth longitudinal studies of the effects of current digital technologies on language learning, child-robot interaction has been studied recently. Breazeal et al.

(2016) looked at children ranging from 3 to 5 years who were introduced to two anthropomorphic robots that provided them with information about unfamiliar animals. This study found that the children treated the robots as interlocutors: they supplied information to the robots and retained what the robots told them. Children also treated the robots as informants from whom they could seek information. Consistent with children's early sensitivity to an interlocutor's non-verbal signals, children were especially attentive and receptive to whichever robot displayed the greater non-verbal contingency. Selective information seeking is consistent with recent findings showing that although young children learn from others, they are selective with respect to the informants that they question or endorse.

Other research in this domain indicates that children readily treat anthropomorphic robots as social companions (Shiomi et al., 2006). Kahn et al. (2013) document that children often respond verbally to robots (beyond what one might give to an automated system). This research also shows that robots are often attributed mental attributes (emotions etc.), and further that young participants readily engage in verbal exchange with (e.g., speak to) robots.

Movellan et al. (2009) assessed learning from a robot. In that study toddlers (18–24 months) interacted with a sociable robot which displayed images of 4 objects. At pre-test the toddlers' choices were a little better than chance. Over a 2-week period a modest learning outcome was observed, in that there was a significant improvement on taught words, but no improvement on control words. Tanaka and Matsuzoe (2012) studied word learning in the context of a social robot in the age range between 3 and 6 years. The robot responded either correctly of incorrectly to test questions about the novel words. Children reacted and spoke to the robot, and tried to teach the novel words to the robot. Furthermore, they learned the meaning of some novel action words in the company of the robot. However, the results of this study remain unclear as the children's utterances were not analyzed.

All of the studies investigating Child-Robot interaction indicate that the features of the robot are important, and that children differentiate among potential informants. Thus, accent (Kintzler et al., 2011), familiarity (Corriveau and Harris, 2009), turn-taking behavior: contingent responsiveness (Murray and Trevarthen, 1985; Nadel et al., 1999) have all been implicated as central for the interaction and learning outcomes. These findings are consistent with factors in early language development. Thus, contingent responsiveness has been shown to be essential for language learning in infancy (Kuhl, 2007), even though earlier studies have suggested that children acquire native competence regardless of whether spoken to by parents or not. Still, this topic has remained largely out of the focus of current research, and the role of child-directed speech is still to be assessed. Other factors with clear impact on language development are joint attention and accompanying gestures (Tomasello and Farrar, 1986; Tomasello, 2006; Esteve-Gibert et al., 2016). Thus, implementing those features in social robots is likely to have a positive effect on language learning as well.

## INTERIM SUMMARY

The current review has revealed the following findings. Children readily interact with robots. While current research has focused on child-parent interaction while engaging with tablets/iPads, as well as learning in educational contexts, little is known about interaction and language learning from digital devices when the child is the sole agent. The level and quality of interaction largely depends on robot features. As pointed out by Belpaeme et al. (2012), for robots to interact effectively with humans, they need to be capable of coordinated and timely behavior in response to social context. Moreover, they need to display adaptive behavior. Children are likely to interact and engage in verbal exchange (e.g., speak to robots), provided robots feature contingency of responses, provide effective feedback and monitor children's learning progress, as well as establish joint attention, and use meaningful gestures. Yet, very few studies document specific advances in language learning. Thus, so far we see only modest language learning and primarily restricted to vocabulary, but only in experimental settings (Westlund et al.). Nothing is known about "outside of laboratory settings." Overall, there is almost no research on language development *per se*.

In a recent detailed review and discussion of educational apps and their affordances, Hirsh-Pasek et al. (2015) emphasize the role of experience and the environment in the process of acquiring knowledge in early development: whether involving language or not. In particular, the path from sensori-motor experience to symbolic learning, as envisaged in approaches influenced by the Piagetan tradition, appears to be of crucial importance for unpacking the impact of digital technologies on the language learning infant. Similar perspectives need to be in focus when assessing the role of tablets (iPads) in early education (Kucirkova, 2014).

## NEW RESEARCH AGENDA

The study of language learning in rich environments, including digital tools, poses specific challenges to theoretical and empirical research. Traditional theories of language acquisition emphasize characteristics of the learner, such as innate structures and maturational constraints, as well as of the input (its quality, quantity, and variation), but typically they do not take into account the different *channels* through which linguistic and contextual data are provided to the learner. The standard channel is human face-to-face interaction, accompanied by books or printed or recorded material later during childhood. However, the digital age is making new channels available to children earlier on. Each such channel provides input to infants and children through multiple sensory modalities simultaneously—not just hearing, but vision, touch etc. Should empirical research show that vocabulary or grammar learning modes or outcomes vary, depending on the channels through which the linguistic input is provided to the child, theories of language acquisition would have to be expanded, so as to include explicit models of how these effects come about. In particular, *learning theories* (modeling

the input and learner) should be accompanied by *transmission theories* (also modeling the input's sources and transmission channels).

Research on language development in the digital age requires us to understand better the standard modes and channels of language transmission, i.e., vertical social learning. In most modern experiments on (artificial) language learning, the learner is exposed to linguistic or related stimuli that are "produced" by machines, e.g., a computer, not by other human beings. Implicitly, much research on language learning involving exposure or training phases is already research on learning from digital tools. There is research on language learning and use in social contexts (Tomasello, 2003), however these two lines of work have not yet been integrated: what is needed are experiments in which learning from others and learning from digital tools are directly compared, i.e., where the learning channel is an explicit experimental factor. This approach may help understanding the cognitive and behavioral consequences of learning in digital ecologies, while keeping other factors under experimental control. For example, one could directly test whether digital tools are simply increasing the *amount of information* that is made available to children, or whether instead they are facilitating or impeding learning (e.g., of new vocabulary) when information quantity is held constant. The same mutatis mutandis would hold for information quality and variation. A further set of questions is whether the effects of digital tools on learning are short-lived or long-lasting, and whether they manifest themselves invariably or only early during development: would the child's brain eventually adapt to the multiplicity of channels and respective modalities through which language is experienced? Longitudinal designs are necessary to answer such questions.

The development of robot tutors to support early language development, as well as L2 language acquisition, offers innovative ways of exploiting the digital age technologies for language tutoring purposes, and in general, for child-robot interaction. Research has consistently demonstrated that the physical presence (embodiment) of a robot (e.g., Kennedy et al., 2015; Cangelosi and Schlesinger, forthcoming), as well as some of its anthropomorphic features (robot appearance with human-like shape; e.g., Walters et al., 2008) and behavior (shared gaze, gestures; e.g., Zanatto et al., 2016), improve the outcome of the tutoring and companionship objectives. Moreover, multimodal approaches to human-robot interaction, such as, those combining tablet-based interfaces with the robot's speech communication capabilities and behavioral feedback strategies, improve the acceptability and efficacy of robot companions (Belpaeme et al., 2012; Di Nuovo et al., 2016). As such, future research directions in robot tutors for language development will benefit from the investigation of hybrid robot and digital technologies, strategically exploiting the benefits from the robot's anthropomorphic features.

Robot companions also offer the opportunity to support language acquisition in children with atypical development. Pioneering studies have looked at social assistive robotics for children with autism spectrum disorder (ASD) (e.g.,

Dautenhahn, 1999; Scassellati et al., 2012). For example, Scassellati et al. (2012) suggest that the improvement of social skills development via robot interaction is the consequence of the fact that robots provide novel sensory stimuli to the ASD child. Robot companions have also been used for the support of children with diabetes (Belpaeme et al., 2012) and with mobility and motor disabilities (Sarabia and Demiris, 2013). Thus, future work combining robot tutors with populations with atypical cognitive and motor development will contribute to the challenges of language skills acquisition in children with disabilities.

Future research should harvest evidence of language development in interaction with digital tools (including social robots). It should compare children who are often exposed to ICT to children who are not. It should investigate how new media/digital tools impact on the development of lower level language skills (e.g., vocabulary, grammar); how new media/digital tools impact on the development of "higher" skills (e.g., discourse comprehension) and explore the development of dimensionality (Language and Reading Research Consortium, 2015), and specifically, the effect of digital technology on oral and reading comprehension, and figurative language skills. A broader and overarching issue is the effect of new digital environments on brain plasticity and learning (Bavelier et al., 2010). Future research on this topic is also in need of novel methods for data analyses.

## AUTHOR CONTRIBUTIONS

MV prepared the initial draft and worked on subsequent editing. GB and AC worked on specific sections. LS completed final editing and alignment of ideas.

## REFERENCES

Allen, M. L., Hartley, C., and Cain, K. (2015). Do iPads promote symbolic understanding and word learning in children with autism? *Front. Psychol.* 6:138. doi: 10.3389/fpsyg.2015.00138

Anderson, D. R. (1998). Educational television is not an oxymoron. *Ann. Am. Acad. Pol. Soc. Sci.* 557, 24–38. doi: 10.1177/0002716298557000003

Anderson, D. R., and Pempek, T. A. (2005). Television and very young children. *Am. Behav. Sci.* 48, 505–522. doi: 10.1177/0002764204271506

Barsalou, L. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Bavelier, D., Green, C. S., and Dye, M. W. (2010). Children, wired – for better and for worse. *Neuron* 67, 692–701. doi: 10.1016/j.neuron.2010.08.035

Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., et al. (2012). Multimodal child-robot interaction: building social bonds. *J. Hum. Robot Interact.* 1, 33–53. doi: 10.5898/JHRI.1.2.Belpaeme

Breazeal, C., Harris, P. L., DeSteno, D., Kory Westlund J. M., Dickens, L., and Jeong, S. (2016). Young children treat robots as informants. *Top. Cogn. Sci.* 8, 1–11. doi: 10.1111/tops.12192

Cangelosi, A., and Schlesinger, M. (2015). *Developmental Robotics: From Babies to Robots.* Cambridge, MA: MIT Press.

Cangelosi, A., and Schlesinger, M. (forthcoming). From babies to robots: the developmental robotics contribution to child psychology. *Child Dev. Perspect.*

Chang, A., and Breazeal, C. (2011). "TinkRBook: shared reading interfaces for storytelling," in *Proceedings of the 10th International Conference on Interaction Design and Children* (Ann Arbor), 145–148.

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

Corriveau, K., and Harris, P. L. (2009), Choosing your informant: weighing familiarity and recent accuracy. *Dev. Sci.* 12, 426–437. doi: 10.1111/j.1467-7687.2008.00792.x

Dautenhahn, K. (1999). "Robots as Social Actors: Aurora and the Case of Autism," in *Proc. CT99, The Third International Cognitive Technology Conference, August,* (San Francisco, CA).

Di Nuovo, A., Wang, N., Broz, F., Belpaeme, T., Jones, R., and Cangelosi, A. (2016). "Experimental evaluation of a multi-modal user interface for a robotic service," in *Towards Autonomous Robotics Systems: 17th Annual Conference, TAROS 2016, Proceedings (LNAI9716),* eds L. Alboul, D. Damian, and J. M. Aitken (Sheffield: Springer), 87–98.

Esteve-Gibert, N., Prieto, P., and Liszkowski, U. (2016). Twelve-month-olds understand social intentions based on prosody and gesture shape. *Infancy* 22, 108–129. doi: 10.1111/infa.12146

Ferguson, C. J., and Donnellan, M. B. (2014). Is the association between children's baby video viewing and poor language development robust? A reanalysis of Zimmerman, Christakis, and Meltzoff (2007). *Dev. Psychol.* 50, 129–137. doi: 10.1037/a0033628

Himmelweit, H., Oppenheim, A., and Vince, P. (1960). *Television and the Child: An Empirical Study of the Effect of Television on the Young.* Oxford: Oxford University Press.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., and Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychol. Sci. Public Interest* 16, 3–34. doi: 10.1177/1529100615569721

Jordan, A. (2004). The role of media in children's development: an ecological perspective. *J. Dev. Behav. Pediatr.* 25, 196–206. doi: 10.1097/00004703-200406000-00009

Kahn, P. H., Gary, H. E., and Shen, S. (2013). Children's social relationships with current and near and future robots. *Child Dev. Perspect.* 7, 32–37. doi: 10.1111/cdep.12011

Kennedy, J., Baxter, P., and Belpaeme, T. (2015). Comparing robot embodiments in a guided discovery learning interaction with children. *Int. J. Soc. Robot.* 7, 293–308. doi: 10.1007/s12369-014-0277-4

Kintzler, K. D., Corriveau, K. L., and Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Dev. Sci.* 14, 106–111. doi: 10.1111/j.1467-7687.2010.00965.x

Kucirkova, N. (2014). iPads in early education: separating assumptions and evidence. *Front. Psychol.* 5:715. doi: 10.3389/fpsyg.2014.00715

Kucirkova N., Messer D., Sheehy K., and Fernández Panadero C. (2014). Children's engagement with educational iPad apps: insights from a Spanish classroom. *Comput. Educ.* 71, 175–184. doi: 10.1016/j.compedu.2013.10.003

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Dev. Sci.* 10, 110–120. doi: 10.1111/j.1467-7687.2007.00572.x

Language and Reading Research Consortium (LARRC) (2015). The dimensionality of language ability in young children. *Child Dev.* 86, 1948–1965. doi: 10.1111/cdev.12450

Linebarger, D. L., and Walker, D. (2005). Infants' and toddlers' television viewing and language outcomes. *Am. Behav. Sci.* 48, 624–645. doi: 10.1177/0002764204271505

McLuhan, M. (1964). *Understanding Media: The Extensions of Man, 1st Edn.* New York, NY: McGraw Hill.

Movellan, J., Eckhardt, M., Virnes, M., and Rodriguez, A. (2009). "Sociable robot improves toddler vocabulary skills," in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction (ACM),* (La Jolla, CA), 307–308.

Murray, L., and Trevarthen, C. (1985). "Emotional regulation of interactions between two-month- olds and their mothers," in *Social Perception in Infants,* eds T. Field and N. Fox, (Norwood, NJ: Ablex Publishers. Na), 177–197.

Nadel, J., Carchon, I., Kervella, C., Marcelli, D., and Reserbat-PIantey, D. (1999). Expectations for social contingency in 2-month-olds. *Dev. Sci.* 2, 164–173. doi: 10.1111/1467-7687.00065

Pinker, S. (1994). *The Language Instinct*. Harper, CA: William Morrow and Company.

Sarabia, M., and Demiris, Y. (2013). "A humanoid robot companion for wheelchair users," in *Social Robotics, Lecture Notes in Computer Science*, eds G. Herrmann, M. J. Pearson, A. Lenz, P. Bremner, A. Spiers, U. Leonards (Berlin: Springer), 432–441.

Scassellati, B. Admoni, H., and Matarić, M. (2012). Robots for use in autism research. *Annu. Rev. Biomed. Eng.* 14, 275–294. doi: 10.1146/annurev-bioeng-071811-150036

Shiomi, M., Kanda, T., Ishiguro, H., and Hagita, N. (2006). "Interactive Humanoid Robots for a Science Museum," in *Paper Presented at the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, (New York, NY).

Tanaka F., and Matsuzoe S. (2012). Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *J. Hum. Robot Interact.* 1, 78–95. doi: 10.5898/JHRI.1.1.Tanaka

Tanimura, M., Okuma, K., and Kyishima, K. (2007). Television viewing, reduced parental utterance and delayed speech development in infants and young children. *Arch. Pediatr. Adolesc. Med.* 161, 618–619. doi: 10.1001/archpedi.161.6.618-b

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2006). "Acquiring linguistic constructions," in *Handbook of Child Psychology*, eds D. Kuhn, and R. Siegler (New York, NY: Wiley).

Tomasello, M., and Farrar, M. J. (1986). Joint attention and early language. *Child Dev.* 57, 1454–1463.

Walters, M. L., Syrdal, D. S., Dautenhahn, K., Te Boekhorst, R., and Koay, K. L. (2008). Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Auton. Robots* 24, 159–178. doi: 10.1007/s10514-007-9058-3

Westlund, K., Lee, J. J., Plummer, L., Faridi, F., Gray, J., Berlin, M., et al. (2016). "Tega: a social robot," in *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, (Cambridge, MA), 561.

Willie, M. (1979). The mass media and language development. *Austral. J. Teacher Educ.* 4, 58–64. doi: 10.14221/ajte.1979v4n2.5

Winn, M. (1977). *The Plug-in Drug*. New York, NY: Viking Press.

Zanatto, D., Goslin, J., Patacchiola, M., and Cangelosi, A. (2016). "Priming anthropomorphism: can our trust in humanlike robots be transferred to non-humanlike robots?," in *Proceeding of the Eleventh ACM/IEEE International Conference on Human Robot Interation* (Christchurch), 543–544.

Zimmerman, F. J., and Christakis, D. A. (2005). Children's television viewing and cognitive outcomes: a longitudinal analysis of national data. *Arch. Pediatr. Adolesc. Med.* 159, 619–625. doi: 10.1001/archpedi.159.7.619

Zimmerman, F. J., Christakis, D. A., and Meltzoff, A. N. (2007a). Television and DVD/Video viewing in children younger than 2 years. *Archiv. Pediatr. Adolesc. Med.* 161, 473–479. doi: 10.1001/archpedi.161.5.473

Zimmerman, F. J., Christakis, D. A., and Meltzoff, A. N. (2007b). Associations between media viewing and language development in children under age 2 years. *J. Pediatr.* 151, 364–368. doi: 10.1016/j.jpeds.2007.04.071

Check for updates

# Flat vs. Expressive Storytelling: Young Children's Learning and Retention of a Social Robot's Narrative

Jacqueline M. Kory Westlund[1]*, Sooyeon Jeong[1], Hae W. Park[1], Samuel Ronfard[2], Aradhana Adhikari[1], Paul L. Harris[2], David DeSteno[3] and Cynthia L. Breazeal[1]

[1]MIT Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA, United States, [2]Harvard Graduate School of Education, Harvard University, Cambridge, MA, United States, [3]Department of Psychology, Northeastern University, Boston, MA, United States

Prior research with preschool children has established that dialogic or active book reading is an effective method for expanding young children's vocabulary. In this exploratory study, we asked whether similar benefits are observed when a robot engages in dialogic reading with preschoolers. Given the established effectiveness of active reading, we also asked whether this effectiveness was critically dependent on the expressive characteristics of the robot. For approximately half the children, the robot's active reading was expressive; the robot's voice included a wide range of intonation and emotion (*Expressive*). For the remaining children, the robot read and conversed with a flat voice, which sounded similar to a classic text-to-speech engine and had little dynamic range (*Flat*). The robot's movements were kept constant across conditions. We performed a verification study using Amazon Mechanical Turk (AMT) to confirm that the *Expressive* robot was viewed as significantly more expressive, more emotional, and less passive than the *Flat* robot. We invited 45 preschoolers with an average age of 5 years who were either English Language Learners (ELL), bilingual, or native English speakers to engage in the reading task with the robot. The robot narrated a story from a picture book, using active reading techniques and including a set of target vocabulary words in the narration. Children were post-tested on the vocabulary words and were also asked to retell the story to a puppet. A subset of 34 children performed a second story retelling 4–6 weeks later. Children reported liking and learning from the robot a similar amount in the *Expressive* and *Flat* conditions. However, as compared to children in the *Flat* condition, children in the *Expressive* condition were more concentrated and engaged as indexed by their facial expressions; they emulated the robot's story more in their story retells; and they told longer stories during their delayed retelling. Furthermore, children who responded to the robot's active reading questions were more likely to correctly identify the target vocabulary words in the *Expressive* condition than in the *Flat* condition. Taken together, these results suggest that children may benefit more from the expressive robot than from the flat robot.

**Keywords: preschool children, emotion, expressiveness, language development, peer modeling, social robotics, storytelling**

# INTRODUCTION

Prior research with preschool children has established that storytelling and story reading can promote oral language development and story comprehension (Isbell et al., 2004; Speaker et al., 2004; Cremin et al., 2016). Participating in storytelling can increase children's verbal fluency, listening skills and vocabulary. Book reading in particular can be an effective method for expanding young children's vocabulary, especially when children are encouraged to actively process the story materials. For example, in an intervention study, middle class parents assigned to an experimental group were instructed to engage in "dialogic" reading with their 2-year-old, i.e., to ask more open-ended and function/attribute questions and to support the efforts of their children to answer these questions; parents in the control group were instructed to read in their usual fashion. In follow-up tests, children in the experimental group scored higher in assessments of expressive vocabulary (Whitehurst et al., 1988). Subsequent studies have replicated and extended this result (e.g., Valdez-Menchaca and Whitehurst, 1992; Hargrave and Sénéchal, 2000; Chang et al., 2012; Nuñez, 2015; Boteanu et al., 2016). Taken together, these studies indicate that dialogic book reading is an effective method for boosting children's vocabulary. Indeed, the studies confirm that such an intervention is quite robust in its effects—it is effective for toddlers as well as preschoolers, for middle class and working class children and for typically developing as well as language-delayed children, when using print or digital storybooks.

In this exploratory study, we asked whether similar benefits could be observed when a social robot engages in dialogic story reading with preschoolers. Social robots share physical spaces with humans and leverage human means of communicating—such as speech, movement and nonverbal cues, including gaze, gestures, and facial expressions—in order to interface with us in more natural ways (Breazeal, 2004; Breazeal et al., 2008; Feil-Seifer and Mataric, 2011). Given our expectation that children would learn from the robot, we also investigated how the emotional expressiveness of the robot's speech might modulate children's learning.

A growing body of research suggests that social robots have potential as learning companions and tutors for young children's early language education. For example, robots have played simple vocabulary games to help children learn new words in their own language or in a second language (Kanda et al., 2004; Movellan et al., 2009; Chang et al., 2010; Tanaka and Matsuzoe, 2012; Gordon et al., 2016; Kennedy et al., 2016). It is plausible that children's successful vocabulary learning in these experiments depended on their relating to the robots as interactive, social beings (Kahn et al., 2013; Breazeal et al., 2016a; Kennedy et al., 2017). Social cues impact children's willingness to engage with and learn from interlocutors (Bloom, 2000; Harris, 2007; Corriveau et al., 2009; Meltzoff et al., 2009; Sage and Baldwin, 2010). Indeed, Kuhl (2007, 2011) has argued that a lack of social interaction with a partner can impair language learning. Thus, infants learn to differentiate new phonemes presented

by a live person, but do not learn this information from a video of a person, or from mere audio. Because robots are seen by children as social agents—a peer, a tutor, or a companion—they seem to be providing the necessary social presence to engage children in a language learning task. Thus, social robots, unlike educational television programs (Naigles and Mayeux, 2001), may allow children to acquire more complex language skills and not just vocabulary. However, existing studies on robots as language learning companions have generally not assessed this possibility. Nearly all of the activities performed with social robots around language learning have been simple, vocabulary-learning tasks, with limited interactivity. For example, the robot might act out new verbs (Tanaka and Matsuzoe, 2012), show flashcard-style questions on a screen (Movellan et al., 2009), or play simple give-and-take games with physical objects (Movellan et al., 2009; see also Gordon et al., 2016).

A few studies have explored other kinds of activities for language learning. For example, Chang et al. (2010) had their robot read stories aloud, ask and answer simple questions, and lead students in reciting vocabulary and sentences. However, they primarily assessed children's engagement with the robot, rather than their language learning. One study used a story-based task in which the robot took turns telling stories with preschool and kindergarten children, for 8 weeks (Kory, 2014; Kory and Breazeal, 2014; Kory Westlund and Breazeal, 2015). In each session, the robot would tell two stories with key vocabulary words embedded, and would ask children to make up their own stories for practice. For half the children, the robot personalized the level of the stories to the child's ability, telling more complex stories for children who had greater ability. This study found increases in vocabulary learning as well as in several metrics assessing the complexity of the stories that children produced, with greater increases when children heard appropriately leveled stories. These findings suggest that a social robot is especially likely to influence language learning if it conveys personal attunement to the child. Indeed, children were more trusting of novel information provided by a social robot whose nonverbal expressiveness was contingent on their behavior (Breazeal et al., 2016a) and showed better recall of a story when the social robot teaching them produced high immediacy gestures in response to drops in children's attention (Szafir and Mutlu, 2012).

In the current study, we focus on a related but hitherto unexplored factor: the emotional expressiveness of the robot's speech. Nearly every study conducted so far on the use of social robots as learning companions for young children has used a computer-generated text-to-speech voice, rather than a more natural, human voice. We know very little about the effects of a more expressive, human-like voice as compared to a less expressive, flatter or synthetic voice on children's learning. Such expressive qualities may have an especially strong impact during storytelling activities. For example, if a potentially engaging story is read with a flat delivery, children might find it anomalous or even aversive. Using robots to study questions about expressivity is quite feasible, because we can carefully control the level of vocal expressiveness across conditions and

between participants. Robots afford a level of control that it is difficult to achieve with human actors with the same consistency.

A small number of human-robot interaction (HRI) studies have investigated the effects of a robot's voice on an interaction. However, these studies tested adults (e.g., Eyssel et al., 2012), compared different synthetic voices (e.g., Walters et al., 2008; Tamagawa et al., 2011; Sandygulova and O'Hare, 2015), or compared qualities of the same voice, such as pitch (e.g., Niculescu et al., 2013; Lubold et al., 2016), rather than varying the expressiveness of a given voice. Eyssel et al. (2012) did compare human voices to synthetic voices, but the adult participants merely watched a short video clip of the robot speaking, and did not interact with it directly. These participants perceived the robot more positively when the voice shared their gender, and anthropomorphized the robot more when the voice was human.

Some related work in speech-language pathology and education has compared children's learning from speakers with normal human voices or voices with a vocal impairment, specifically, dysphonic voices. Children ages 8–11 years performed better on language comprehension measures after hearing passages read by a normal human voice than when the passages were read by a dysphonic voice (Morton et al., 2001; Rogerson and Dodd, 2005; Lyberg-Åhlander et al., 2015). These studies suggest that vocal impairment can be detrimental to children's speech processing, and may force children to allocate processing to the voice signal at the expense of comprehension. However, it is unclear whether a lack of expressivity or the use of a synthetic voice would impair processing relative to a normal human voice.

Given the lack of research in this area, we compared the effect of an expressive as compared to a flat delivery by a social robot. We also focused on a more diverse population as compared to much prior work with regard to both age and language proficiency. Previous studies have tended to focus on just one population of children—either native speakers of the language, or children learning a second language—whereas we included both. In addition, few previous studies have included preschool children (Movellan et al., 2009; Tanaka and Matsuzoe, 2012; Kory, 2014); the majority of studies have targeted older children. More generally, young children comprise an age group that is typically less studied in HRI (Baxter et al., 2016).

In this study, we invited preschoolers with an average age of 5 years and considerable variation in language proficiency to engage in a dialogic reading task with a social robot. Thus, some children were English Language Learners (ELL), some were bilingual, and some were monolingual, native English speakers. All children were introduced to a robot who first engaged them in a brief conversation and then proceeded to narrate a story from a picture book using dialogic reading techniques. Two versions of the study were created; each version contained a unique set of three novel words. In post-story testing, children's comprehension of the novel words they had heard was compared to their comprehension of the novel words embedded in the story version they had not heard. We predicted that children would

display superior comprehension of the novel words that they had heard.

Given the established effectiveness of dialogic reading with young children, the robot always asked dialogic questions. We asked two related questions: first, we asked whether children would learn from a dialogic storytelling robot. Second, we asked whether its effectiveness was critically dependent on the expressiveness of the robot's voice—how might the robot's vocal expressivity impact children's engagement and learning? For approximately half the children, the robot's dialogic reading was expressive in the sense that the robot's voice included a wide range of intonation and emotion. For the remaining children, the robot read and conversed with a flat voice, which sounded similar to a classic text-to-speech engine and had little dynamic range. To control for the many differences that computer-generated voices have from human voices (e.g., pronunciation and quality), an actress recorded both voices, and we performed a manipulation check to ensure the expressive recording was perceived to be sufficiently more emotional and expressive than the flat recording. We anticipated that children would be more attentive, show greater gain in vocabulary, and use more of the target vocabulary words themselves if the dialogic reading was delivered by the expressive as compared to the flat robot. To further assess the potentially distinct impact of the two robots, children were also invited to retell the picture-book story that the robot had narrated. More specifically, they were invited to retell the story to a puppet who had allegedly fallen asleep during the robot's narration and was disappointed at having missed the story. Finally, a subset of children was given a second opportunity to retell the story approximately 4–6 weeks later.

## MATERIALS AND METHODS

### Design

The experiment was designed to include two between-subjects conditions: Robot expressiveness (*Expressive* voice vs. *Flat* voice) and Robot redirection behaviors (*Present* vs. *Absent*). Regarding the robot's voice, the expressive voice included a wide range of intonation and emotion, whereas the flat voice sounded similar to a classic text-to-speech engine with little dynamic range. The robot redirection behaviors were a set of re-engagement phrases that the robot could employ to redirect a distracted child's attention back to the task at hand. However, the conditions under which the robot would use redirection behaviors did not arise—i.e., all the children were attentive and the opportunity to redirect their attention did not occur. Thus, the experiment ultimately had a two-condition, between-subjects design (*Expressive* vs. *Flat*).

### Participants

This study was carried out in accordance with the recommendations of the MIT Committee on the Use of Humans as Experimental Subjects. Children's parents gave written informed consent prior to the start of the study and all children assented to participate, in accordance with the

Declaration of Helsinki. The protocol was approved by the MIT Committee on the Use of Humans as Experimental Subjects and by the Boston Public Schools Office of Data and Accountability.

We recruited 50 children aged 4–7 (23 female, 27 male) from a Boston-area school (36 children) and the general Boston area (14 children) to participate in the study. Five children were removed from the analysis because they did not complete the study. The children in the final sample included 45 children (22 female, 23 male; 34 from the school and 11 from the general Boston area) with a mean age of 5.2 years ($SD = 0.77$). Seventeen children were ELL, eight were bilingual, 18 were native English speakers, and three were not reported.

Children were randomly assigned to conditions. There were 23 children (14 male, 9 female; 10 ELL, 6 Native English, 5 bilingual, 2 unknown) in the *Expressive* condition and 22 children (9 male, 13 female; 6 ELL, 12 Native English, 3 bilingual, 1 unknown) in the *Flat* condition. The two conditions were not perfectly balanced due to the fact that we did not obtain information about children's language learning status until the completion of the study, and thus could not assign children evenly between conditions.

We created two versions of the story that the robot told (version A and version B); each version was identical except for the inclusion of a different set of target vocabulary words. Approximately half of the participants heard story version A (*Expressive*: 11, *Flat*: 10); the other half heard story version B (*Expressive*: 13, *Flat*: 11).

We used the Peabody Picture Vocabulary Test, 4th edition (PPVT; Dunn and Dunn, 2007), to verify that the children in the *Expressive* and *Flat* conditions did not have significantly different language abilities. The PPVT is commonly used to measure receptive language ability for standard American English. On each test item, the child is shown a page with four pictures, and is asked to point to the picture showing the target word. PPVT scores for three of the 45 children could not be computed due to missing data regarding their ages. For the remaining 42 children, there were, as expected, no significant differences between the *Expressive* and *Flat* conditions in PPVT scores, $t_{(40)} = 0.64$, $p = 0.53$. A one-way analysis of variance (ANOVA) with age as a covariate revealed that children's PPVT scores were significantly related to their age, $F_{(3,37)} = 5.83$, $p = 0.021$, $\eta^2 = 0.114$, as well as to their language status, $F_{(3,37)} = 2.72$, $p = 0.058$, $\eta^2 = 0.160$. As expected, *post hoc* pairwise comparisons indicated that children who were native English speakers had higher PPVT scores ($M = 109.4$, $SD = 18.2$) than ELL children ($M = 92.0$, $SD = 14.6$), $p = 0.004$. There were no differences between the bilingual children ($M = 103.5$, $SD = 15.7$) and either the native English-speaking children or the ELL children.

## Hypotheses

The effects of the robot's expressivity might be transient or long-term, subtle or wide-ranging. Accordingly, we used a variety of measures, including immediate assessments as well as the delayed retelling task, to explore whether the effect of the robot's expressivity was immediate and stable and whether it impacted all measures, or selected measures only.

We tentatively expected the following results:

### Learning
- In both conditions, children would learn the target vocabulary words presented in the story version that they heard.
- Children who learned the target words would also use them in their story retells.
- The robot's expressivity would lead to differences in children's long-term retention of the story. Children in the *Expressive* condition would better retain the story, and thus tell longer stories, incorporating the phrases that appeared in the initial story into their delayed retells.

### Behavior
- In both conditions, children would typically respond to the robot's dialogic reading questions, but children who responded more often to the dialogic reading questions would show greater learning gains.
- The *Expressive* robot would promote greater modeling by the children of the robot's story. Children in the *Expressive* condition would produce more vocabulary and phrase mirroring.

### Engagement
- Although most children would express liking for the robot, indirect behavioral measures would show that children were more attentive and engaged with the *Expressive* robot than with the *Flat* robot.
- The surprising moments in the story would have greater impact on children in the *Expressive* condition, because suspense and surprise in the story were strongly reflected in the robot's voice.

## Procedure

Each child was greeted by an experimenter and led into the study area. The experimenter wore a hand puppet, a purple Toucan, which she introduced to the child: "This is my friend, Toucan." Then the puppet spoke: "Hi, I'm Toucan!" The experimenter used the puppet to invite the child to do a standard vocabulary test, the PPVT, by saying "I love word games. Want to play a word game with me?" The experimenter then administered the PPVT.

For the children who participated in the study at their school, the PPVT was administered during an initial session. The children were brought back on a different day for the robot interaction. This second session began with the puppet asking children if they remembered it: "Remember me? I'm Toucan!" Children who participated in the lab first completed the PPVT, and were then given a 5-min break before returning to interact with the robot.

For the robot interaction, the experimenter led the child into the robot area. The robot sat on a low table facing a chair, in which children were directed to sit. A tablet was positioned in an upright position in a tablet stand on the robot's right side. A smartphone sat in front of the robot; it ran software to track children's emotional expressions (see **Figure 1**). The experimenter sat to the side and slightly behind the children with the puppet. The interaction began with the puppet introducing
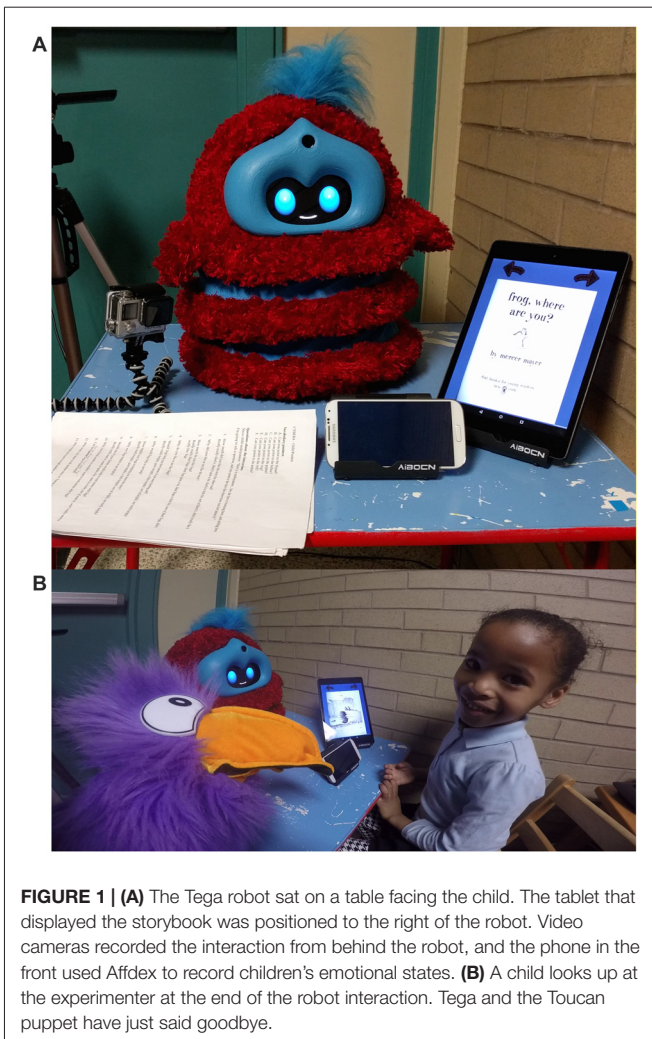
**FIGURE 1 | (A)** The Tega robot sat on a table facing the child. The tablet that displayed the storybook was positioned to the right of the robot. Video cameras recorded the interaction from behind the robot, and the phone in the front used Affdex to record children's emotional states. **(B)** A child looks up at the experimenter at the end of the robot interaction. Tega and the Toucan puppet have just said goodbye.

the robot, Tega: "This is my friend, Tega!" The robot introduced itself, shared personal information, and prompted children to do the same, e.g., "Hi, I'm Tega! My favorite color is blue. What is your favorite color?" and "Do you like to dance? I like to dance!"

After this brief introductory conversation, the robot asked the children if they wanted to hear a story. At this point, the puppet interjected that it was sleepy, but would try to stay awake for the story. The experimenter made the puppet yawn and fall asleep; it stayed asleep for the duration of the story. The robot then told the story which consisted of a 22-page subset of the wordless picture book "Frog, Where Are you?" by Mercer Mayer. This book has been used before in numerous studies, especially in research on speech pathology (e.g., Boudreau and Hedberg, 1999; Greenhalgh and Strong, 2001; Diehl et al., 2006; Heilmann et al., 2010).

The pages of the book were shown one at a time on the tablet screen. Each page was accompanied by 1–2 sentences of text, which the robot read in either an expressive or a flat voice depending on the condition. For every other page, the robot asked a dialogic reading comprehension question about

the events in the story, e.g., "What is the frog doing?", "Why did the boy and the dog fall?", and "How do you think the boy feels now?" (11 questions total). The robot responded to children's answers with encouraging, but non-committal, phrases such as "Mmhm", "Good thought" and "You may be right".

We embedded three target vocabulary words (all nouns) into the story. We did not test children on their knowledge of these words prior to the storytelling activity because we did not want to prime children to pay attention to these words, since that could bias our results regarding whether or not children would learn or use the words after hearing them in the context of the robot's story. Instead, in order to assess whether children were more likely to know or use the words after hearing the robot use them in the story, two versions of the story (version A and version B) were created with different sets of target words. The two versions of the story were otherwise identical. We identified six key nouns in the original story: animal, rock, log, hole, deer and hill. Then, in each of our two story versions, we replaced three of the words with our target words, so that each story version included three target words and three original words. Version A included the target words "gopher" (original word: animal), "crag" (rock), and "lilypad" (log); version B included the words "hollow" (hole), "antlers" (deer), and "cliff" (hill). We anticipated that children would display selective learning and/or use of these words, depending on which story they heard. We looked both at children's later receptive knowledge of the words as well as expressive or productive abilities, since children who can recognize a word may or may not be able to produce it themselves.

At the end of the story, the Toucan woke up and exclaimed, "Oh no! Did I miss the story?" This presented an opportunity for children to retell the story to the puppet, thereby providing a measure of their story recall. Children were allowed to go through the story on the tablet during their retelling. Thus, the depictions on each page could serve as a reminder during retelling.

After the story-retelling task, the experimenter administered a PPVT-style vocabulary test for the six target words used across the two versions of the story. For each word, four pictures taken from the story's illustrations were shown to children and they were asked to point to the picture matching the target word. Finally, the experimenter asked children a set of questions regarding their perception of the robot and their enjoyment of the story. These questions were as follows:

1. How much did you like the story the robot read? Really really liked it, liked it quite a lot, liked it a little bit, sort of liked it, didn't really like it.
2. Why did you like or not like the story?
3. How much do you like Tega? Really really liked Tega, liked Tega quite a lot, liked Tega a little bit, sort of liked Tega, didn't really like Tega.
4. Why do you like or not like Tega?
5. Would Tega help you feel better if you were feeling sad? Really really helpful, quite helpful, a little helpful, sort of helpful, not really helpful.

6. Why would Tega help or not help?

7. How helpful was Tega in helping you learn the story? Really really helpful, quite helpful, a little helpful, sort of helpful, not really helpful.

8. Why was Tega helpful or not helpful?

9. Would one of your friends would want to read stories with Tega? Really really want to, want to quite a lot, want to a little, sort of want to, won't really want to.

10. Why your friend would or wouldn't want to read stories with Tega?

11. Can you describe Tega to your friend?

12. Who would you want to tell another story to: Toucan or Tega?

13. Why would you want to read another story to: Toucan or Tega?

Where appropriate, we used a Smiley-o-meter to gather responses on a 1–5 scale (Read and MacFarlane, 2006). Although Read and MacFarlane (2006) suggest that this measure is not useful with children younger than 10 years, previous research has successfully used it, or similar measures, with modest pre-training (Harris et al., 1985; Leite et al., 2014). Thus, we did a practice item before the test questions so children could learn how the measure worked. Children were also asked to explain their answers, such as "Why do you like or not like Tega?" and "Why was Tega helpful or not helpful?" Children's parents or teachers provided demographic data regarding the children's age and language status (ELL, bilingual, or native English speaker).

A subset of 34 children from the school sample participated in a second, follow-up session approximately 4–6 weeks later at their school. Children who participated in the lab did not have a follow-up session due to logistical reasons. During this follow-up session, we administered the PPVT a second time, then asked children to retell the story to the puppet. The puppet prompted children by saying, "I tried to tell the story to my friend last week, but I forgot most of it! Can you tell it to me again?" This allowed us to observe children's long-term memory for the story.

Four different experimenters (three female adults and one male adult) ran the study in pairs. One experimenter interacted with the child. The other experimenter acted as the robot teleoperator and equipment manager; she could be seen by the children, but she did not interact directly with them.

## Materials
We used the Tega robot, a squash and stretch robot designed for educational activities with young children (Kory Westlund et al., 2016). The robot is shown in **Figure 1**. It uses an Android phone to run its control software as well as display an animated face. The face has two blue oval eyes and a white mouth, which can all morph into different shapes. This allows the face to show different facial expressions and to show appropriate visemes (i.e., mouth shapes) when speech is played back. The robot can move up and down, tilt its head sideways or forward/backward, twist to the side, and lean forward or backward. Some animations played on the robot use only the face; others incorporate both facial expressions and physical movements of the body. The robot is covered in red fur with blue stripes, giving it

a whimsical, friendly appearance. The robot was referred to in a non-gendered way by the experimenters throughout the study.

A female adult recorded the robot's speech. These utterances were shifted into a higher pitch to make them sound child-like. For the *Expressive* condition, the utterances were emotive with a larger dynamic range; the actress was instructed to speak in an expressive, human-like way. For the *Flat* condition, the actress imitated a computer-generated text-to-speech voice, keeping her intonation very flat. We did not use an actual computer-generated voice for the *Flat* voice because there would have been many differences in pronunciation and quality compared to the *Expressive* voice. Similarly, we did not use a computer-generated voice for the *Expressive* voice because no computer-generated voices can currently imitate the dynamic, expressive range that human voices are capable of.

Many of the physical actions the robot can perform are expressive. We used the same physical movements in both conditions; however, in the *Expressive* condition, some movements were accompanied by expressive sounds (such as "Mm hm!"), whereas in the *Flat* condition, these movements were either accompanied by a flat sound ("Mm hm.") or, in cases where the sound was a short, non-linguistic expressive utterance, no sound.

We used a Google Nexus 9 8.9″ tablet to display the storybook. Touchscreen tablets have been shown to effectively engage children and social robots in a shared task (Park et al., 2014). We used custom software to display the story pages that allowed a teleoperator to control when the pages were turned; this software is open-source and available online under the MIT License at https://github.com/mitmedialab/SAR-opal-base/.

We used a Samsung Galaxy S4 android smartphone to run Affdex, which is emotion measurement software from Affectiva, Inc., Boston, MA, USA[1]. Affdex performs automatic facial coding in four steps: face and facial landmark detection, face feature extraction, facial action, and emotion expression modeling based on the EMFACS emotional facial action coding system (Ekman and Friesen, 1978; Friesen and Ekman, 1983; McDuff et al., 2016). Although no data has been published yet specifically comparing the performance of the software on adults vs. children, FACS coding is generally the same for adults and for children and has been used with children as young as 2 years (e.g., Camras et al., 2006; LoBue and Thrasher, 2015; also see Ekman and Rosenberg, 1997). Furthermore, this software has been trained and tested on tens of thousands of manually coded images of faces from around the world (McDuff et al., 2013, 2015; Senechal et al., 2015).

## Teleoperation
We used a custom teleoperation interface to control the robot and the digital storybook. Using teleoperation allowed the robot to appear autonomous to participants while removing technical barriers such as natural language understanding, because the teleoperator could be in the loop as the language parser. The teleoperator used the interface to trigger when the robot should

---

[1]http://affectiva.com/ (retrieved September 19, 2016).

begin its next sequence of actions (a list of speech, physical motions, and gaze) and also when the storybook should proceed to the next page. Thus, the teleoperator needed to pay attention to timing in order to trigger the robot's next action sequence at the appropriate times relative to when the experimenter spoke (i.e., when introducing the robot to the child), or when the child responded to one of the robot's questions. Since the teleoperator did not manage the timing of actions within each sequence, the robot's behavior was highly consistent for all children.

The four experimenters were all trained to control the robot by an expert teleoperator; they had all controlled robots before in multiple prior studies.

## Manipulation Check

To check that the *Expressive* robot was, in fact, perceived to be more expressive than the *Flat* robot, we performed a verification study using Amazon Mechanical Turk (AMT). We recorded video of the robot performing all the speech and behavior used in the main study. We then selected samples of the robot's speech and behavior from the introductory conversation, the beginning, middle, and end of the story, and the closing of the interaction to create a video clip that was approximately two and a half minutes in length. We created one video of the *Flat* robot and one video of the *Expressive* robot. In the two videos, we used the same speech and behavior samples such that the only difference was the expressiveness of the robot's voice.

We recruited 40 AMT workers from the United States. Half the participants (11 male, 9 female) viewed the video of the *Flat* robot and half (13 male, 7 female) viewed the video of the *Expressive* robot. After viewing the video, participants were asked to rate their impression of the robot and report demographic information. We used the following questions, each of which was measured on a 1–5 Likert-type scale anchored with "*1: Not___at all*" and "*5: Extremely___*":

1. Overall, how expressive or not expressive was the robot in the video?
2. Overall, how emotional or not emotional was the robot in the video?
3. Overall, how passive or not passive was the robot in the video?
4. How expressive or not expressive was the robot's <u>voice</u> in the video?
5. How emotional or not emotional was the robot's <u>voice</u> in the video?
6. How passive or not passive was the robot's <u>voice</u> in the video?
7. How expressive or not expressive was the robot's <u>movement</u> in the video?
8. How emotional or not emotional was the robot's <u>movement</u> in the video?
9. How passive or not passive was the robot's <u>movement</u> in the video?

**Table 1** shows a summary of participant responses. We found that participants who watched the *Expressive* robot rated the robot as significantly more emotional overall than participants who watched the *Flat* robot video, $t_{(39)} = 2.39$, $p = 0.022$. Participants who watched the *Expressive* robot

video rated the robot's voice as significantly more expressive, $t_{(39)} = 4.44$, $p < 0.001$; more emotional, $t_{(39)} = 5.15$, $p < 0.001$; and less passive, $t_{(39)} = 2.96$, $p = 0.005$, than participants who watched the *Flat* robot video. There were no statistically significant differences in participants' ratings of the robot's movement.

The results demonstrate that the *Expressive* and *Flat* robot conditions were indeed sufficiently different from each other, with the voice of the *Expressive* robot being viewed as more expressive, more emotional, and less passive than the *Flat* robot.

## Data

We recorded video and audio data for each session using two different cameras set up on tripods behind the robot, facing the child. We recorded children's facial expressions using Affdex, emotion measurement software from Affectiva, Inc., Boston, MA, USA. Children's responses to the PPVT, target word vocabulary test, and interview questions were recorded on article during the experiment and later transferred to a spreadsheet.

## Data Analysis

We coded whether or not children responded to each of the questions the robot asked during the initial conversation and during the story, and if they did respond, how many words their response consisted of. We also counted the number of questions that children asked the puppet when retelling the story.

To assess how children perceived Tega as a function of their assignment to the *Expressive* and *Flat* conditions, we coded children's responses to the open-ended question inviting them to describe Tega to a friend (i.e., "Can you describe Tega to your friend?") for positive traits (e.g., nice, helpful, smart, fun). All children provided a response to this question. Children's responses to the Smiley-o-meter questions were coded on a 1–5 scale.

Children's transcribed story retells were analyzed in terms of their story length, overall word usage and target word usage, and phrase similarity compared to the robot's original story. Automatic tools were developed such that each word was converted into its original form for comparison (stemming), words with no significant information (i.e., stopwords) were removed, and an N-gram algorithm was implemented to match phrases between the child's and the robot's stories. N-gram refers to a contiguous sequence of N items from a given sequence of text. In our analysis, we used $N = 3$ for matching and comparison. We chose $N = 3$ because a smaller N (e.g., $N = 2$) often retains too little information to constitute actual phrase matching, and a larger N may encompass more information than would constitute a single phrase. For example, the robot's story included the section, "The frog jumped out of an open window. When the boy and the dog woke up the next morning, they saw that the jar was empty". After stemming and stopword removal, this section would be converted to "frog jump open window boy dog wake next morning see jar empty". One child retold this section of the story by saying "Frog was going to jump out the window. So whe... then the boy and the dog woke up, the jar was empty". This was converted to "frog jump window boy dog wake jar empty". The N-gram phrase matching for this

**TABLE 1 |** Summary of participant responses for the *Expressive* vs. *Flat* robot verification study.

| Question | Condition | Mean | Median | Mode | Range | Inter-quartile range |
|---|---|---|---|---|---|---|
| Overall expressive | *Flat* | 3.75 | 4 | 4 | 1–5 | 1 |
| | *Expressive* | 3.60 | 4 | 4 | 2–5 | 1 |
| Overall emotional | *Flat* | 2.90 | 3 | 2 | 1–5 | 2 |
| | *Expressive* | 3.60 | 4 | 4 | 2–5 | 1 |
| Overall passive | *Flat* | 3.15 | 3 | 3 | 1–5 | 2 |
| | *Expressive* | 2.56 | 3 | 3 | 1–5 | 1 |
| Expressive voice | *Flat* | 2.65 | 3 | 1 | 1–5 | 2.25 |
| | *Expressive* | 4.05 | 4 | 4 | 3–5 | 0 |
| Emotional voice | *Flat* | 2.15 | 2 | 1 | 1–5 | 2 |
| | *Expressive* | 3.85 | 4 | 4 | 3–5 | 1 |
| Passive voice | *Flat* | 3.45 | 3.5 | 3 | 1–5 | 2 |
| | *Expressive* | 2.30 | 2 | 2 | 1–5 | 1 |

*The Expressive robot was viewed as more expressive, more emotional, and less passive than the Flat robot.*

segment reveals multiple phrase matches, e.g., (robot) "*frog jump open window*"/(child) "*frog jump window*", and (robot) "*boy dog wake* next morning see *jar empty*"/(child) "*boy dog wake jar empty*".

Children's affect data were collected using Affdex whenever a face was detected with the front-facing camera on the Samsung Galaxy S4 device (McDuff et al., 2016). Affdex is capable of measuring 15 expressions, which are used to calculate the likelihood that the detected face is displaying each of nine different affective states. We analyzed the four affective states most relevant to our research questions: attention, concentration, surprise and engagement. *Attention* is a measure of focus based on head orientation—i.e., is the child attending to the task or not. The likelihood of *concentration* is increased by *brow furrow* and *smirk*, and decreased by *smile*. Thus, *concentration* reflects the effort and affective states associated with attending, rather than merely whether the child is looking in the correct direction or not. *Surprise* is increased by *inner brow raise*, *brow raise* and *mouth open*, and decreased by *brow furrow*. *Engagement* measures facial muscle activation reflective of the subject's expressiveness, and is calculated as a weighted sum of the *brow raise*, *brow lower*, *nose wrinkle*, *lip corner depressor*, *chin raise*, *lip pucker*, *lip press*, *lips part*, *lip suck* and *smile*. Thus, the *Engagement* score reflects total facial muscle activation during the task. On every video frame (up to 32 frames per second), each of these affective states was scored by Affdex in the range 0 (no expression present) to 100 (expression fully present). Values in the middle (e.g., 43 or 59) indicate that the expression is somewhat present; these values are relative and Affdex does not indicate what the exact difference is between each score. See Senechal et al. (2015) for more detail regarding the algorithms uses for classification.

For the story retelling, the audio quality of 40 out of 45 participants was sufficiently good enough for transcription (22 female, 18 male; age $M = 5.2$, $SD = 0.76$; 14 ELL, 7 bilingual, 16 native English, 3 unknown). There were 21 children (10 female, 11 male; age $M = 5.3$, $SD = 0.80$; 9 ELL, 4 bilingual, 6 native English, 2 unknown) in the *Expressive* condition and 19 children (12 female, 7 male; age $M = 5.1$, $SD = 0.71$; 5 ELL, 3 bilingual, 10 native English, 1 unknown) in the *Flat* condition. Half of the participants had heard story version A (*Expressive*: 10,

*Flat*: 10); the other half had heard story version B (*Expressive*: 11, *Flat*: 9).

To perform analyses across the two sessions, immediate and delayed retell pairs from 29 children were used (14 female, 15 male; age $M = 5.2$, $SD = 0.68$; 14 ELL, 3 bilingual, 12 native English). There were 15 children (6 female, 9 male; age $M = 5.3$, $SD = 0.70$; 9 ELL, 2 bilingual, 4 native English) from the *Expressive* condition and 14 children (8 female, 6 male; age $M = 5.1$, $SD = 0.66$; 5 ELL, 1 bilingual, 8 native English) from the *Flat* condition. Half of the participants heard story version A (*Expressive*: 8, *Flat*: 8); the other half heard story version B (*Expressive*: 7, *Flat*: 6).

In the following analyses, we ran Shapiro-Wilk (S-W) tests to check for normality and Levene's test to check for equal variance, where applicable. Levene's null hypothesis was rejected for all data in our dataset ($p > 0.05$) and constant variance was assumed across conditions and sessions. Parametric (paired/unpaired $t$-test) and non-parametric (Wilcoxon signed-rank and Mann-Whitney's U) tests were used based on the S-W result.

## RESULTS

We present our results in three parts, with each part addressing one of our three main hypotheses: (1) *Learning*: our primary question was whether children would learn from a robot that led a dialogic storytelling activity, and specifically whether the expressiveness of the robot's voice would impact children's learning; (2) *Behavior*: we asked whether children would learn more if they responded to the dialogic reading questions, and whether the robot's expressiveness would produce greater lexical and phrase modeling; and (3) *Engagement*: we asked whether the robot's expressiveness would lead to greater attention or engagement. Finally, we also examined whether children's learning was impacted by their language status.

### Learning

#### Target Vocabulary Word Identification
Overall, children correctly identified a mean of 4.0 of the six target vocabulary words ($SD = 1.38$). A $2 \times 2 \times 2$ mixed ANOVA with condition (*Expressive* vs. *Flat*), the story children

| Age | Number of children | Mean target words correct *(SD)* |
|---|---|---|
| 4 years | 9 | 3.22 *(1.30)* |
| 5 years | 21 | 4.14 *(1.28)* |
| 6 years | 14 | 4.21 *(1.53)* |
| 7 years | 1 | 5.00 *(N/A)* |

heard (version A vs. version B), and the words correctly identified (number of version A words correct vs. number of version B words correct, where children were asked to identify both sets of words), with age as a covariate, revealed a trend toward age affecting how many words children identified correctly, $F_{(1,81)} = 3.40$, $p = 0.069$, $\eta^2 = 0.045$. *Post hoc* pairwise comparisons showed that older children identified more target words correctly, with 4-year-olds identifying fewer words than 5-year-olds ($p = 0.016$), 6-year-olds ($p = 0.016$), and the 7-year-old ($p = 0.077$; **Table 2**). There was no difference between the total number of target vocabulary words that children identified correctly in the *Expressive* ($M = 3.8$ correct of 6, $SD = 1.48$) vs. *Flat* ($M = 4.23$, $SD = 1.27$) conditions.

We also found the expected interaction between story version heard and number of words correctly identified from each version (**Figure 2**). Children who heard story version A were likely to correctly identify more version A words ($M = 2.00$ correct of 3, $SD = 0.853$) than version B words ($M = 1.62$, $SD = 0.813$), whereas children who heard story version B were more likely to correctly identify more version B words ($M = 2.21$ correct of 3, $SD = 1.03$) than version A words ($M = 1.92$, $SD = 0.626$), $F_{(1,81)} = 4.21$, $p = 0.043$.

In summary, performance in the vocabulary test improved with age. Nevertheless, there was evidence of learning from the story in that children performed better on those items they had encountered in the story version they heard.
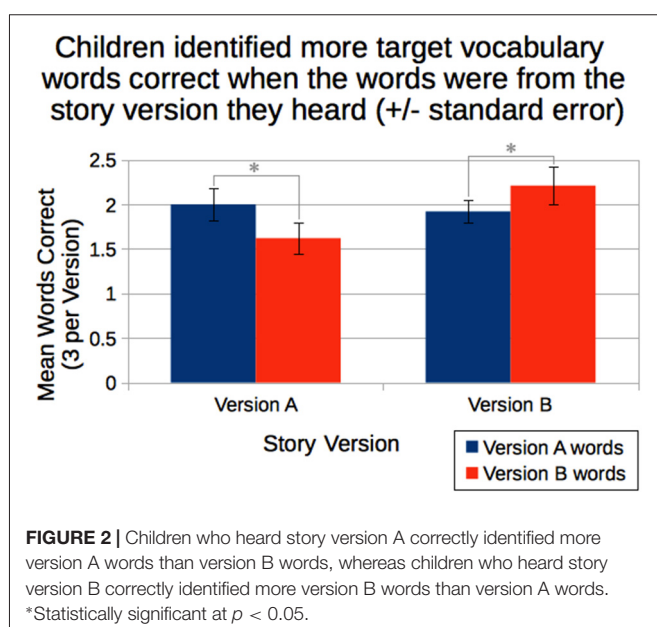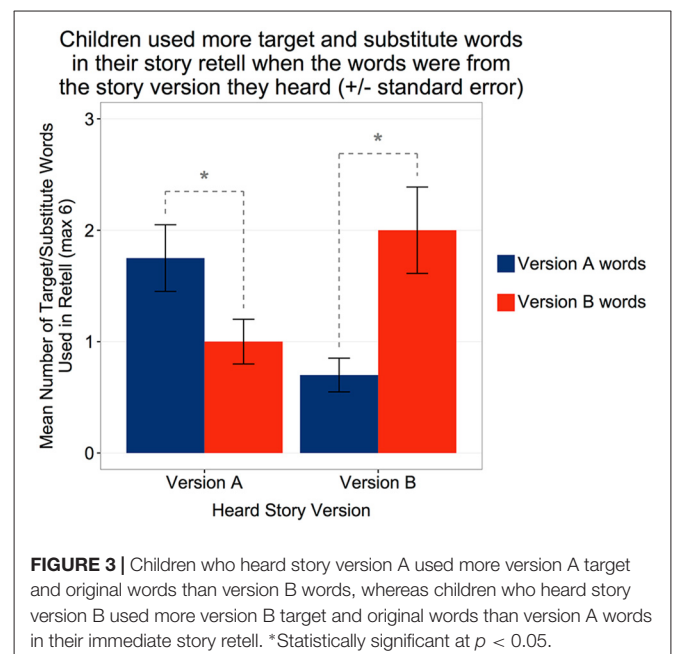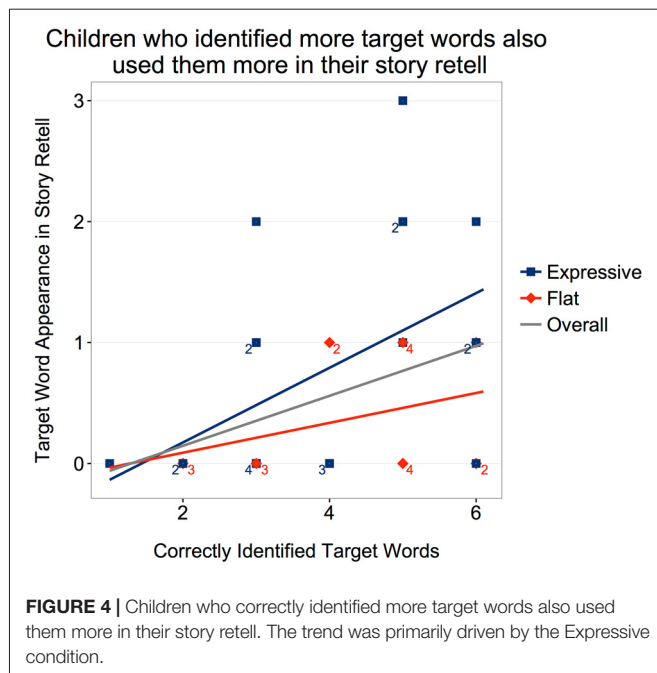
## Target Word Use

First, because the two story versions (A and B) differed both in terms of the target words included and the original words (i.e., the lower level words that we replaced with the target words), we analyzed how often children used either type of word. This was to provide context in terms of children's overall word reuse rates after hearing the words in the robot's story. Thus, among 40 children, 35 children either used the target words or the original words in their story retelling ($M = 2.15$ out of 6, $SD = 1.48$). As in the target word identification, we also found significant differences in children's word usage behavior based on the story version they heard. A Wilcoxon signed-rank test revealed that children who heard story version A were more likely to use version A words in their story retelling ($M = 1.75$, $SD = 1.37$) than version B words ($M = 1.00$, $SD = 0.920$); $W = 12$, $Z = -2.34$, $p = 0.019$, $r = 0.52$, whereas children who heard story version B were more likely to use version B words ($M = 2.00$, $SD = 1.69$) than version A words ($M = 0.700$, $SD = 0.660$); $W = 12.5$, $Z = -2.87$, $p = 0.004$, $r = 0.64$ (**Figure 3**).

Then, to analyze children's learning of new words from the robot, we focused on children's reuse of the target words. There was no significant difference in overall target word usage between the *Flat* and *Expressive* conditions. In the immediate retell, children used a mean of 0.45 target words (out of 3), $SD = 0.69$. However, out of the 17 children who used at least one of the target words in their retell (*Expressive*: 10 children, *Flat*: 7), children in the *Expressive* condition used significantly more target words ($M = 1.6$, $SD = 0.70$) than children in the *Flat* condition ($M = 1.00$, $SD = 0.00$), $t_{(15)} = 2.248$, $p = 0.040$.

A trend toward older children using more target words than younger children was also observed; age 4 ($M = 0.14$, $SD = 0.38$),



**FIGURE 2 |** Children who heard story version A correctly identified more version A words than version B words, whereas children who heard story version B correctly identified more version B words than version A words. *Statistically significant at $p < 0.05$.



**FIGURE 3 |** Children who heard story version A used more version A target and original words than version B words, whereas children who heard story version B used more version B target and original words than version A words in their immediate story retell. *Statistically significant at $p < 0.05$.

**FIGURE 4 |** Children who correctly identified more target words also used them more in their story retell. The trend was primarily driven by the Expressive condition.



**FIGURE 5 |** Children's story-retell length significantly reduced after 1–2 months in the Flat condition, but not in the Expressive condition. *Statistically significant at $p < 0.05$.

age 5 ($M = 0.58$, $SD = 0.77$), age 6 ($M = 0.62$, $SD = 0.65$), age 7 ($M = 3.0$, $SD = 0.00$); Kendall's rank correlation $\tau_{(38)} = 0.274$, $p = 0.059$. In the delayed retell, time was significant ($M = 0.21$, $SD = 0.49$; $W = 10$, $Z = -2.77$, $p = 0.05$). The correlation between the number of target words that children used in the immediate retell and their score on the target-word test was significant, $\tau_{(38)} = 0.348$, $p = 0.011$. This trend was significant in the *Expressive* condition ($\tau_{(19)} = 0.406$, $p = 0.031$), but not in the *Flat* condition ($\tau_{(17)} = 0.246$, $p = 0.251$; **Figure 4**).

In summary, children tended to use more of the target words encountered in the story version they heard, and older children tended to use more of the target words.

### Story Length

The length of the story told by the robot was 365 words. In the immediate retell, the mean length of children's stories was 200.7 words ($SD = 80.8$). No statistically significant difference in story length was observed between the two conditions (*Expressive*: $M = 191.8$ words, $SD = 82.5$, *Flat*: $M = 210.6$, $SD = 79.9$), $t_{(38)} = -0.73$, $p = 0.47$. Story length also did not vary with age, Pearson's $r_{(7)} = 0.06$, $p = 0.71$.

A $2 \times 2$ mixed ANOVA with time (within: Immediate vs. Delayed) and condition (between: *Expressive* vs. *Flat*) for the subset of children who produced both immediate and delayed retells revealed significant main effects of time, $F_{(1,27)} = 17.9$, $p < 0.001$, $\eta^2 = 0.398$, as well as a significant interaction between time and condition, $F_{(1,27)} = 15.0$, $p < 0.001$, $\eta^2 = 0.357$. In the delayed retell, the overall length of children's story decreased to $M = 147.9$ ($SD = 58.3$; $t_{(13)} = 5.35$, $p < 0.001$). Children in the *Flat* condition showed a significant decrease (Immediate: $M = 210.9$, $SD = 85.4$, Delayed: $M = 125.4$, $SD = 57.2$), while in the *Expressive* condition, the decrease was not statistically significant (Immediate: $M = 173.3$, $SD = 79.33$, Delayed: $M = 168.9$, $SD = 52.8$; $t_{(14)} = 0.33$, $p = 0.75$). Furthermore, the length
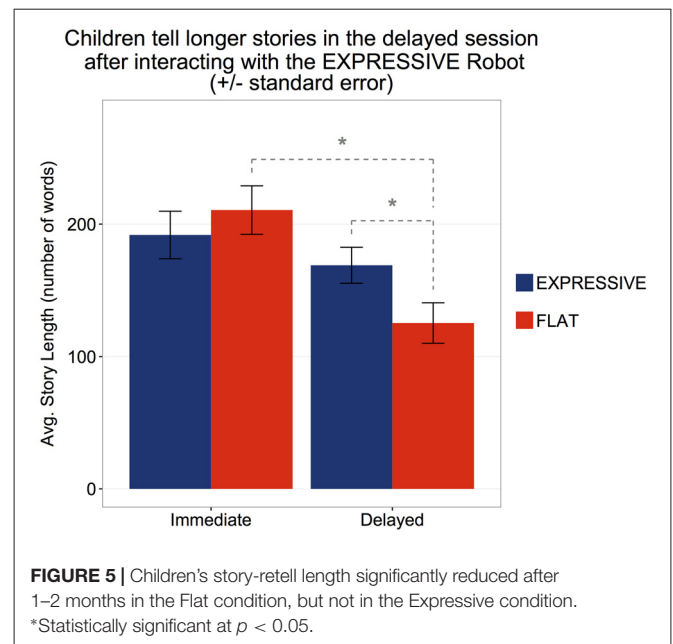
of stories in the two conditions were significantly different at the delayed retelling (*Expressive*: $M = 168.9$, $SD = 52.8$, *Flat*: $M = 125.4$, $SD = 57.2$), $t_{(27)} = 2.13$, $p = 0.043$.

Thus, children in the *Flat* condition told shorter stories at the delayed retell as compared to the immediate retell whereas no such reduction was seen among children in the *Expressive* condition. Their stories were just as lengthy after 1–2 months (**Figure 5**). To further understand the impact of expressivity on retelling, we analyzed children's phrase production as reported in the following section.

## Behavior

### Responses to the Robot's Dialogic Questions

Forty-two children had data regarding their responses to the robot-posed dialogic reading questions. Thirty-five (83.3%) responded to at least some of the questions; 23 (54%) responded to all 11 questions; seven (16.7%) responded to none. There was no significant difference between the number of questions responded to by children in the *Expressive* and *Flat* conditions.

A simple linear regression model revealed that children who had responded to the robot's dialogic questions were likely to correctly identify more of the target vocabulary words, $F_{(1,38)} = 5.84$, $p = 0.021$, $\eta^2 = 0.118$. The interaction between the condition and the number of questions responded showed a trend, $F_{(1,38)} = 4.094$, $p = 0.0501$, $\eta^2 = 0.083$, such that question answering in the *Expressive* condition was related to correct identification of target words, while question answering was not related to correct identification of words in the *Flat* condition. The correlation was driven primarily by the *Expressive* condition, $r_{(20)} = 0.619$, $p = 0.002$, i.e., children in the *Expressive* condition who answered the robot's questions were more likely to identify more of the target words; there was no significant correlation for the *Flat* condition, $r_{(18)} = 0.134$, $p = 0.57$ (**Figure 6**). Thus, answering the dialogic questions was linked to
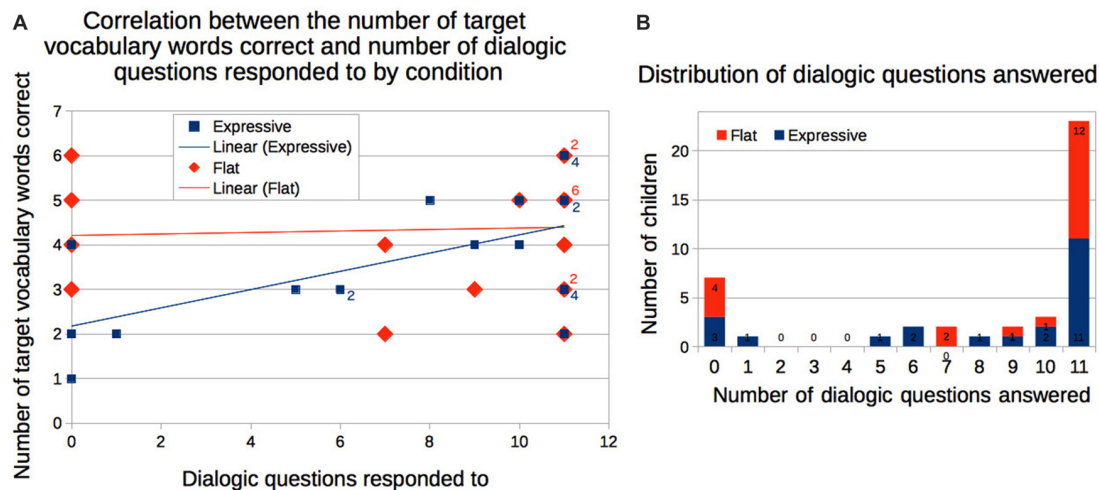
**FIGURE 6 | (A)** Children who responded to the robot's dialogic questions were also more likely to correctly identify more of the target vocabulary words. The correlation was primarily driven by children in the Expressive condition. **(B)** The majority of children responded to most or all of the robot's dialogic questions.

better vocabulary learning, but this link was only found in the *Expressive* condition.

Children who answered more dialogic questions also used significantly more target words in the immediate story retell as indicated by a Spearman's rank-order correlation $r_{s(38)} = 0.352$, $p = 0.026$. These children also told longer stories, $r_{s(38)} = 0.447$, $p = 0.003$ (**Figure 7A**). They displayed greater emulation of the robot in terms of phrase usage, $r_{s(38)} = 0.320$, $p = 0.044$, but again this was driven primarily by the *Expressive* condition, $r_{s(19)} = 0.437$, $p = 0.048$, and not by the *Flat* condition, $r_{s(17)} = 0.274$, $p = 0.257$. Children in the *Expressive* condition also showed significant correlation to phrase usage in the delayed retell, $r_{s(13)} = 0.554$, $p = 0.032$ (**Figure 7B**).

From the above observations, we can conclude that children were, in general, actively engaged in the robot's storytelling. When children were more engaged, as indexed by how often they responded to the robot's questions, their vocabulary learning was greater, and they were more likely to emulate the robot. However, these links between engagement and learning were evident in the *Expressive* rather than the *Flat* condition.

### Emulating the Robot's Story

An analysis of children's overall word usage reveals their word-level mirroring of the robot's story. In total, the robot used 96 unique words after stopword removal and the calculation of non-overlapping words. In the immediate retell, children used a mean of 58.7 words (*SD* = 12.4) emulating the robot. There was no significant difference between conditions. In the delayed retell, however, children in the *Expressive* condition used more words emulating the robot than children in the *Flat* condition (*Expressive*: *M* = 48.6, *SD* = 13.5, *Flat*: *M* = 38.7, *SD* = 8.62; $t_{(27)} = 2.33$, $p = 0.028$).

We also analyzed the phrase-level similarity between the robot's story and the children's stories. In the immediate retell, a mean of 5.63 phrases (*SD* = 3.55) were matched.

A statistically significant difference was observed between conditions (*Expressive*: *M* = 6.67, *SD* = 3.98, *Flat*: *M* = 4.47, *SD* = 2.65; $t_{(38)} = 2.03$, $p = 0.049$) with robot's expressivity increasing children's phrase-level similarity. In the delayed retell, the overall usage of matched phrases decreased (*M* = 3.34, *SD* = 2.26), $t_{(28)} = 5.87$, $p < 0.001$. However, a Mann-Whitney U test showed that participants in the *Expressive* condition (*M* = 4.20, *SD* = 2.40) continued to use more similar phrases than participants in the *Flat* condition (*M* = 2.42, *SD* = 1.74), *Z* = 2.07, $p = 0.039$, $r = 0.38$ (**Figure 8**). Thus, at both retellings, children were more likely to echo the expressive than the flat robot in terms of their phrasing.

The overall correlation between children's score on the target-word vocabulary test and the number of matched phrases they used in the retell was significant both for the immediate retell, $r_{s(38)} = 0.375$, $p = 0.017$; and for the delayed retell, $r_{s(27)} = 0.397$, $p = 0.033$ (**Figure 9**). However, further analysis showed that this link was significant in the *Expressive* condition (Immediate: $r_{s(19)} = 0.497$, $p = 0.022$; Delayed: $r_{s(13)} = 0.482$, $p = 0.031$), but not in the *Flat* condition (Immediate: $r_{s(19)} = 0.317$, $p = 0.186$; Delayed: $r_{s(12)} = 0.519$, $p = 0.067$).

In summary, children were more likely to use similar words and phrases as the robot in the *Expressive* than in the *Flat* condition during both retellings. Furthermore, given that scores on the target-word vocabulary test were not significantly different between the two conditions, the correlation results suggest that the robot's expressivity did not impact initial encoding, but did encourage children to emulate the robot in their subsequent retelling of the story.

## Engagement

### Interview Questions

We found no difference between conditions in children's responses to the interview questions. Children reported that
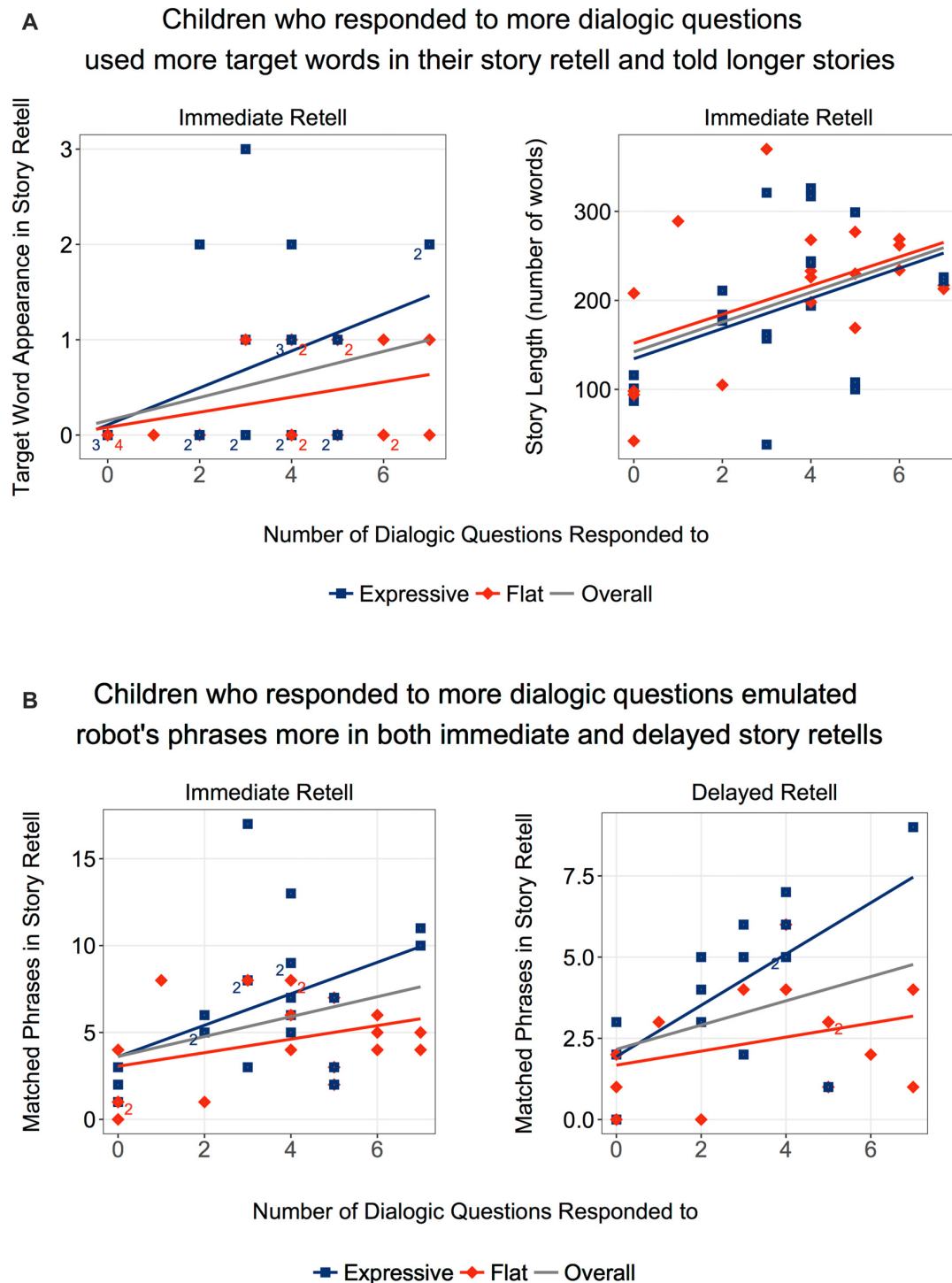
**FIGURE 7 | (A)** Children who responded to the robot's dialogic questions were more likely to use the target words in their retells and tell longer stories. **(B)** These children were also more likely to emulate robot's story in terms of phrase similarity, in both immediate and delayed retell. The trend was primarily driven by the *Expressive* condition.

they liked the story (*Median* = 5, *Mode* = 5, *Range* = 1–5, *Inter-Quartile Range (IQR)* = 1) and that they liked Tega (*Median* = 5, *Mode* = 5, *Range* = 3–5, *IQR* = 0). For example,

one child said said he liked the story because "in the end they found a new pet frog". Children's reasons for liking Tega included physical characteristics, such as "furry", "cute",

**FIGURE 8 |** Children in the Expressive condition showed stronger emulation of the robot's story in terms of phrase similarity, in both immediate and delayed retell. *Statistically significant at $p < 0.05$.
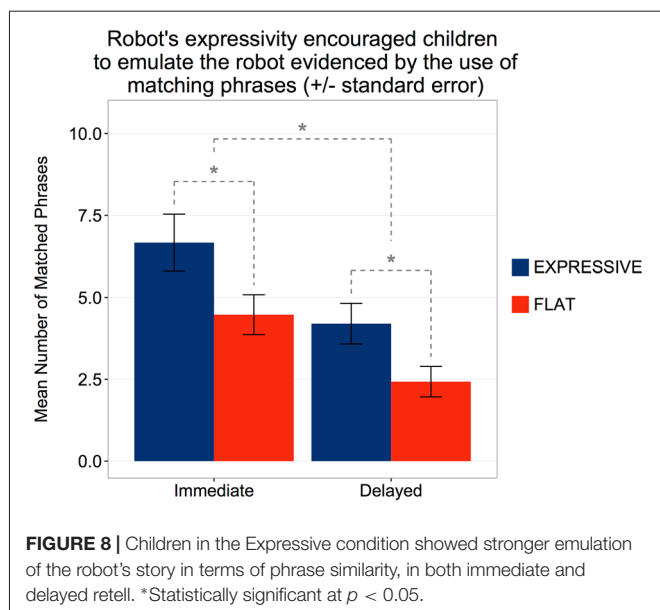
and "red", as well as personality traits including "kind" and "nice".

Children thought Tega could help them feel better (*Median* = 5, *Mode* = 5, *Range* = 1–5, *IQR* = 0), saying, for example, that "he's cute, funny, and makes me smile", and "would give a big hug". They thought Tega helped them learn

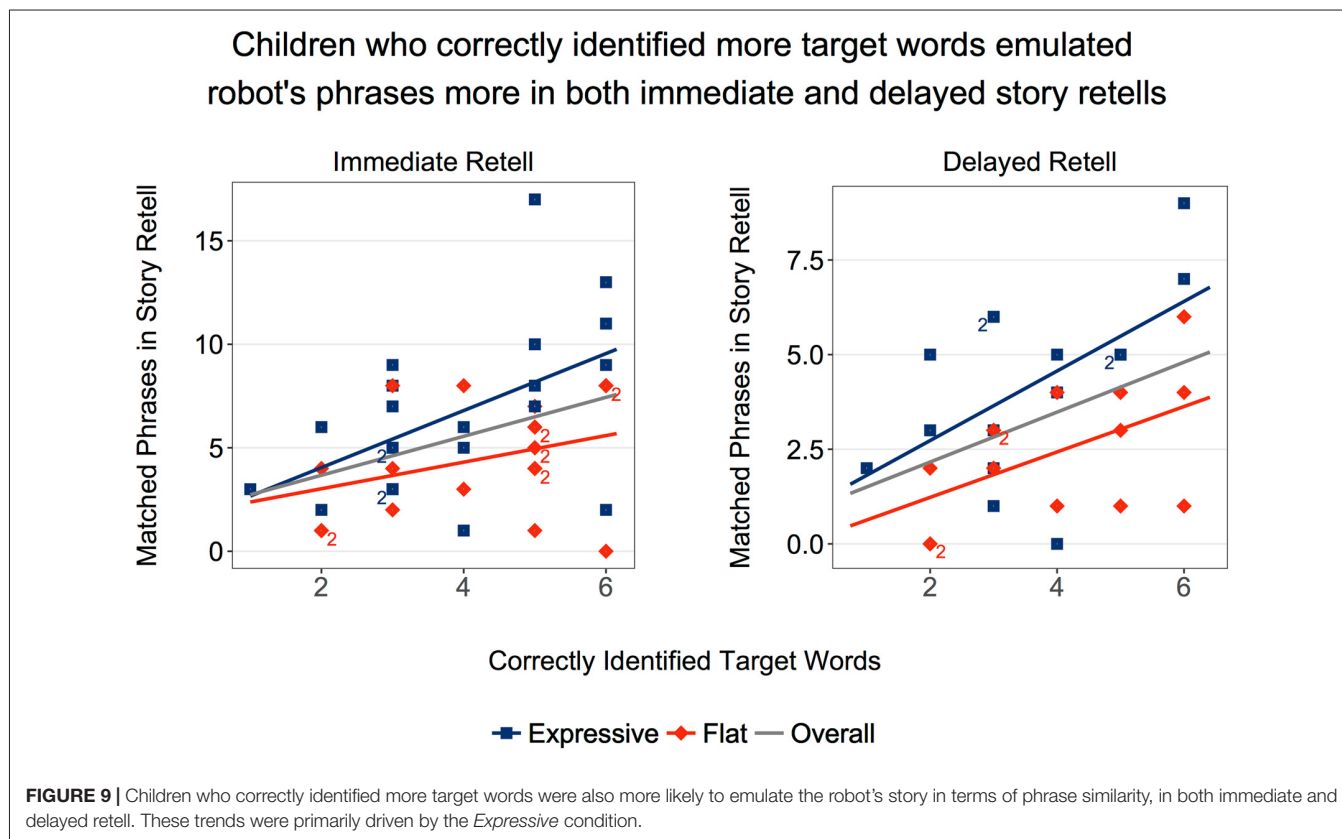the story (*Median* = 5, *Mode* = 5, *Range* = 2–5, *IQR* = 0). One child reported Tega was helpful because "the story was a little bit long", while another said "because he asked me what happened in the story". Another child also noted the questions, saying "stopped to ask questions and talked slowly so I could understand". Children thought their friends would like reading with Tega (*Median* = 5, *Mode* = 5, *Range* = 1–5, *IQR* = 0), because "he's a nice robot and will be nice to them", and "Tega's got a lot of good stories, and is good at telling them".

When asked if they would prefer to play again with Tega or with the Toucan puppet, 26 children picked Tega, 11 picked Toucan, and 8 either said "both", "not sure", or did not respond. They justified picking Toucan with reasons such as "Toucan didn't hear the story", "because he fell asleep and is super, super soft", and "because she's very sleepy and never listens". They justified picking Tega with various reasons including "because Tega can listen and Toucan is just a puppet", "because she read the story to me", "because he's fun", and "I like her". Thus, we see that children felt the desire to be fair in making sure Toucan got a chance to hear the story, and a desire to reciprocate Tega's sharing of a story with them, as well as expressing general liking for the robot.

When asked to describe Tega to a friend, 44% of children described the robot using positive traits (e.g., nice, helpful, smart, fun) in the *Expressive* condition and 48% in the *Flat* condition, *ns*. For example, one child said, "he told me about antlers. Tega is



**FIGURE 9 |** Children who correctly identified more target words were also more likely to emulate the robot's story in terms of phrase similarity, in both immediate and delayed retell. These trends were primarily driven by the *Expressive* condition.

very helpful", while another reported "that he read me a story and will be a nice robot to them". In sum, the expressiveness of the robot did not influence how children described the robot to a peer. Many of the other 56% of children in the *Expressive* condition and the 52% of children in the *Flat* condition focused on the robot's physical characteristics, for example, "red and blue, stripes, big eyes, tuft of blue hair, phone for face, fuzzy, cute smile". One child said Tega "looks like a rock star".
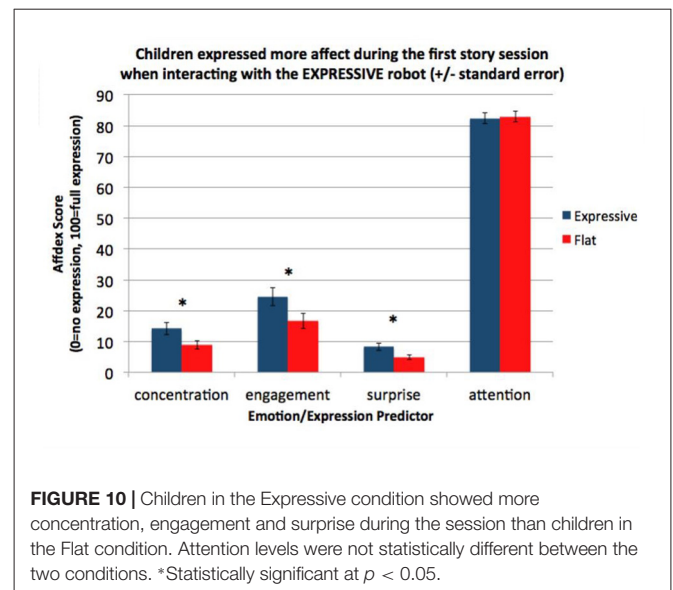
## Children's Expressivity

We analyzed affect data for 36 children (19 in the *Expressive* condition and 17 in the *Flat* condition). For the remaining nine children, no affect data were collected either because the children's faces were not detected by the system, or because of other system failures.

As described earlier, we focused our analysis on the four affective states most relevant to our research questions: concentration, engagement, surprise and attention. All other affective states were measured by Affdex very rarely (less than 5% of the time). We found that overall, children maintained attention throughout most of the session, were engaged by the robot, showed some concentration, and displayed surprise during the story (**Table 3**).

To evaluate whether the robot's vocal expressiveness influenced children's facial expressiveness, we examined the mean levels of the four affective states across the entire session by condition. We conducted a one-way ANCOVA with condition (*Expressive* vs. *Flat*) for each Affdex score, with age as a covariate. The analysis revealed that children in the *Expressive* condition showed significantly higher mean levels of concentration, $F_{(1,32)} = 4.77$, $p = 0.036$, $\eta^2 = 0.127$; engagement, $F_{(1,32)} = 4.15$, $p = 0.049$, $\eta^2 = 0.112$; and surprise, $F_{(1,32)} = 5.21$, $p = 0.029$, $\eta^2 = 0.13$, than children in the *Flat* condition, but that children's attention was not significantly different, $F_{(1,32)} = 0.111$, $p = 0.741$. Furthermore, these differences were not affected by children's age (**Table 3**, **Figure 10**). The lack of difference in children's attention demonstrated that the differences in the concentration, engagement and surprise levels across the two conditions were not a result of children paying less attention to the Flat robot's story.

Next, we asked whether children's affect changed during the session. We split the affect data into two halves—the first half of the session and the second half of the session—using the data timestamps to determine the session halfway point We ran a 2 × 2 mixed design ANOVA with time (within: first half vs. second half) × condition (between: *Expressive* vs. *Flat*) for each of the affect scores. These analyses revealed main effects



**FIGURE 10 |** Children in the Expressive condition showed more concentration, engagement and surprise during the session than children in the Flat condition. Attention levels were not statistically different between the two conditions. *Statistically significant at $p < 0.05$.

of condition on children's concentration scores, $F_{(1,34)} = 4.71$, $p = 0.037$, $\eta^2 = 0.067$; engagement scores, $F_{(1,34)} = 4.16$, $p = 0.049$, $\eta^2 = 0.075$; and surprise scores, $F_{(1,34)} = 5.36$, $p = 0.027$, $\eta^2 = 0.090$. In all three cases, children displayed greater affect in the *Expressive* condition than the *Flat* condition (see **Figures 11B–D**). There were no main effects of time or any significant interactions for these affect measures. However, we did see a main effect of time for children's attention scores, $F_{(1,34)} = 7.84$, $p = 0.008$, $\eta^2 = 0.044$. In both conditions, children's attention scores declined over time (**Figure 11A**).

In summary, although all children were less attentive over time, they showed more facial expressiveness throughout the whole session with the expressive robot than with the flat robot.

## Language Status

We completed our analyses by checking whether the results were stronger or weaker based on children's language status (i.e., native English speakers, ELL, or bilingual). The differences were modest and are reported here.

First, with regards to learning new vocabulary, a one-way ANOVA with age as a covariate revealed that children's language status affected how many target vocabulary words they identified correctly, $F_{(3,37)} = 4.10$, $p = 0.012$, $\eta^2 = 0.230$, but vocabulary learning was not affected by age (**Figure 12**). *Post hoc* pairwise comparisons showed that children who were native English speakers correctly identified more words ($M = 4.53$ correct, $SD = 1.23$) than ELL children ($M = 3.13$ correct, $SD = 1.30$), $p = 0.002$. Bilingual speakers also identified more words correctly ($M = 4.86$, $SD = 1.07$) than ELL children, $p = 0.005$, but were not significantly different from the native English speakers.

Second, native English speakers used more target words ($M = 0.94$, $SD = 0.93$) than ELL students ($M = 0.14$, $SD = 0.36$) in the immediate retell, $t_{(20)} = -3.16$, $p = 0.005$. Bilingual students were in-between ($M = 0.29$, $SD = 0.49$). This trend

**TABLE 3 |** Analysis of four facial expressions during the interaction by condition.

| Expression | Overall mean (SD) | Expressive mean (SD) | Flat mean (SD) |
|---|---|---|---|
| Concentration | 11.7 (7.63) | 14.1 (8.33) | 8.93 (5.83) |
| Engagement | 20.8 (12.1) | 24.5 (12.6) | 16.6 (10.3) |
| Surprise | 6.71 (4.57) | 8.28 (4.99) | 4.95 (3.39) |
| Attention | 82.6 (7.45) | 82.4 (7.47) | 83.0 (7.63) |

*Values can range from 0 (no expression present) to 100 (expression fully present).*
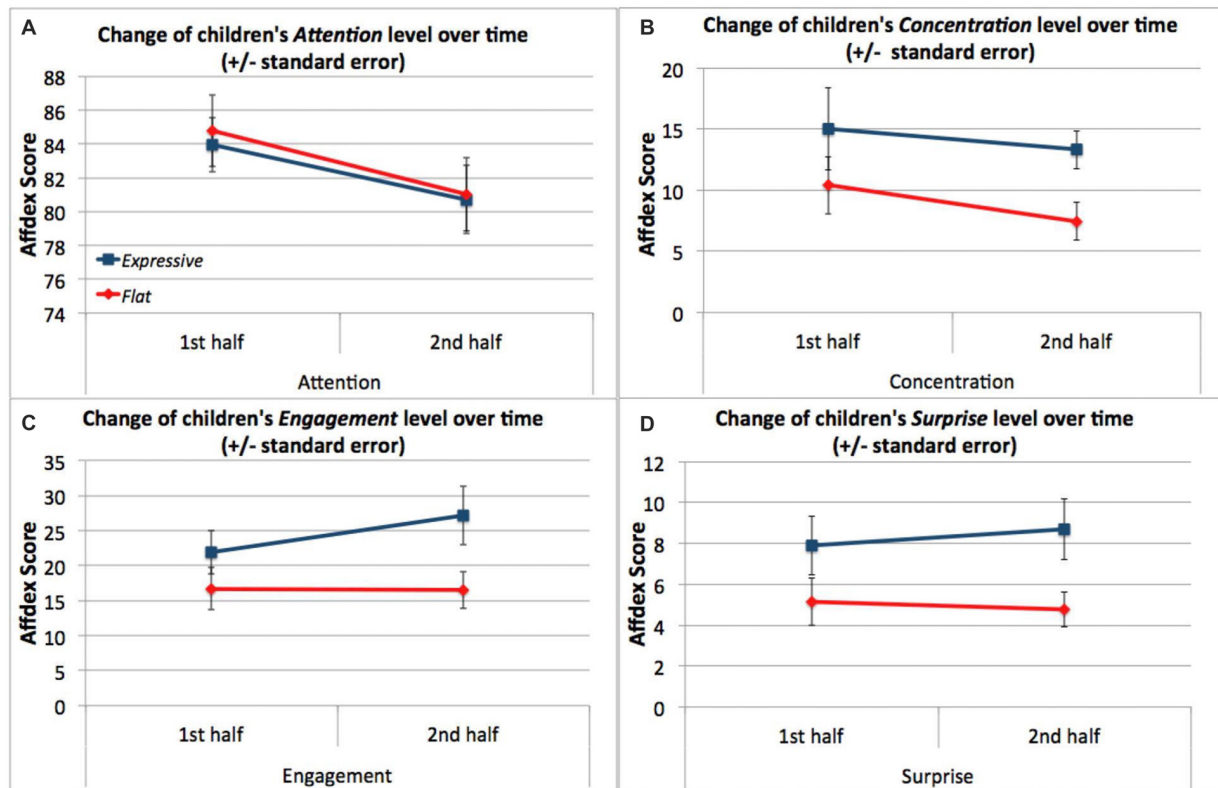
**FIGURE 11 | (A)** Children's level of attention decreased in the course of the session but showed no difference between conditions. **(B)** The concentration level of children in the Expressive condition was consistently higher than that of children in the Flat condition. **(C)** The engagement level of children in the Expressive condition was consistently higher than that of children in the Flat condition. **(D)** The surprise level of children in the Expressive condition was consistently higher than that of children in the Flat condition.

was primarily driven by the *Expressive* condition, $F_{(2,34)} = 5.458$, $p = 0.009$, rather than the *Flat* condition. *Post hoc* pairwise comparison within the *Expressive* condition showed that native speakers used more target words than bilingual speakers,



**FIGURE 12 |** Children who were native English speakers or bilingual correctly identified more of the target vocabulary words than did English Language Learners (ELL) children. *Statistically significant at $p < 0.05$.

$t_{(8)} = 3.00$, $p = 0.017$, and more than ELL speakers, $t_{(13)} = 7.45$, $p < 0.001$. Bilingual speakers also used more target words than the ELL group in the immediate retell, $t_{(11)} = 2.75$, $p = 0.019$.

Lastly, native English speakers showed stronger phrase mirroring behavior ($M = 6.56$, $SD = 4.49$) than ELL students ($M = 4.43$, $SD = 2.38$) in the *Expressive* condition in the immediate retell, $t_{(13)} = 3.41$, $p = 0.005$. The robot's expressivity had a significant effect on native English speakers' usage of similar phrases in both the immediate retell (*Expressive* $M = 10.17$, $SD = 4.40$, *Flat* $M = 4.40$, $SD = 2.99$), $t_{(14)} = 3.139$, $p = 0.007$; and in the delayed retell (*Expressive* $M = 5.50$, $SD = 2.52$, *Flat* $M = 2.38$, $SD = 1.30$), $t_{(10)} = 2.904$, $p = 0.016$. Though not significant, ELL children trended toward also using more similar phrases when they heard the story from the *Expressive* robot in both the immediate retell (*Expressive* $M = 4.55$, $SD = 1.94$, *Flat* $M = 4.20$, $SD = 3.27$) and the delayed retell (*Expressive* $M = 3.78$, $SD = 1.86$, *Flat* $M = 2.20$, $SD = 2.49$).

In summary, as might be expected, native English speakers performed better on the vocabulary test, used more target words, and showed more phrase matching than either ELL or bilingual children.

## DISCUSSION

We asked whether children would learn from a dialogic, storytelling robot and whether the robot's effectiveness as a narrator and teacher would vary with the expressiveness of the robot's voice. We hypothesized that a more expressive voice would lead to greater engagement and greater learning. Below, we review the main findings pertinent to each of these questions and then turn to their implications.

Whether the robot spoke with a flat or expressive voice, children were highly attentive in listening to the robot—as indexed by their head orientation—when it was recounting the picture book story. Moreover, irrespective of the robot's voice, children were able to acquire new vocabulary items embedded in the story. Although some children may have already known some of the target words, as indicated by their above-zero recognition of the target words from the story version they did *not* hear, the interaction between story version heard and scores on each set of words (shown in **Figure 2**) shows that genuine learning did occur. Children could also retell the story (with the help of the picture book) both immediately afterwards and some weeks later. At their initial retelling, children typically produced a story about half as long as the one they had heard, sometimes including a newly acquired vocabulary item. Finally, when they were invited to provide both an explicit evaluation and a free-form description, children were equally positive about the robot whether they had listened to the flat or the expressive robot.

Despite this equivalence with respect to attentiveness, encoding and evaluation, there were several indications that children's mode of listening was different for the two robots. First, as they listened to the expressive rather than the flat robot, children's facial expressions betrayed more concentration (i.e., more brow furrowing and less smiling), more engagement (i.e., greater overall muscle activation) and more surprise (i.e., more brow raising with open mouth). Thus, children were not only attentive to what the robot was saying, they also displayed signs of greater emotional engagement.

Furthermore, inclusion of the newly acquired vocabulary items in the initial retelling was more frequent among children who listened to the expressive rather than the flat robot. Note that children's score on the target-word test was not significantly different between the two conditions, suggesting that children who correctly identified the target words in the *Expressive* condition tended to also use them in their story retell whereas children who correctly identified the target words in the *Flat* condition were less likely to use them in their story recall. Thus, although children were able to acquire new vocabulary from either robot (*receptive* vocabulary knowledge), they were more likely to subsequently use that vocabulary in their stories if the expressive robot had been the narrator (i.e., *productive* vocabulary knowledge). This pattern of findings implies that children could encode and retain new input from either robot, but they were more likely to engage with the expressive robot during the narration and more likely to emulate the expressive

robot's narrative vocabulary in their own recounting. That is, interacting with the expressive robot led to greater *behavioral* outcomes—producing new words rather than merely identifying them.

Further signs of the differential impact of the two robots were found at the delayed retelling. Whereas there was a considerable decline in story length among children who had heard the flat robot, there was no such decline among children who had heard the expressive robot. Again, we cannot ascribe this difference to differences in encoding. Children in each condition had told equally long stories on their initial retelling. A more plausible interpretation is that children who had heard the expressive robot were more inclined to emulate its narrative than children who had heard the flat robot. More detailed support for this interpretation emerged in children's story phrasing. At both retellings, children were more likely to echo the expressive rather than the flat robot in terms of using parallel phrases. This may also indicate that children were engaging with the expressive robot as a more socially dynamic agent, since past research has shown that children are more likely to use particular syntactic forms when primed by an adult (e.g., Huttenlocher et al., 2004). In addition, recent work by Kennedy et al. (2017) showed that a robot that used more nonverbal immediacy behaviors (e.g., gestures, gaze, vocal prosody, facial expressions, proximity and body orientation, touch) led to greater short story recall by children. The difference in the expressive vs. flat robot's vocal qualities (e.g., intonation and prosody) could have led to a difference in the perceived nonverbal immediacy of the robot, which may have led to the differences in children's engagement with the robot as a socially dynamic agent.

Both the expressive and the flat robot asked dialogic questions about the story as they narrated it. The more often children answered these dialogic questions the more vocabulary items they learned. Here too, however, the robot's voice made a difference. The link between question answering and vocabulary acquisition was only significant for the *Expressive* condition. Children who answered more dialogic questions also displayed greater fidelity to the robot's story in terms of phrase usage when they retold it, but again this link was only significant for the *Expressive* condition. Thus, answering more of the robot's questions was associated with the acquisition of more vocabulary and greater phrase emulation but only for the expressive robot. Finally children's score on the target-word vocabulary test correlated with the number of matched phrases they used at both retellings. However, this correlation emerged only for children in the expressive condition, again consistent with the idea that robot expressivity enhanced emulation but not initial encoding.

In sum, we obtained two broad patterns of results. On the one hand, both robots were equally successful in capturing children's attention, telling a story that children were subsequently able to narrate, and teaching the children new vocabulary items. On the other hand, as compared to the flat robot, the expressive robot provoked stronger emotional engagement in the story as it was being narrated, greater inclusion of the newly learned vocabulary into the retelling of the story and greater fidelity to the original story during the retelling. A plausible interpretation of these two

patterns is that story narration *per se* was sufficient to capture children's attention and sufficient to ensure encoding both of the story itself and of the new vocabulary. By contrast the mode in which the story was narrated—expressive or flat—impacted the extent to which the child eventually cast him or herself into the role enacted by the narrator. More specifically, it is plausible that children who were emotionally engaged by the expressive robot were more prone to re-enact the story-telling mode of the robot when it was their turn to tell the story to the puppet: they were more likely to reproduce some of the unfamiliar nouns that they had heard the robot use and more likely to mimic the specific phrases included in the robot's narrative.

It is tempting to conclude that children identified more with the expressive robot and found it more appealing. It is important to emphasize, however, that no signs of that differentiation were apparent either in children's explicit verbal judgments about the two robots or in the open-ended descriptions. In either case, children were quite positive about both of the robots. An important implication of these findings, therefore, is that children's verbal ratings of the robots are not a completely accurate guide to the effectiveness of the robots as role models. Future research on social robots as companions and pedagogues should pay heed to such findings. More generally, the results indicate that it is important to assess the influence and impact of a robot via a multiplicity of measures rather than via questionnaires or self-report.

## Language Status

When analyzing children's learning and performance based on their language status, we saw only modest differences. These differences—in which native English speakers and bilingual children correctly identified more target vocabulary words with both robots, and showed stronger phrase mirroring and use more target words than ELL students with the expressive robot—were not unexpected, given that bilingual and native English speaking children have greater familiarity with the language. Nevertheless, it is important to note that both native English speakers and ELL children who heard the story from the *Expressive* robot reused and retained more information from the robot's story. Thus, despite the limitations listed in the following section, these results suggest that the storytelling activity was an effective intervention for all the native English speakers, the bilingual and the ELL children, leading to learning and engagement by all groups. This is an important finding given that ELL children arguably need the most additional support for their language development (Páez et al., 2007). Effective and engaging language learning interventions like this one that can benefit the entire classroom—native English speakers, bilingual, and ELL children alike—will be important educational tools in years to come.

## Limitations

We should note several limitations of this study. First, some potentially important individual differences among children, such as their learning ability, socio-economic status and sociability were not controlled. Second, although 45 children participated in the study ranging from 4 to 7 years, we did not have an equal number of children at each age. We also did not have an equal number of children with each language status. In future work, it will be important to assess a more homogenous sample, as well as the degree to which our results remain stable across these individual differences and across the preschool and elementary school years.

In addition, we did not have complete story retelling data for all children. As reported earlier, the audio quality of some of the recordings of children's retells prevented analysis, and not all children performed delayed retellings. As a result of this and the aforementioned imbalances in age and language, the analyses we report here are under-powered. This is exploratory work, and the result should be interpreted in light of this fact. Future work should take greater effort to collect quality audio recordings and to see all children at the delayed test.

Finally, while the target vocabulary words used were uncommon, some children may still have known them—particularly older children, given the correlation between age and target words identified. The rarity of the words may have also increased their saliency, being a cue for children to pay attention to the words. Follow-up studies should either consider using nonce words or include a vocabulary pretest for the target words.

## Future Directions

The technology landscape continues to rapidly evolve from passively consumed content such as television and radio to interactive and social experiences enabled through digital technology and the internet. Each new technology transforms the ways we interact with one another, how we communicate and share, how we learn, tell stories and experience imaginary worlds.

Today, the linguistic and interpersonal environment of children is comprised of other people, yet children are increasingly growing up talking *with* AI-based technologies, too. Despite the proliferation of such technologies, very little is understood about children's language acquisition in this emerging social-technological landscape. While it has been argued in the past that children cannot learn language from impersonal media because language acquisition is socially gated (e.g., Kuhl, 2007, 2011), the reality of social robots forces us to revisit our past assumptions. These assumptions need revisiting because numerous studies have now shown that children and adults interact with social robots as social others (Breazeal et al., 2008, 2016b; DeSteno et al., 2012). Social robots represent a new and provocative psychological category betwixt and between inanimate things and socially animate beings. They bridge the digital world of content and information to the physically co-present and interpersonal world of people. Because of this, we are likely to interact with social robots differently than prior technologies. As such, social robots open new opportunities for how educational content and experiences can be brought to the general public, just as their technological predecessors have.

Therefore, how *should* social robots be designed to best foster the learning, development, and benefit to children? This

is very new territory, indeed. This work explores three key avenues, although there are many others to explore, and to explore deeply.

In the context of language learning for preschool age children, we begin by applying knowledge and taking inspiration from how children learn language through storytelling with a peer-like companion. Children learn quite a lot from interacting with and socially modeling the behavior and attitudes of their peers, and in prior work, we have seen behaviors suggesting that children also socially model or emulate the behavior of social robots. For instance, we have found that children become more emotively expressive when a robot is more expressive (Spaulding et al., 2016). We see this effect again in this work. We have also observed that when children play with a "curious" robot that exhibits pro-curious behaviors and attitudes, children express and engage in more curious behaviors (Gordon et al., 2015) and are more willing to teach new tasks to a robot peer (Park and Howard, 2015). In the present study, we found evidence of this social modeling effect in terms of children emulating the linguistic phrases and vocabulary a robot uses. This peer-learning dynamic is quite different from how children learn with other technologies.

Emotional expressivity is another characteristic that social robots bring to interaction. Understanding the impact of emotion and expressive behavior on learning with young children is an area worth further systematic investigation. It is generally accepted that telling a story more expressively will make it more engaging. Social robots enable us to study the impact of expressivity on children's behavior and learning in a more systematic and carefully controlled way. Because of these attributes, social robots could serve as a compelling tool to gain insights into children's social development and learning.

In this work, we observed a greater tendency for children to emulate a storytelling robot's phrasing when the robot was more vocally expressive; children reproduced this pattern after a month-long delay. Further research is warranted to understand whether children are encoding the information differently when the delivery is more expressive, or whether they are simply more apt to emulate the robot when it is more expressive.

We see growing evidence that the more socially expressive and interactive a robot is, the more it "opens the spigot" to children's social engagement and learning. This suggests a new paradigm for educational technology and how it promotes children's learning and development. It is increasingly clear that it is not just the introduction of a social robot into an educational context that matters, *but how socially designed* the robot is that impacts children's behavior and learning.

Finally, for social robots to have a large-scale impact in the educational realm, research should extend beyond the context of 1:1 interaction of a social robot with a child. We need to also understand how to design social robots to support and foster peer learning among groups of children; we need to understand how social robots can best support and include the participation of adults, such as teachers and parents, or facilitate classroom orchestration (e.g., Dillenbourg and Jermann, 2010); and we need to understand how to effectively integrate robots into the broader educational context of the classroom and continued learning at home. Much work remains to be done in order to understand how to best design social robots that can successfully engage and support learning over longitudinal time scales, where the opportunity to deeply attune to the individual child exists—not only in terms of curricular goals, but in order to foster positive attitudes toward learning and challenge, and to build trust and rapport as well. Finally, as research matures and social robots become an affordable mass consumer technology, there exists many opportunities for social robots to help support and augment learning experiences for children who are underserved, at-risk, or have other learning challenges.

## AUTHOR CONTRIBUTIONS

The study was conceived and designed by JMKW, SJ, HWP, SR, PLH, DD and CLB. Data analysis was performed by JMKW, SJ, HWP, SR and AA. The article was drafted, written, revised and approved by JMKW, SJ, HWP, SR, AA, PLH, DD and CLB.

## ACKNOWLEDGMENTS

## REFERENCES

Baxter, P., Kennedy, J., Senft, E., Lemaignan, S., and Belpaeme, T. (2016). "From characterising three years of HRI to methodology and reporting recommendations," in *The 11th ACM/IEEE International Conference on Human Robot Interaction HRI'16* (Piscataway, NJ: IEEE Press), 391–398. Available online at: http://dl.acm.org/citation.cfm?id=2906831.2906897. [Accessed on August 23, 2016].

Bloom, P. (2000). *How Children Learn the Meanings of Words.* Cambridge, MA: MIT press.

Boteanu, A., Chernova, S., Nuñez, D. S., and Breazeal, C. (2016). Fostering parent-child dialog through automated discussion suggestions. *User Model. User-Adap. Inter.* 26, 393–423. doi: 10.1007/s11257-016-9176-8

Boudreau, D. M., and Hedberg, N. L. (1999). A comparison of early literacy skills in children with specific language impairment and their typically developing peers. *Am. J. Speech-Lang. Pathol.* 8, 249–260. doi: 10.1044/1058-0360.0803.249

Breazeal, C. L. (2004). *Designing Sociable Robots.* Cambridge, MA: MIT Press.

Breazeal, C., Dautenhahn, K., and Kanda, T. (2016a). "Social robotics," in *Springer Handbook of Robotics*, eds B. Siciliano and O. Khatib (New York, NY: Springer International Publishing), 1935–1972. doi: 10.1007/978-3-319-32552-1_72.

Breazeal, C., Harris, P. L., DeSteno, D., Kory Westlund, J. M., Dickens, L., and Jeong, S. (2016b). Young children treat robots as informants. *Top. Cogn. Sci.* 8, 481–491. doi: 10.1111/tops.12192

Breazeal, C., Takanishi, A., and Kobayashi, T. (2008). "Social robots that interact with people," in *Springer Handbook of Robotics*, eds B. Siciliano and O. Khatib (Berlin, Heidelberg: Springer), 1349–1369. doi: 10.1007/978-3-540-30301-5_59

Camras, L. A., Bakeman, R., Chen, Y., Norris, K., and Cain, T. R. (2006). Culture, ethnicity and children's facial expressions: a study of European American, mainland Chinese, Chinese American, and adopted Chinese girls. *Emotion* 6, 103–114. doi: 10.1037/1528-3542.6.1.103

Chang, A., Breazeal, C., Faridi, F., Roberts, T., Davenport, G., Lieberman, H., et al. (2012). "Textual tinkerability: encouraging storytelling behaviors to foster emergent literacy," in *CHI'12 Extended Abstracts on Human Factors in Computing Systems CHI EA'12* (New York, NY: ACM), 505–520. doi: 10.1145/2212776.2212826

Chang, C.-W., Lee, J.-H., Chao, P.-Y., Wang, C.-Y., and Chen, G.-D. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educ. Technol. Soc.* 13, 13–24.

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., et al. (2009). Young childrens' trust in their mothers' claims: longitudinal links with attachment security in infancy. *Child Dev.* 80, 750–761. doi: 10.1111/j.1467-8624.2009.01295.x

Cremin, T., Flewitt, R., Mardell, B., and Swann, J. (2016). *Storytelling in Early Childhood: Enriching Language, Literacy and Classroom Culture.* New York, NY: Routledge.

DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., et al. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychol. Sci.* 23, 1549–1556. doi: 10.1177/0956797612448793

Diehl, J. J., Bennetto, L., and Young, E. C. (2006). Story recall and narrative coherence of high-functioning children with autism spectrum disorders. *J. Abnorm. Child Psychol.* 34, 83–98. doi: 10.1007/s10802-005-9003-x

Dillenbourg, P., and Jermann, P. (2010). "Technology for classroom orchestration," in *New Science of Learning*, eds M. S. Khine and I. M. Saleh (New York, NY: Springer), 525–552. doi: 10.1007/978-1-4419-5716-0_26

Dunn, L. M., and Dunn, L. M. (2007). *Peabody Picture Vocabulary Test.* 4th Edn. Minneapolis, MN: Pearson Assessments.

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System (FACS). A Technique for the Measurement of Facial Action.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., and Rosenberg, E. L. (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS).* New York, NY: Oxford University Press.

Eyssel, F., Kuchenbrandt, D., Bobinger, S., de Ruiter, L., and Hegel, F. (2012). "If you sound like me, you must be more human: on the interplay of robot and user features on human-robot acceptance and anthropomorphism," in *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction HRI'12* (New York, NY: ACM), 125–126. doi: 10.1145/2157689.2157717

Feil-Seifer, D., and Mataric, M. J. (2011). Socially assistive robotics. *IEEE Robot. Autom. Mag.* 18, 24–31. doi: 10.1109/MRA.2010.940150

Friesen, W. V., and Ekman, P. (1983). *EMFACS-7: Emotional Facial Action Coding System.* Unpublished Manuscript. San Francisco, CA: University of California.

Gordon, G., Breazeal, C., and Engel, S. (2015). "Can children catch curiosity from a social robot?," in *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction HRI '15* (New York, NY: ACM), 91–98. doi: 10.1145/2696454.2696469

Gordon, G., Spaulding, S., Kory Westlund, J., Lee, J. J., Plummer, L., Martinez, M., et al. (2016). "Affective personalization of a social robot tutor for children's second language skill," in *Proceedings of the 30th AAAI Conference on Artificial Intelligence* (Palo Alto, CA).

Greenhalgh, K. S., and Strong, C. J. (2001). Literate language features in spoken narratives of children with typical language and children with language impairments. *Lang. Speech Hear. Serv. Sch.* 32, 114–125. doi: 10.1044/0161-1461(2001/010)

Hargrave, A. C., and Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: the benefits of regular reading and dialogic reading. *Early Child. Res. Q.* 15, 75–90. doi: 10.1016/S0885-2006(99)00038-1

Harris, P. L. (2007). Trust. *Dev. Sci.* 10, 135–138. doi: 10.1111/j.1467-7687.2007.00575.x

Harris, P. L., Guz, G. R., Lipian, M. S., and Man-Shu, Z. (1985). Insight into the time course of emotion among western and chinese children. *Child Dev.* 56, 972–988. doi: 10.1111/j.1467-8624.1985.tb00170.x

Heilmann, J., Miller, J. F., Nockerts, A., and Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *Am. J. Speech Lang. Pathol.* 19, 154–166. doi: 10.1044/1058-0360(2009/08-0024)

Huttenlocher, J., Vasilyeva, M., and Shimpi, P. (2004). Syntactic priming in young children. *J. Mem. Lang.* 50, 182–195. doi: 10.1016/j.jml.2003.09.003

Isbell, R., Sobol, J., Lindauer, L., and Lowrance, A. (2004). The effects of storytelling and story reading on the oral language complexity and story comprehension of young children. *Early Child. Educ. J.* 32, 157–163. doi: 10.1023/b:ecej.0000048967.94189.a3

Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Hum.-Comput. Interact.* 19, 61–84. doi: 10.1207/s15327051hci1901&2_4

Kahn, P. H., Gary, H. E., and Shen, S. (2013). Children's social relationships with current and near and future robots. *Child Dev. Perspect.* 7, 32–37. doi: 10.1111/cdep.12011

Kennedy, J., Baxter, P., and Belpaeme, T. (2017). Nonverbal immediacy as a characterisation of social behaviour for human-robot interaction. *Int. J. Soc. Robot.* 9, 109–128. doi: 10.1007/s12369-016-0378-3

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016). "Social robot tutoring for child second language learning," in *Proceedings of the 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch, New Zealand), 231–238. Available online at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7451757 [Accessed on May 20, 2016].

Kory, J. (2014). *Storytelling with Robots: Effects of Robot Language Level on Children's Language Learning.* Master's Thesis, Cambridge, MA: Media Arts and Sciences, Massachusetts Institute of Technology.

Kory, J., and Breazeal, C. (2014). "Storytelling with robots: learning companions for preschool children's language development," in *Proceedings of the 2014 RO-MAN: The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (Edinburgh, UK), 643–648. doi: 10.1109/ROMAN.2014.6926325

Kory Westlund, J., and Breazeal, C. (2015). "The interplay of robot language level with children's language learning during storytelling," in *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts HRI'15 Extended Abstracts* (New York, NY: ACM), 65–66. doi: 10.1145/2701973.2701989

Kory Westlund, J. M., Lee, J. J., Plummer, L., Faridi, F., Gray, J., Berlin, M., et al. (2016). "Tega: a social robot," in *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Cambridge, MA), 561. doi: 10.1109/HRI.2016.7451856

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Dev. Sci.* 10, 110–120. doi: 10.1111/j.1467-7687.2007.00572.x

Kuhl, P. K. (2011). "Social mechanisms in early language acquisition: understanding integrated brain systems supporting language," in *The Oxford Handbook of Social Neuroscience*, eds J. Decety and J. Cacioppo (Oxford: Oxford University Press), 649–667. Available online at: http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780195342161.001.0001/oxfordhb-9780195342161-e-043 [Accessed September 12, 2016].

Leite, I., Castellano, G., Pereira, A., Martinho, C., and Paiva, A. (2014). Empathic robots for long-term interaction. *Int. J. Soc. Robot.* 6, 329–341. doi: 10.1007/s12369-014-0227-1

LoBue, V., and Thrasher, C. (2015). The child affective facial expression (CAFE) set: validity and reliability from untrained adults. *Front. Psychol.* 5:1532. doi: 10.3389/fpsyg.2014.01532

Lubold, N., Walker, E., and Pon-Barry, H. (2016). "Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction HRI'16* (Piscataway, NJ: IEEE Press), 255–262. Available online at:

http://dl.acm.org/citation.cfm?id=2906831.2906876 [Accessed September 14, 2016].

Lyberg-Åhlander, V., Haake, M., Brännström, J., Schötz, S., and Sahlén, B. (2015). Does the speaker's voice quality influence children's performance on a language comprehension test? *Int. J. Speech Lang. Pathol.* 17, 63–73. doi: 10.3109/17549507.2014.898098

McDuff, D., Kaliouby, R., and Picard, R. W. (2015). "Crowdsourcing facial responses to online videos: extended abstract," in *Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (Xi'an), 512–518. doi: 10.1109/ACII.2015.7344618

McDuff, D., Kaliouby, R., Senechal, T., Amr, M., Cohn, J. F., and Picard, R. (2013). "Affectiva-MIT facial expression dataset (AM-FED): naturalistic and spontaneous facial expressions collected in-the-wild," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops* 881–888. doi: 10.1109/CVPRW.2013.130

McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., and Kaliouby, R. (2016). "AFFDEX SDK: a cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems CHI EA'16* (New York, NY: ACM), 3723–3726. doi: 10.1145/2851581.2890247

Meltzoff, A. N., Kuhl, P. K., Movellan, J., and Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science* 325, 284–288. doi: 10.1126/science.1175626

Morton, V., Watson, D. R., and Watson, D. R. (2001). The impact of impaired vocal quality on children's ability to process spoken language. *Logoped. Phoniatr. Vocol.* 26, 17–25. doi: 10.1080/14015430118232

Movellan, J., Eckhardt, M., Virnes, M., and Rodriguez, A. (2009). "Sociable robot improves toddler vocabulary skills," in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction (ACM)* (La Jolla, CA), 307–308.

Naigles, L. R., and Mayeux, L. (2001). "Television as incidental language teacher," in *Handbook of Children And The Media*, eds D. G. Singer and J. L. Singer (Thousand Oaks, CA: Sage Publications), 135–152.

Niculescu, A., van Dijk, B., Nijholt, A., Li, H., and See, S. L. (2013). Making social robots more attractive: the effects of voice pitch, humor and empathy. *Int. J. Soc. Robot.* 5, 171–191. doi: 10.1007/s12369-012-0171-x

Nuñez, D. S. (2015). *GlobalLit: A Platform for Collecting, Analyzing, and Reacting to Children's Usage Data on Tablet Computers.* Cambridge, MA: Massachusetts Institute of Technology. Available online at: http://dspace.mit.edu/handle/1721.1/98622 [Accessed September 12, 2016].

Páez, M. M., Tabors, P. O., and López, L. M. (2007). Dual language and literacy development of Spanish-speaking preschool children. *J. Appl. Dev. Psychol.* 28, 85–102. doi: 10.1016/j.appdev.2006.12.007

Park, H. W., Coogle, R. A., and Howard, A. (2014). "Using a shared tablet workspace for interactive demonstrations during human-robot learning scenarios," in *Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA)* (Hong Kong), 2713–2719. doi: 10.1109/ICRA.2014.6907248

Park, H. W., and Howard, A. M. (2015). "Retrieving experience: interactive instance-based learning methods for building robot companions," in *Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA)*, (Seattle, WA) 6140–6145. doi: 10.1109/ICRA.2015.7140061

Read, J. C., and MacFarlane, S. (2006). "Using the fun toolkit and other survey methods to gather opinions in child computer interaction," in *Proceedings of the 2006 Conference on Interaction Design and Children IDC'06* (New York, NY: ACM), 81–88. doi: 10.1145/1139073.1139096

Rogerson, J., and Dodd, B. (2005). Is there an effect of dysphonic teachers' voices on children's processing of spoken language? *J. Voice* 19, 47–60. doi: 10.1016/j.jvoice.2004.02.007

Sage, K. D., and Baldwin, D. (2010). Social gating and pedagogy: mechanisms for learning and implications for robotics. *Neural Netw.* 23, 1091–1098. doi: 10.1016/j.neunet.2010.09.004

Sandygulova, A., and O'Hare, G. M. P. (2015). "Children's perception of synthesized voice: robot's gender, age and accent," in *Social Robotics Lecture Notes in Computer Science*, eds A. Tapus, E. André, J.-C. Martin, F. Ferland and M. Ammi (Cham: Springer International Publishing), 594–602. doi: 10.1007/978-3-319-25554-5_59

Senechal, T., McDuff, D., and Kaliouby, R. (2015). "Facial action unit detection using active learning and an efficient non-linear kernel approximation," in *Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW) ICCVW'15* (Washington, DC: IEEE Computer Society), 10–18. doi: 10.1109/ICCVW.2015.11

Spaulding, S., Gordon, G., and Breazeal, C. (2016). "Affect-aware student models for robot tutors," in *Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems AAMAS'16* (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), 864–872. Available online at: http://dl.acm.org/citation.cfm?id=2937029.2937050

Speaker, K. M., Taylor, D., and Kamen, R. (2004). Storytelling: enhancing language acquisition in young children. *Education* 125, 3–14.

Szafir, D., and Mutlu, B. (2012). "Pay attention!: designing adaptive agents that monitor and improve user engagement," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'12* (New York: ACM), 11–20. doi: 10.1145/2207676.2207679

Tamagawa, R., Watson, C. I., Kuo, I. H., MacDonald, B. A., and Broadbent, E. (2011). The effects of synthesized voice accents on user perceptions of robots. *Int. J. Soc. Robot.* 3, 253–262. doi: 10.1007/s12369-011-0100-4

Tanaka, F., and Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *J. Hum. Robot Interact.* 1, 78–95. doi: 10.5898/jhri.1.1.tanaka

Valdez-Menchaca, M. C., and Whitehurst, G. J. (1992). Accelerating language development through picture book reading: a systematic extension to Mexican day care. *Dev. Psychol.* 28, 1106–1114. doi: 10.1037/0012-1649.28.6.1106

Walters, M. L., Syrdal, D. S., Koay, K. L., Dautenhahn, K., and te Boekhorst, R. (2008). "Human approach distances to a mechanical-looking robot with different robot voice styles," in *RO-MAN 2008–The 17th IEEE International Symposium on Robot and Human Interactive Communication* (Munich), 707–712. doi: 10.1109/ROMAN.2008.4600750

Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., et al. (1988). Accelerating language development through picture book reading. *Dev. Psychol.* 24, 552–559. doi: 10.1037/0012-1649.24.4.552

# Social Interaction Affects Neural Outcomes of Sign Language Learning As a Foreign Language in Adults

Noriaki Yusa[1]*, Jungho Kim[2], Masatoshi Koizumi[3], Motoaki Sugiura[4] and Ryuta Kawashima[4]

[1] Department of English, Miyagi Gakuin Women's University, Sendai, Japan, [2] Department of Foreign Languages, Kyoto Women's University, Kyoto, Japan, [3] Graduate School of Arts and Letters, Tohoku University, Sendai, Japan, [4] Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan

Children naturally acquire a language in social contexts where they interact with their caregivers. Indeed, research shows that social interaction facilitates lexical and phonological development at the early stages of child language acquisition. It is not clear, however, whether the relationship between social interaction and learning applies to adult second language acquisition of syntactic rules. Does learning second language syntactic rules through social interactions with a native speaker or without such interactions impact behavior and the brain? The current study aims to answer this question. Adult Japanese participants learned a new foreign language, Japanese sign language (JSL), either through a native deaf signer or via DVDs. Neural correlates of acquiring new linguistic knowledge were investigated using functional magnetic resonance imaging (fMRI). The participants in each group were indistinguishable in terms of their behavioral data after the instruction. The fMRI data, however, revealed significant differences in the neural activities between two groups. Significant activations in the left inferior frontal gyrus (IFG) were found for the participants who learned JSL through interactions with the native signer. In contrast, no cortical activation change in the left IFG was found for the group who experienced the same visual input for the same duration via the DVD presentation. Given that the left IFG is involved in the syntactic processing of language, spoken or signed, learning through social interactions resulted in an fMRI signature typical of native speakers: activation of the left IFG. Thus, broadly speaking, availability of communicative interaction is necessary for second language acquisition and this results in observed changes in the brain.

Keywords: social interaction, foreign language learning, fMRI, Japanese sign language, syntax, left inferior frontal gyrus

## INTRODUCTION

It is a trivial fact that all normal children effortlessly acquire a particular language used around them. Less trivial is the fact that children do so through social interactions: children cannot acquire a language from linguistic input such as TV, or computer presentations (Sachs et al., 1981; Baker, 2001; Kuhl et al., 2003). This fact is all the more worth remarking, considering that other cognitive

systems such as the visual system do not require human interaction for them to develop properly from birth. In this sense, language is uniquely human in that it is inherently social (de Saussure, 1916/1972).

In addition to the atypical cases of children raised in social isolation such as the wild boy of Aveyron (Lane, 1976) and Genie (Curtiss, 1977), the importance of a communicative partner in language acquisition has been illustrated by Sachs et al. (1981): the case of hearing children raised by deaf parents, who attempted in vain to teach them spoken English via television. Kuhl et al. (2003) provide more direct evidence for the experimental effects of social interactions on phonetic learning (discrimination) in a foreign language. Infants less than 6 months old of age can discriminate various speech contrasts in the world that do not exist in their mother tongues (Eimas et al., 1971; Werker and Tees, 1984), but they lose the discriminating ability between 6 and 12 months of age (Werker and Tees, 1984). During this period, they grow into "native listeners" from "universal listeners." In Kuhl et al.'s experiment, 9-to-10-old month American babies were exposed to a new language, Mandarin Chinese, over 4–6 weeks through four different speakers of Mandarin Chinese or via televised recordings of Mandarin Chinese speakers. After exposure, the researchers performed a head-turn phonetic discrimination task of a Mandarin fricative-affricate contrast that does not exist in English. Only infants exposed to Mandarin Chinese speakers retained their sensitivity to distinguish the non-native Mandarin speech contrast and showed the same level of phonetic discrimination as native speakers of Mandarin Chinese. The result clearly indicates that phonetic learning is not triggered by simple exposure to linguistic input, but that infants must be exposed to a language in socially interactive situations to develop speech perception (Kuhl, 2007). TV programs or DVDs cannot be substitutes for human instruction in the early periods of phonetic learning.

Social interactions provide a variety of information needed for language development, so that several explanations have been offered for the findings in Kuhl et al. (2003). Social interactions may "attract more attention and increase motivation" in infants (Verga and Kotz, 2013, p. 3) resulting in phonetic learning; joint attention may provide more referential information needed for the association of a word and its referent (Kuhl et al., 2003); social contingency or back-and-forth feedback from humans may play a vital role in language development (Kuhl, 2007; Roseberry et al., 2014); infants may not be familiar or experienced with DVD presentations. These explanations are not mutually exclusive or implausible in that infants acquire a language through social interactions with their caregivers that involve child-directed speech (Bruner, 1983). The reader is referred to Hoff (2006) and Verga and Kotz (2013) for the review of relevant studies showing that social interaction influences language learning in infants.

Despite the alleged importance of social interaction in language development, previous language learning studies on social interaction only focused on vocabulary learning (Kuhl, 2007) and phonetic discrimination (Kuhl et al., 2003; Kuhl, 2007) in a foreign language during childhood, and word learning in a first language (Krcmar et al., 2007; Roseberry et al., 2009; Verga and Kotz, 2013). Language is, however, more than words

and sounds. Human language is a computational system of connecting meaning and sound (or a visual-manual channel in sign languages) by means of syntactic structure. Syntactic structure has not been observed in other species (Hauser et al., 2002), so structure dependence in this sense is the most characterizing feature of human language (Chomsky, 2013a,b; Everaert et al., 2015; Berwick and Chomsky, 2016).

In spite of the fact that syntax is "the basic property" of human language (Chomsky, 2013a,b) and that social interaction plays a key role in early language development, syntax has not been discussed in adult second language acquisition research from the perspective of social interaction. It is true that the number of neuroscience studies on social interactions has exponentially increased over the last decade (for a research review, see Verga and Kotz, 2013), little social neuroscience research has until recently dealt with adult second language acquisition. One of the few studies on second language acquisition in different social settings is Jeong et al. (2010). Jeong and her colleagues tested the effects of social interactions on the acquisition of second language vocabulary by adult learners. They compared the retrieval of words learned from text-based learning (written translations) and that of words learned from situation-based learning (real-life situations). The result shows that the comprehension of words learned through movie-clips depicting a social situation elicited activity in the right supramarginal gyrus similar to that evoked by the comprehension of vocabulary in one's native language. The result indicates the effects of social interaction in second language acquisition of vocabulary on the brain, but it should be noted that participants in situation-based learning contexts learned foreign language vocabulary through "artificial" movie-clips of a dialogue. Therefore, it remains to be elucidated what differences *natural* social interaction with a teacher makes in second language acquisition in comparison to learning a second language with artificial interaction such as DVDs (Verga and Kotz, 2013). In fact, no study, to our knowledge, has yet investigated how social interaction during foreign language learning in adulthood will affect neural mechanisms. Thus, whether adult learners benefit from learning in social contexts is still an open issue.

Given that linguistic knowledge is internalized in the brain and that more people are using computer-assisted learning without human interaction, we reasonably address the non-trivial question of whether social interaction will have distinctive effects on the brains of adults learning foreign language syntax, which is more complex than vocabulary and phonetic learning. Due to resource constraints, non-interactive learning through a combination of audio and video is common among second language learners who have few opportunities to interact with native speakers of a target language. It should be noted, however, that there is no clear evidence that computer-supported learning without social interaction has the same effects on the learning of syntactic rules in a foreign language as learning through human interaction. Most studies on social interactions are based on behavior or performance data, but behavioral data have some limitations. First, behavioral scores of linguistic knowledge are blurred by numerous factors such as attention, cognition, and perception. It is, therefore, extremely difficult, if not impossible,

to tease them apart, which in turn makes the interpretation of the performance data inconclusive (Raizada et al., 2008). Second, behavioral data do not reveal the neurocognitive mechanisms responsible for the processing of second language knowledge. Third, similar behavioral data do "not necessarily implicate reliance on similar neural mechanisms" (Morgan-Short et al., 2012, p. 934). Indeed, several brain imaging studies (Musso et al., 2003; Osterhout et al., 2008; Sakai et al., 2009) have reported the evidence for the difference between performance data and their respective neuroscience data.

The present paper discusses whether presence or absence of a human being has distinct effects on neural (fMRI) and behavioral (performance) measures of syntactic processing of a foreign language in adults. As a foreign language, we tested the acquisition of Japanese sign language (JSL) by Japanese adults who had not learned JSL. A sign language is mistakenly conceived to be a kind of artificial pantomime-like gesture lacking linguistic structure or at least a variant of a spoken language, but neither is well-grounded. A great deal of research in recent years demonstrates that sign language is a natural language with rich grammatical properties that characterize other natural languages such as spoken English or Japanese (Sandler and Lillo-Martin, 2006). Cecchetto et al. (2012), for example, demonstrate that Italian sign language respects structure dependence based on abstract hierarchical syntactic structure characteristic of only human languages (Chomsky, 2013a,b; Everaert et al., 2015). Studies of language development have also provided evidence that deaf children experience almost the same stages of language development as hearing children (Petitto and Marentette, 1991). Deaf babies, for instance, experience a stage of manual babbling during the same period as hearing children go through a stage of vocalization babbling. This confirms that irrespective of superficial speech modality differences, the same mechanism applies to core functions of sign and spoken languages. Differences between the two languages derive from the modalities in which they are produced and comprehended (MacSweeney et al., 2008). Furthermore, neuroimaging studies show that comprehension of spoken and sign languages activates the classical language brain regions including the left inferior frontal gyrus (IFG) (Sakai et al., 2005) in addition to the left superior temporal gyrus and sulcus (for a relevant literature review, see MacSweeney et al., 2008).

Areas in the left IFG, specifically the posterior pars opercularis (BA 44) and the more anterior pars triangularis (BA 45) of Broca's area, are known to be involved in processing linguistic and non-linguistic information (e.g., Koechlin and Jubault, 2006; Tettamanti and Weniger, 2006). This leads to the suggestion that Broca's area works as a "supramodal processor of hierarchical structures" (Tettamanti and Weniger, 2006). The "supramodal syntactic processor" (Clerget et al., 2013) has been localized either in BA 44 (Bahlmann et al., 2009; Fazio et al., 2009) or in BA 45 (Santi and Grodzinsky, 2010; Pallier et al., 2011). We will not go into the issue of which region, BA 44 or BA 45, is selectively responsible for processing syntactic structure (Musso et al., 2003; Pallier et al., 2011; Yusa, 2012; Goucha and Friederici, 2015; Zaccarella and Friederici, 2015; Zaccarella et al., 2017), but instead follow previous research showing that

syntactic processing in a first language and a second language activates Broca's area in the left IFG (Perani and Abutalebi, 2005; Abutalebi, 2008). In particular, syntactic rules satisfying structure dependence selectively activate the language area of the brain, specifically the left IFG, while syntactic rules violating structure-dependent rules do not (Musso et al., 2003; Yusa et al., 2011). In addition, instruction effects of syntax in a second language are reflected in the left IFG (Musso et al., 2003; Sakai et al., 2009; Yusa et al., 2011). Recent extensive research on syntax processing also validates the claim that the left IFG is responsible for processing syntactic structure (Moro et al., 2001; Musso et al., 2003; Friederici et al., 2011; Goucha and Friederici, 2015). All taken together, we assume that activation of the left IFG is indicative of the acquisition of syntactic rules respecting structure-dependence.

We show, by examining the acquisition of JSL under two different social learning conditions, that learning through interaction with a deaf signer resulted in a stronger activation of the left IFG than learning through identical input via DVD presentations, though behavioral data did not show distinct differences.

## JAPANESE SIGN LANGUAGE

JSL has the basic word or constituent order of SOV (subject-object-verb), but exhibits free word order as spoken Japanese does. The basic word order SOV can be changed into its topicalized order OSV with the topicalized O accompanied by a set of non-manual markers (NMM) such as eyebrow raising and nodding. There are, however, some restrictions on constituent order. Consider the following wh-cleft sentence "/PT-I/ /FATHER/ /OCCUPATION/ /WHAT/ /DOCTOR/," which means "What my father is is a doctor." (Following conventions, signs are written as glosses in capital letters and PT stands for "pointing to the nose or chest with the index finger of either hand"). In JSL, possessives cannot be moved from their modifying head nouns, whose phenomenon in spoken languages has been discussed in terms of the Left Branch Condition since Ross (1967). We call this the Possessive Construction Restriction. For example, possessive pronoun MY indicated by /PT-I/ cannot be separated from *FATHER* as in "/FATHER/ /OCCUPATION/ /PT-I/ /WHAT/ /DOCTOR/," which is ungrammatical. Although languages differ as to whether they allow left-branch extraction (Bošković, 2005), it suffices to note for the purpose of the present paper that left-branch extraction is disallowed in JSL. What matters here is the syntactic difference between the optionality of topicalization of objects and the prohibition of the movement of possessives from their modifying nouns. In this sense, movement of a constituent respects structure dependence in a sense that movement of a constituent depends on the syntactic structure of the moved constituent.

It is interesting to note at this point that even a native speaker of JSL in our experiment had not considered the possessive construction restriction until it was pointed out, so it is natural that no book on JSL we know of refers to any aspects of the possessive construction restriction.
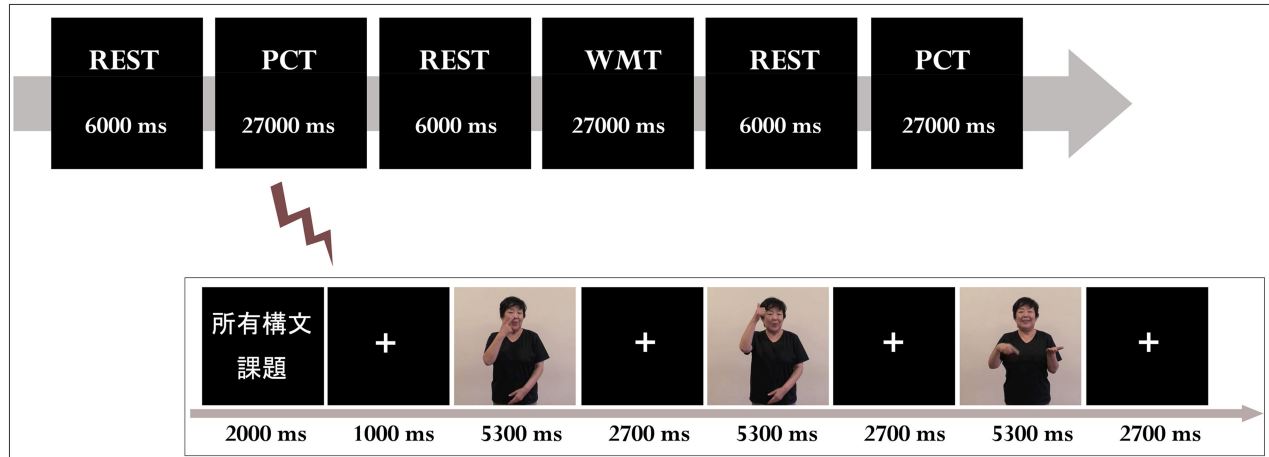
**FIGURE 1 | Timeline in the experimental task.** PCT, Possessive Construction Task; WMT, Working Memory Task; REST, Rest Task. The experiment was performed in a block design. Participants were asked to judge whether the JSL they saw on the screen was correct. Response time was recorded from the beginning of each stimulus sentence until the button was pressed. E-prime ver. 2.0 (Psychology Software Tools) was used to present the stimuli and obtain the behavioral data.

## MATERIALS AND METHODS

### Participants

Forty six adult Japanese without any knowledge of JSL participated in our experiment. Participants were all recruited from Miyagi Gakuin Women's University, Sendai, Japan. They were divided into the Live-Exposure Group and the DVD-Exposure Group on the basis of working memory measured by the reading span test. As a result, those groups were indistinguishable on working memory before JSL lessons ($t_{(44)} = 0.249$, $p = 0.80$). Before the experiment, all participants were provided with minute explanations of the experiment and its safety. They gave written informed consent for the study and right-handedness was verified using the Edinburgh Inventory (Oldfield, 1971). All experiments were performed in compliance with the relevant institutional guidelines approved by Tohoku University. Approval for the study was obtained from the Ethics Committee of the Institute of Development, Aging and Cancer, Tohoku University.

### Procedure and Stimuli

The Live-Exposure Group and the DVD-Exposure Group learned JSL in two different contexts. Twenty two participants in the Live-Exposure Group learned JSL through social interactions with a native signer of JSL in ten 80-min classes in 1 month, where they learned the JSL expressions related to self-introduction, numbers, family, transportation, weather, hobbies, food and so on. A native signer did not teach the participants the grammar of JSL, but a large number of expressions in JSL in an implicit way. On the other hand, 24 participants in the DVD-Exposure Group learned JSL in the same number of classes during the same period through the DVDs that recorded the class lessons in the Live-Exposure Group. Therefore, the difference between the Live-Exposure Group and the DVD-Exposure Group was the existence/absence of social interchanges through a deaf signer.

The participants in both groups underwent two sets of fMRI measurements after the 4th class (TEST 1) and the 10th class (TEST 2). Stimuli were visually presented to the participants in a block design (**Figure 1**). The total number of stimuli was 72, which was divided into two sessions with 36 stimuli each. Each session consisted of three conditions: Possessive Construction Task (correct/incorrect), Working Memory Task (correct/incorrect), and Rest Task (**Table 1**). In the Possessive Construction Task (PCT), the participants were visually presented with both possible and impossible JSL in random order on a screen; they had to judge the grammaticality of the JSL by pressing a button. The second task was the Working Memory Task (WMT): the participants were presented with three signs in sequence and had to judge whether the sequence included three different signs: A stimulus with three different signs was judged as a "grammatical JSL," whereas a stimulus involving two identical signs was regarded as an "ungrammatical JSL."

The Rest Task (REST) required the participants to gaze at a fixation cross. All stimuli were controlled using E-prime ver. 2.0 (Psychology Software Tools). **Figure 1** shows how the experiments proceeded. Following Hashimoto and Sakai (2002), we employed the WMT in our experiment. Its rationale was to disassociate working memory effects from the comprehension of JSL. The comprehension of a language is based on structure-dependent operations. Moreover, language comprehension is incremental in that linguistic information of a lexical item is processed immediately every time it is encountered (Neville et al., 1991; Phillips, 2003). Therefore, the PCM task implicitly required the participants to encode linguistic information of signs and decode it from working memory when they judged the JSL.

### Image Acquisition

Functional neuroimaging data were acquired with a 3.0 Tesla MRI scanner (Philips Achieva Quasar Dual, Philips Medical Systems, Best, The Netherlands) using a gradient echo planar image (EPI) sequence ([TE] = 30 ms, field of view [FOV] =

**TABLE 1 | Sample stimuli used in the fMRI experiment.**

Possessive Construction Task (PCT)

What my father is is a doctor.

- Grammatical JSL

  /PT-I/ /FATHER/ /OCCUPATION/ /WHAT/ /DOCTOR/

  "What my father is is a doctor"

- Ungrammatical JSL

  /FATHER/ /OCCUPATION/ /PT-I/ /WHAT/ /DOCTOR/

  "What my father is is a doctor"

Working Memory Task (WMT)

- Grammatical JSL

  /JOB/ /EAT/ /STUDY/

- Ungrammatical JSL

  /WRITE/ /READ/ /WRITE/

*PT; Pointing Sign.*

*Japanese Sign language (JSL) has the basic word or constituent order of SOV (subject-object-verb), but exhibits free word order as spoken Japanese does. There are, however, some restrictions on constituent order. Consider "/PT-I/ /FATHER/ /OCCUPATION/ /WHAT/ /DOCTOR/," which means "What my father is is a doctor." In JSL, possessives cannot be moved from their modifying head nouns, whose phenomenon in spoken languages has been discussed in terms of the Left Branch Condition since Ross (1967). For example, possessive pronoun MY indicated by /PT-I/ cannot be separated from FATHER as in "/FATHER/ /OCCUPATION/ /PT-I/ /WHAT/ /DOCTOR/," which is ungrammatical. In the Possessive Construction Task (PCT), where the participants were visually presented with both possible and impossible JSL in random order on a screen, they had to judge the grammaticality of the JSL by pressing a button. In the Working Memory Task (WMT), the participants were instructed to determine whether the sequence just presented included three different signs. A stimulus with different three signs was treated as a "grammatical JSL" stimulus, whereas a stimulus including two identical signs was regarded as an "ungrammatical JSL" sequence.*

192 mm, flip angle [FA] = 70°, slice thickness = 5 mm, slice gap = 0 mm). Thirty-two axial slices spanning the entire brain were obtained every 2 s. After the attainment of functional imaging, T1-weighted anatomical images were also acquired from each participant.

## Analysis

All data processing and group analyses were performed using MATLAB (The Mathworks Inc., Natick, MA, USA) and SPM8 (Wellcome Department of Cognitive Neurology, London, UK). The acquisition timing of each slice was corrected using the middle (16th in time) slice as a reference for EPI data. In order to correct for head movement artifacts, functional images were first resliced and subsequently realigned with the first scan of the subjects. After alignment to the AC-PC line, each participant's T1-weighted image was coregistered to the mean functional EPI image and segmented using the standard tissue probability maps provided in SPM8. The coregistered structural image was spatially normalized to the Montreal Neurological Institute (MNI) standard brain template. All normalized functional images were then smoothed with a Gaussian kernel of 8 mm full-width at half-maximum (FWHM). An analysis of the tasks for each participant was conducted at the first statistical stage and a group statistical analysis at the second stage. Contrasts in the PCT − WMT condition was calculated using a one sample $t$-test. The threshold for significant activation of each contrast was set at $p < 0.001$, uncorrected. The spatial extent threshold was set at $k = 10$ voxels. Finally, we performed a region of interest (ROI)

**TABLE 2 | Error rates (%) and reaction times (ms) for PCT.**

| | Error rates (%) | | Reaction times (ms) | |
|---|---|---|---|---|
| | TEST 1 | TEST 2 | TEST 1 | TEST 2 |
| DVD-Exposure Group | 42.5 (11.7) | 26.2 (19.5) | 6,117 (494) | 5,984 (583) |
| Live-Exposure Group | 39.1 (21.2) | 15.5 (19.2) | 5,899 (588) | 5,788 (524) |
| | $p = 0.54$, ns | $p = 0.09$, ns | $p = 0.22$, ns | $p = 0.28$, ns |

*There was no significant difference in the percentage of error rates and reaction times in TEST 1 or TEST 2 between the Live-Exposure Group and the DVD-Exposure Group. A significant performance improvement was, however, found in the DVD-Exposure Group as well as the Live-Exposure Group between TEST 1 and TEST 2: the percentage of errors in TEST 2 significantly decreased with both groups as compared to that in TEST 1 [Live-Exposure Group; $t_{(17)} = 4.79$, $p < 0.001$; DVD-Exposure Group; $t_{(20)} = 4.82$, $p < 0.001$].*

analysis in the brain area obtained from the comparison [PCT − WMT$_{(TEST\ 2)}$] − [PCT − WMT $_{(TEST\ 1)}$]. Activation maxima are reported as MNI-coordinates and anatomical regions are based on the Talairach Client (Lancaster and Fox, Research Imaging Center, University of Texas Health Science Center San Antonio; Talairach and Tournoux, 1988; Lancaster et al., 2000).

## Predictions

If linguistic input is sufficient to induce JSL learning in adults, then exposure to JSL via a deaf signer or DVDs should result in the same changes in behavioral and imaging data. Instead, if social interaction is required and is an important factor in JSL learning, the Live-Exposure Group and the DVD-Exposure Group should show a different pattern of activation in the brain, or more specifically, the former group should have greater activation in the left IFG than the latter group.

## RESULTS

All data processing analyses were performed using SPM8. The threshold for significant activation of each contrast was set at $p < 0.001$, uncorrected. We analyzed data from 18 participants (mean age ± SD: 20.7 ± 0.76 years) in the Live-Exposure Group and 21 participants (mean age ± SD: 20.6 ± 0.76 years) in the DVD-Exposure Group. There was no significant difference in the percentage of error rates in TEST 1 or TEST 2 between the Live-Exposure Group and the DVD-Exposure Group [TEST 1, $t_{(37)} = -0.62$, $p = 0.54$; TEST 2, $t_{(37)} = -1.71$, $p = 0.09$] (**Table 2**). The result indicates that the participants in both groups developed the same level of knowledge of the Possessive Construction at the 4th and 10th trainings; their performance or behavior results were not significantly different. No significant difference in reaction times was observed in TEST 1 or TEST 2 between the Live-Exposure Group and the DVD-Exposure Group, either [TEST 1, $t_{(37)} = -1.26$, $p = 0.22$; TEST 2, $t_{(37)} = -1.09$, $p = 0.28$].

A significant performance improvement was, however, found in the DVD-Exposure Group as well as the Live-Exposure Group between TEST 1 and TEST 2 (**Table 2**): the percentage of errors in TEST2 significantly decreased with both groups as compared to that in TEST 1, indicating that teaching JSL through a native signer or DVDs had significant effects on the acquisition of the

Possessive Construction Restriction [Live-Exposure Group; $t_{(17)}$ = 4.79, $p < 0.001$; DVD-Exposure Group; $t_{(20)} = 4.82$, $p < 0.001$].

## Imaging Data

To identify cortical activation generated in two different learning contexts (i.e., via social interactions with a deaf signer and through DVDs), we subtracted [PCT – WMT$_{(TEST 1)}$] from [PCT – WMT$_{(TEST 2)}$]. **Table 3** shows the activated regions in the comparison of [PCT – WMT$_{(TEST 2)}$] – [PCT – WMT$_{(TEST 1)}$]. For the DVD-Exposure Group, we found increased activations in the right middle frontal gyrus, the bilateral cuneus, the right superior temporal gyrus, the right middle temporal gyrus, the right IFG. For the Live-Exposure Group, activations in the right parietal lobule, the right IFG, the right middle frontal gyrus, the left IFG, the left Inferior parietal gyrus, and the middle frontal gyrus increased.

A ROI analysis of each cluster was conducted using the SPSS 19 (SPSS Inc., IBM, Armonk, NY, USA) on the value of the single voxel of the peak coordinate, which was obtained using an in-house SPM-compatible MATLAB script. The ROI was set at the activated area in the contrast [PCT – WMT$_{(TEST 2)}$] – [PCT – WMT$_{(TEST 1)}$] pooling the data from two groups. Activity

in this ROI was compared in each group between [PCT – WMT$_{(TEST 1)}$] and [PCT – WMT$_{(TEST 2)}$] using a paired $t$-test. Significant activations in the left IFG, an area assumed to be involved in the processing of syntactic rules (Musso et al., 2003; Abutalebi, 2008; Yusa, 2012; Zaccarella et al., 2017), were found only for the Live-Exposure Group [paired $t$-test: $t_{(17)} = -4.88$, $p < 0.001$]. No significant cortical activation change in the left IFG, by contrast, was found for the DVD-Exposure Group, who experienced the same visual input for the same duration via the DVD presentations [paired $t$-test: $t_{(20)} = -0.29$, $p = 0.78$, *n.s.*] (**Figures 2**, **3**; **Table 3**). This result shows that (superficially) similar performance between the groups "does not necessarily implicate reliance on similar neural mechanisms" (Morgan-Short et al., 2012, p. 934). Given that the LIFG is involved in the syntactic processing of language, spoken or signed, only training in an interactional setting resulted in an fMRI signature typical of native speakers: activation of the left IFG.

## DISCUSSION

The aim of the current investigation was to investigate the effects of social interaction on JSL learning in adult speakers. To examine social impacts on learning, we set up two types of learning contexts (that is, learning JSL through a deaf signer or through DVDs). Our results show that participants learned JSL equally in terms of behavioral data in both contexts, but that social interaction caused significant changes in the brain, particularly in the left IFG. This suggests that in addition to early speech learning in infants (Kuhl, 2007), social interaction is crucial in order for adult second language learners to come to rely on native-like neural mechanisms in processing syntactic rules or their efficient use. Social interaction through the interchanges
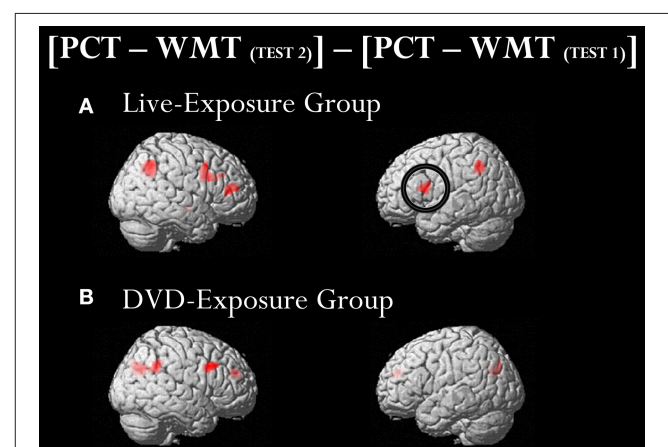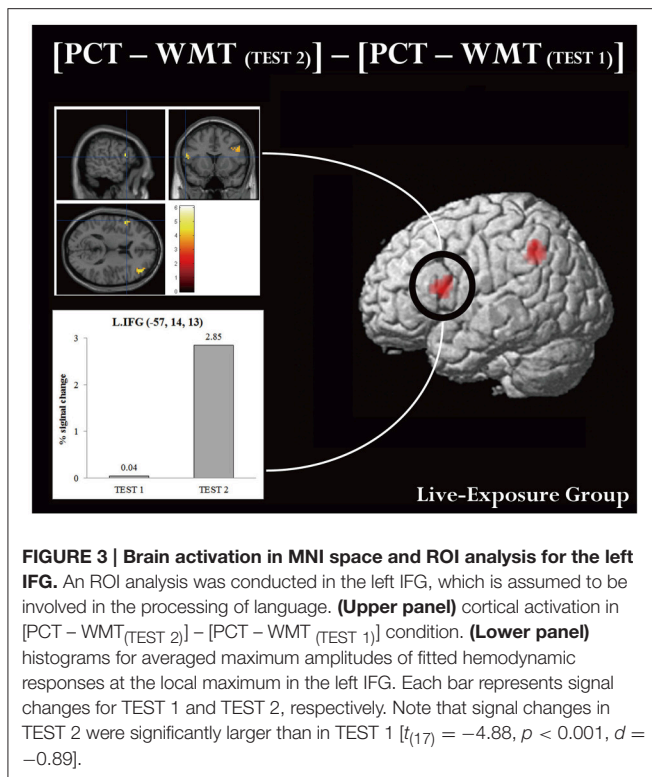
**TABLE 3 | Activated regions in the contrast [PCT – WMT$_{(TEST 2)}$] – [PCT – WMT $_{(TEST 1)}$].**

| Hemisphere | Anatomical region | MNI | | | *t*-value |
|---|---|---|---|---|---|
| | | x | y | z | |
| **DVD-EXPOSURE GROUP** | | | | | |
| R | Medial frontal gyrus (BA 9) | 9 | 44 | 25 | 5.92 |
| R | Cuneus (BA 7) | 9 | −70 | 31 | 5.52 |
| R | Cuneus (BA 7) | 3 | −64 | 31 | 4.49 |
| L | Cuneus (BA 18) | −9 | −76 | 28 | 4.38 |
| R | Superior temporal gyrus (BA 39) | 36 | −49 | 28 | 4.99 |
| R | Middle temporal gyrus (BA 39) | 33 | −61 | 31 | 4.51 |
| R | Supramarginal gyrus (BA 40) | 45 | −49 | 34 | 4.32 |
| R | Inferior frontal gyrus (BA 9) | 45 | 11 | 31 | 4.85 |
| R | Middle frontal gyrus (BA 9) | 36 | 17 | 34 | 4.37 |
| R | Midbrain | 6 | −13 | −5 | 4.08 |
| R | Thalamus | 3 | −4 | 4 | 3.82 |
| **LIVE-EXPOSURE GROUP** | | | | | |
| R | Superior parietal lobule (BA 7) | 30 | −58 | 40 | 6.11 |
| R | Precuneus (BA 7) | 15 | −64 | 37 | 3.89 |
| R | Inferior frontal gyrus (BA 46) | 45 | 38 | 10 | 5.66 |
| R | Middle frontal gyrus (BA46) | 42 | 47 | 16 | 4.47 |
| R | Midbrain | 9 | −13 | −11 | 5.17 |
| R | Midbrain | 3 | −16 | −17 | 4.72 |
| L | Inferior frontal gyrus (BA 44) | −57 | 14 | 13 | 4.88 |
| L | Inferior parietal lobule (BA 40) | −42 | −52 | 40 | 4.65 |
| R | Middle frontal gyrus (BA 9) | 36 | 20 | 25 | 4.20 |
| R | Middle frontal gyrus (BA 9) | 45 | 8 | 37 | 4.15 |
| R | Middle frontal gyrus (BA 9) | 48 | 14 | 31 | 4.06 |

*Respective activated anatomic region, approximate Brodmann's area, right or left (R, L), t-values. Stereotactic coordinates (x, y, z) as defined by MNI are shown for each voxel with a local maximum of t-values in the contrasts indicated (p < 0.001, uncorrected).*



**FIGURE 2 | Brain activated regions in the contrast [PCT – WMT$_{(TEST 2)}$] – [PCT – WMT $_{(TEST 1)}$].** The participants in both groups underwent two sets of fMRI measurements after the 4th class (TEST 1) and the 10th class (TEST 2). To identify cortical activation generated after the instruction, we subtracted [PCT – WMT$_{(TEST 1)}$] from [PCT – WMT$_{(TEST 2)}$]. Significant activations in the left inferior frontal gyrus (IFG) were found only for the Live-Exposure Group **(A)**. No significant cortical activation change, by contrast, was found for the DVD-Exposure Group, who experienced the same visual input for the same duration via the DVD presentations **(B)**.

**FIGURE 3 | Brain activation in MNI space and ROI analysis for the left IFG.** An ROI analysis was conducted in the left IFG, which is assumed to be involved in the processing of language. **(Upper panel)** cortical activation in [PCT – WMT$_{(TEST 2)}$] – [PCT – WMT$_{(TEST 1)}$] condition. **(Lower panel)** histograms for averaged maximum amplitudes of fitted hemodynamic responses at the local maximum in the left IFG. Each bar represents signal changes for TEST 1 and TEST 2, respectively. Note that signal changes in TEST 2 were significantly larger than in TEST 1 [$t_{(17)} = -4.88$, $p < 0.001$, $d = -0.89$].

with a deaf native signer may make it easier to "crack the JSL code," neurologically supporting the view that language is inherently social (de Saussure, 1916/1972). Thus, learning accompanied by changes in brain functions is not triggered solely by linguistic input such as DVDs, but is enhanced by social interaction. The current research provides a significant platform for studies on second language learning in adults: linguistic input is necessary for second language learning, but influences of a social partner are different from the ones exerted from the source without social interactions.

Numerous studies reveal that JSL has linguistic characteristics distinct from spoken Japanese (Fischer, 1996, 2017; Matsuoka, 2015), which are to be discussed below. One might, however, object that the participants in our experiment simply transferred the knowledge of the possessive construction restriction in JSL from spoken Japanese, since extraction of possessives from their modifying nouns is also prohibited in spoken Japanese. This objection is plausible in light of the finding that in bilingualism both languages unconsciously influence each other (Kroll et al., 2006; Jarvis and Pavlenko, 2008), but it cannot explain why only the Live-Exposure Group experienced functional changes in the left IFG. If transfer from spoken Japanese had been a crucial factor in the learning of JSL in our experiment, learning JSL via DVD presentations would also have elicited similar activity in the left IFG. However, the lack of activation in the left IFG in the DVD Group rules out this possibility. Thus, the differences in the left IFG suggest that the two groups employed different mechanisms to learn JSL.

This raises an interesting question of what the DVD Group actually learned in our experiment. On this question, the

activation of the right supramarginal gyrus in the DVD Group is suggestive in terms of the result in Jeong et al. (2010): the right supramarginal gyrus is crucially involved in the retrieval of words learned by means of situation-based learning using media-clips of a dialogue. Note here that situation-based learning in Jeong et al. (2010) roughly corresponds to learning via DVD recordings in our experiment. The right supramarginal gyrus is part of the right parietal lobule, which is considered to play a key role in incorporating multimodal information from different senses (Macaluso and Driver, 2003). Jeong et al. (2010) suggest that the activation of the right supramarginal gyrus is associated with imitation learning, since the area is proposed to constitute a part of human mirror neuron systems (Chong et al., 2008). Mirror neurons are active not only during the execution of an action but also during the observation of the same action (Gallese, 2008). Learners in the DVD Group might have developed the knowledge of JSL only by observing the DVD recordings, inferring the intentions of a signer recorded there and imitating JSL to adapt to a given situation in learning sessions. The imitation of familiar gestures is also known to invoke activation in the right supramarginal gyrus (Peigneux et al., 2004). The right IFG [45, 11, 31] can also be considered the anterior component of the mirror neuron system. Putting these together, it might be reasonable to conclude that participants in the DVD Group developed the knowledge of JSL through imitation learning.

We have assumed, following the generative tradition (Chomsky, 2013a,b; Everaert et al., 2015; Berwick and Chomsky, 2016), that aside from externalization at the sensory-motor level (sign language or speech), the brain contains a universal computational system, which merges or combines smaller elements into larger elements or constituents in a hierarchical manner, generating hierarchical structures. This structure-building operation called Merge is universal, so that it does not need to be learned. If JSL and spoken Japanese differ only in "their modality of externalization" with their syntactic operations the same, one might ask what participants in the Live-Exposure Group learned. It is interesting to note here that activity in the right middle frontal gyrus ([45, 8, 37], [48, 14, 31]) in the Live-Exposure Group might show the involvement of the anterior component of the mirror neuron system, suggesting the role of the mirror neuron in the acquisition of JSL in the Live-Exposure Group. It is natural to think that the Live-Exposure Group learned JSL through observing a teacher use JSL, but second language acquisition involves much more than imitation.

Successful second language acquisition involves assembling or mapping syntactic, semantic and phonological features into new configurations, that is, second language acquisition learners are required to reconfigure features from the way they are coded in the first language into the new configuration where they are represented in the second language; this is a proposal termed "Feature Reassembly Hypothesis" (Lardiere, 2009). On this hypothesis, second language learners of JSL must develop the knowledge of which signs and non-manual markers such as facial expressions, and their variants represent which syntactic, semantic, and phonological features. In addition, they must acquire the knowledge of whether such signs are obligatory, optional or prohibited under

which syntactic, semantic, phonological, lexical and pragmatic conditions (Hwang and Lardiere, 2013; Slabakova, 2016). Assuming the Feature Reassembly Hypothesis, we assume that what developed in the Live-Exposure Group is the knowledge of reassembling relevant features in spoken Japanese into new configurations in JSL by means of associating abstract features carrying grammatical information in spoken Japanese and their exponents (signs) in JSL.

To be more specific, at least two points are relevant to the question of the relation between learning second language syntactic rules and feature-reassembly. One is the knowledge of the wh-cleft in JSL and the other is the knowledge of the possessive construction in JSL.

The wh-cleft in JSL is different at least in three points from the wh-cleft in spoken Japanse (for the wh-cleft in American Sign Language, see Caponigro and Davidson, 2011). The wh-phrase in JSL must be accompanied by NMMs such as "a repeated weak headshake and furrowed eyebrows" (Matsuoka, 2015) as well as "the following fixation of the head" (Ichida, 2005). Following the analysis of wh-interrogatives in JSL by Uchibori and Matsuoka (2016), we assume that the wh-element in JSL is morphologically made up of a wh-phrase (represented by a wh-sign) and a Q-particle or a wh-interrogative marker (represented by wh-NMMs) (Uchibori and Matsuoka, 2016). The lack of these NMMs results in ungrammatical wh-cleft sentences. The wh-phrase and the Q-particle *ka* in spoken Japanese are pronounced in different positions, while in JSL the wh-phrase and the Q-particle must co-occur. Therefore, the participants had to reassembly the Q or wh-interrogative feature into the NMM in JSL and to express the wh-phrase and the NMM simultaneously. Incidentally, it is interesting to note here that *Shushi Nihongo* or *Nihongotaiou Shuwa* "Signed Japanese," a variant of spoken Japanese, lacks NMMs (Kimura, 2011).

Second, semantics is different: the element following the wh-phrase in JSL does not receive a focus interpretation, while the counterpart in spoken Japanese is in focus. Third, pragmatics is different; the wh-cleft in JSL is commonly used and does not sound "orotund" unlike the wh-cleft in spoken Japanese (Matsuoka, 2015). These differences are what the participants learned in our experiments.

Regarding the possessive construction in JSL, nominative "I," expressed by POINTING AT THE SPEAKER, is not accompanied by the NMM of nodding. When nodding co-occurs with pointing at the speaker, it means "and." Thus, the difference between "my father" and "I and father" depends on the NMM (nodding). Therefore, the participants had to learn that the possessive pronoun is morphologically composed of two parts: the sign meaning the first person and the absence of nodding (NMM). It is clear that learning of the wh-cleft and the possessive construction is related to externalization, which is in turn related to the fact that a sign language can use more than one articulator simultaneously.

Second language acquisition is influenced by similarities and differences between the feature arrays charactering the first language and those in the second language input. Consequently, the magnitude of feature reassembly depends on the nature of the input: "feature reassembly may occur slowly or not at all if the relevant evidence is rare or ambiguous in the input" (Slabakova et al., 2014, p. 602). Thus, the knowledge of the association interacts with structure-building operations to result in the knowledge of specific constructions such as the possessive construction in JSL. From this perspective, it is more appropriate to say that as a result of feature reassembly the participants in Live-Exposure Group learned several constructions including the possessive construction. Even so, it is noteworthy that learning JSL through social interactions with a communicative partner had a different impact on the left IFG from learning it via DVD presentations without such interactions. Our result also suggests that the association might be influenced by the source of information, human or non-human, at least in the early stages of foreign language acquisition in adults.

We conclude the paper by pointing out four remaining issues. The first issue is concerned with the fMRI data of the Live-Exposure Group. The difference between TEST 1 and TEST 2 was found at uncorrected thresholds. One possible explanation for this result is that the Live-Exposure Group had already learned the possessive construction at the time of TEST 1, which was conducted just after the fourth class; knowledge of the possessive construction at TEST 1 could have washed away clear instruction effects at TEST 2, leading to the result at uncorrected thresholds. Had TEST 1 been carried out before the instruction of JSL started, more significant results at corrected thresholds should have been obtained.

The second concerns the behavioral results in TEST 2 obtained just after the tenth class, which did not show any significant differences in error rates between the Live-Exposure Group and the DVD-Exposure Group. This result seems strange but it is consistent with previous research showing that the same performance outcomes do not show the use of the same brain system (Poldrack et al., 2001; Foerde et al., 2006; Morgan-Short et al., 2012). Greater changes in the brain may be needed to show the corresponding changes in the behavior (Boyke et al., 2008). It is not clear from our experiment whether knowledge acquired from DVD-Exposure learning is as durable as knowledge obtained from social interactions with a deaf signer. The impact of social interaction on the long-term retention of newly acquired knowledge in adults is an issue for future research.

The third has to do with interactive learning tools such as video chatting with the properties of social interactions and video, as well as interactive media such as Skype or FaceTime. The interactive situation resembles a natural learning situation between a teacher and a student. Positive effects of interactive media use on second-language learning, if confirmed, will provide new insights into the issue of quality and quantity of input in second-language learning, thereby rethinking the issue of critical or sensitive periods in second-language learning. In birds, richer social interaction can delay the critical period closure for learning (Brainard and Knudsen, 1998). Even adults beyond sensitive periods in second language acquisition may also benefit from richer social interaction (Zhang et al., 2009). In addition to the quantity of input, its quality, not age, matters in the

attainment of native-like processing of a second language (Piske and Young-Scholten, 2009).

The last is concerned with the relation of linguistic experience (input) and innate mechanisms in language acquisition. Whatever approaches to language acquisition, there is some consensus that language grows in the brain from the interaction of several factors, including at least three factors: genetic endowment (innate mechanisms), experience (linguistic input) and language-independent properties (Chomsky, 2005; Everaert et al., 2015). Although the importance of the second factor (input) for the ontogenesis of language in an individual is not controversial, what properties are attributed to innate mechanisms characterize two approaches to language acquisition. One approach (called generative approach) assumes that a human is born with the language-dedicated cognitive system (called Universal Grammar), which grows into knowledge of a particular language through the interaction of linguistic experiences; the other (called a general or nativist emergent approach) denies a language-specific innate mechanism, but instead proposes the innate domain-general learning mechanism including statistical learning. On the latter account, linguistic knowledge emerges as a result of linguistic experiences or linguistic usage through statistical learning (O'Grady, 2005). However, the current minimalist program in generative grammar has dramatically minimized the innate language-specific properties by reducing them to other cognitive systems (see Chomsky, 2005, 2013a,b). As a result, the two approaches just mentioned are not as mutually exclusive as they used to be (Yang, 2004; Kirby, 2014). Further research needs to examine whether and to what extent the two approaches converge. It should be noted that generative grammar has never claimed that social interaction or frequency of words is not responsible for language acquisition. Then, what effects does social interaction have on second language acquisition? Our data show that learning second language syntax in social and non-social contexts can lead to differences in brain processing that cannot be reflected by behavioral data. Future research will be needed to characterize the details of the relationship between social interaction and adult second language learning, and thereby to maximize the brain development responsible for learning.

## CONCLUSION

The current study investigated effects of social interaction on the acquisition of syntax in adult second language learners. We found that learning JSL through interactions with a deaf signer resulted in a stronger activation of the left IFG than learning through identical input via DVD presentations, though behavioral data did not show distinct differences. This study provides the first neuroimaging data to show that interaction with a human being aids acquiring syntactic rules and in turn causes significant changes in the brain. If the activation in the left IFG is indicative of native-like processing of syntax, one implication for second language learning is that learning second language syntax in a richer social context may well lead to native-like attainment of second language processing. This implication calls for further studies on whether interactive media such as Skype or FaceTime will induce distinct changes than traditional learning media such as DVDs and TV programs.

## REFERENCES

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychol.* 128, 466–478. doi: 10.1016/j.actpsy.2008.03.014

Bahlmann, J., Schubotz, R. I., Mueller, J. L., Koester, D., and Friederici, A. D. (2009). Neural circuits of hierarchical visuo-spatial sequence processing. *Brain Res.* 1298, 161–170. doi: 10.1016/j.brainres.2009.08.017

Baker, M. (2001). *The Atoms of Language: The Mind's Hidden Rules of Grammar.* New York, NY: Basic Books.

Berwick, R., and Chomsky, N. (2016). *Why Only Us: Language and Evolution.* Cambridge, MA: MIT Press.

Bošković, Ž. (2005). On the locality of left branch extraction and the structure of NP. *Stud. Linguist.* 59, 1–45. doi: 10.1111/j.1467-9582.2005.00118.x

Boyke, J., Driemeyer, J., Gaser, C., Büchel, C., and May, A. (2008). Training-induced brain structure changes in the elderly. *J. Neurosci.* 28, 7031–7035. doi: 10.1523/JNEUROSCI.0742-08.2008

Brainard, M. S., and Knudsen, E. I. (1998). Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *J. Neurosci.* 18, 3929–3942.

Bruner, J. S. (1983). *Child's Talk: Learning to Use Language*. New York, NY: W.W. Norton.

Caponigro, I., and Davidson, K. (2011). Ask, and tell as well: question–answer clauses in American Sign Language. *Natural Lang. Semant.* 19, 323–371. doi: 10.1007/s11050-011-9071-0

Cecchetto, C., Checchetto, A., Geraci, C., and Zucchi, S. (2012). The same constituent structure. *Formal Exp. Adv. Sign Lang.* 2012, 22.

Chomsky, N. (2005). Three factors in language design. *Linguist. Inq.* 36, 1–22. doi: 10.1162/0024389052993655

Chomsky, N. (2013a). What kind of creatures are we? *J. Philos.* CX, 645–662. doi: 10.5840/jphil2013110121

Chomsky, N. (2013b). Problems of projection. *Lingua* 130, 33–49. doi: 10.1016/j.lingua.2012.12.003

Chong, T. T. J., Cunnington, R., Williams, M. A., Kanwisher, N., and Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Curr. Biol.* 18, 1576–1580. doi: 10.1016/j.cub.2008.08.068

Clerget, E., Andre, M. A., and Olivier, E. (2013). Deficit in complex sequence processing after a virtual lesion of BA 45. *PLoS ONE* 8:e63722. doi: 10.1371/journal.pone.0063722

Curtiss, S. (1977). *Genie: A Psycholinguistic Study of Modern-day "Wild Child."* New York, NY: Academic Press.

de Saussure, F. (1916/1972). *Cours de Linguistique Générale*. Paris: Payot.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science* 171, 303–306. doi: 10.1126/science.171.3968.303

Everaert, M. B. H., Huybregts, M. A. C., Chomsky, N., Berwick, R. C., and Bolhuis, Johan, J. (2015). Structures, not strings: linguistics as part of the cognitive sciences. *Trends Cogn. Sci.* 19, 729–743. doi: 10.1016/j.tics.2015.09.008

Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A. C., Pozzo, T., et al. (2009). Encoding of human action in Broca's area. *Brain* 132, 1980–1988. doi: 10.1093/brain/awp118

Fischer, S. (1996). The role of agreement and auxiliaries in sign language. *Lingua* 98, 103–119. doi: 10.1016/0024-3841(95)00034-8

Fischer, S. (2017). Crosslinguistic variation in sign language syntax. *Annu. Rev. Linguist.* 3, 1.1–1.23. doi: 10.1146/annurev-linguistics-011516-034150

Foerde, K., Knowlton, B. J., and Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11778–11783. doi: 10.1073/pnas.0602659103

Friederici, A. D., Bahlmann, J., Friedrich, R., and Makuuchi, M. (2011). The neural basis of recursion of complex syntactic hierarchy. *Biolinguistics* 5, 87–104.

Gallese, V. (2008). Mirror neurons and the social nature of language: the neural exploitation hypothesis. *Soc. Neurosci.* 3, 317–333. doi: 10.1080/17470910701563608

Goucha, T., and Friederici, A. D. (2015). The language skeleton after dissecting meaning: a functional segregation within Broca's Area. *Neuroimage* 114, 294–302. doi: 10.1016/j.neuroimage.2015.04.011

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

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

Hoff, E. (2006). How social contexts support and shape language development. *Dev. Rev* 26, 55–88. doi: 10.1016/j.dr.2005.11.002

Hwang, S. H., and Lardiere, D. (2013). Plural-marking in L2 Korean: a feature-based approach. *Second Lang. Res.* 29, 57–86. doi: 10.1177/0267658312461496

Ichida, Y. (2005). Bun koozoo to atama no ugoki: Nihon shuwa no bunpoo (6), gojun, hobun, kankei setu (Clause structures and head movements: the grammar of Japanese Sign Language (6), word orders, complement clauses and relative clauses). *Gekkan Gengo* 34, 91–99.

Jarvis, S., and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. London: Routledge.

Jeong, H., Sugiura, M., Sassa, Y., Wakusawa, K., Horie, K., Sato, S., et al. (2010). Learning second language vocabulary: neural dissociation of situation-based learning and text-based learning. *Neuroimage* 50, 802–809. doi: 10.1016/j.neuroimage.2009.12.038

Kimura, H. (2011). *Nihon Shuwa to Nihongo Taioo Shuwa (Shushi Nihongo): Aida ni aru "fukai tani" (Japanese Sign Language and Signed Japanese: A "deep gap" Between Them)*. Tokyo: Seikatsu Shoin.

Kirby, S. (2014). "Major theories in acquisition of syntax research," in *The Routledge Handbook of Syntax*, eds A. Carnie, Y. Sato, and D. Siddiqi (London; New York, NY: Routledge), 426–445.

Koechlin, E., and Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron* 50, 963–974. doi: 10.1016/j.neuron.2006.05.017

Krcmar, M., Grela, B., and Lin, K. (2007). Can toddlers learn vocabulary from television? An experimental approach. *Media Psychol.* 10, 41–63. doi: 10.1080/15213260701300931

Kroll, J. F., Bobb, S., and Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: arguments against a fixed locus of language selection in bilingual speech. *Bilingualism Lang. Cogn.* 9, 119–135. doi: 10.1017/S1366728906002483

Kuhl, P. (2007). Is speech learning 'gated' by the social brain? *Dev. Sci.* 10, 110–120. doi: 10.1111/j.1467-7687.2007.00572.x

Kuhl, P. K., Tsao, F. M., and Liu, H. M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Nat. Acad. Sci. U.S.A.* 100, 9096–9101. doi: 10.1073/pnas.1532872100

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10, 120–131. doi: 10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8

Lane, H. (1976). *The Wild Boy of Aveyron*. Cambridge, MA: Harvard University Press.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Lang. Res.* 25, 173–227. doi: 10.1177/0267658308100283

Macaluso, E., and Driver, J. (2003). Multimodal spatial representations in the human parietal cortex: evidence from functional imaging. *Adv. Neurol.* 93, 219–233.

MacSweeney, M., Capek, C. M., Campbell, R., and Woll, B. (2008). The signing brain: the neurobiology of sign language. *Trends. Cogn. Sci.* 12, 432–440. doi: 10.1016/j.tics.2008.07.010

Matsuoka, K. (2015). *Nihonshuwa de Manabu Shuwa Gengogaku no Kiso (Foundations of Sign Language Linguistics with a Special Reference to Japanese Sign Language)*. Tokyo: Kuroshio.

Morgan-Short, K., Steinhauer, K., Sanz, C., and Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *J. Cogn. Neurosci.* 24, 933–947. doi: 10.1162/jocn_a_00119

Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S. F., and Fazio, F. (2001). Syntax and the brain: disentangling grammar by selective anomalies. *Neuroimage* 13, 110–118. doi: 10.1006/nimg.2000.0668

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

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., and Garrett, M. F. (1991). Syntactically based sentence processing classes: evidence from event-related brain potentials. *J. Cogn. Neurosci.* 3, 151–165. doi: 10.1162/jocn.1991.3.2.151

O'Grady, W. (2005). *Syntactic Carpentry: An Emergentist Approach to Syntax*. Mahwah, NJ: Lawrence Erlbaum.

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

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., et al. (2008). Second-language learning and changes in the brain. *J. Neuroling.* 21, 509–521. doi: 10.1016/j.jneuroling.2008.01.001

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

Peigneux, P., Van der Linden, M., Garraux, G., Laureys, S., Degueldre, C., Aerts, J., et al. (2004). Imaging a cognitive model of apraxia: the neural substrate of gesture-specific cognitive processes. *Hum. Brain Mapp.* 21, 119–142. doi: 10.1002/hbm.10161

Perani, D., and Abutalebi, J. (2005). The neural basis of first and second language processing. *Curr. Opin. Neurobiol.* 15, 202–206. doi: 10.1016/j.conb.2005.03.007

Petitto, L. A., and Marentette, P. F. (1991). Babbling in the manual mode: evidence for the ontogeny of language. *Science* 251, 1493–1496. doi: 10.1126/science.2006424

Phillips, C. (2003). Linear order and constituency. *Linguist. Inq.* 34, 37–90. doi: 10.1162/002438903763255922

Piske, T., and Young-Scholten, M. (eds.) (2009). *Input Matters in SLA*. Bristol: Multilingual Matters.

Poldrack, R., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature* 414, 546–550. doi: 10.1038/35107080

Raizada, R. D. S., Richards, T. L., Meltzoff, A. N., and Kuhl, P. K. (2008). Socioeconomic status predicts hemispheric specialization of the left inferior frontal gyrus in young children. *Neuroimage* 40, 1392–1401. doi: 10.1016/j.neuroimage.2008.01.021

Roseberry, S., Hirsh-Pasek, K., and Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Chi. Dev.* 85, 956–970. doi: 10.1111/cdev.12166

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., and Golinkoff, R. M. (2009). Live action: can young children learn verbs from video? *Chi. Dev.* 80, 1360–1375. doi: 10.1111/j.1467-8624.2009.01338.x

Ross, J. R. (1967). *Constraints on Variables in Syntax*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Sachs, J., Bard, B., and Johnson, M. L. (1981). Language learning with restricted input: case studies of two hearing children of deaf parents. *Appl. Psycholinguist.* 2, 33–54. doi: 10.1017/S0142716400000643

Sakai, K. L., Nauchi, A., Tatsuno, Y., Hirano, K., Muraishi, Y., Kimura, M., et al. (2009). Distinct roles of left inferior frontal regions that explain individual differences in second language acquisition. *Hum. Brain. Mapp.* 30, 2440–2452. doi: 10.1002/hbm.20681

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

Sandler, W., and Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.

Santi, A., and Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *Neuroimage* 51, 1285–1293. doi: 10.1016/j.neuroimage.2010.03.034

Slabakova, R. (2016). *Second Language Acquisition*. London: Oxford University Press.

Slabakova, R., Leal, T. L., and Liskin-Gasparro, J. (2014). We have moved on: current concepts and positions in generative SLA. *Appl. Linguisti.* 35, 601–606. doi: 10.1093/applin/amu027

Talairach, J., and Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain*. New York, NY: Thieme.

Tettamanti, M., and Weniger, D. (2006). Broca's area: a supramodal hierarchical processor? *Cortex* 42, 491–494. doi: 10.1016/S0010-9452(08)70384-8

Uchibori, A., and Matsuoka, K. (2016). Split movement of wh-elements in Japanese Sign Language: a preliminary study. *Lingua* 183, 107–125. doi: 10.1016/j.lingua.2016.05.008

Verga, L., and Kotz, S. A. (2013). How relevant is social interaction in second language learning? *Front. Hum. Neurosci.* 7:550. doi: 10.3389/fnhum.2013.00550

Werker, J. F., and Tees, R. C. (1984). Cross language speech perception-evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63. doi: 10.1016/S0163-6383(84)80022-3

Yang, C. (2004). Universal grammar, statistics, or both? *Trends Cogn. Sci.* 8. 451–456. doi: 10.1016/j.tics.2004.08.006

Yusa, N. (2012). "Structure dependence in the brain, " in *Five Approaches to Language Evolution. Proceedings of the Workshops of the 9th International Conference on the Evolution of Language*, eds L. McCrohon, K. Fujita, R. Martin, K. Okanoya, R, Suzuki (Kyoto: Evolang9 Organizing Committee), 25–26.

Yusa, N., Koizumi, M., Kim, J., Kimura, N., Uchida, S., Yokoyama, S., et al. (2011). Second-language instinct and instruction effects: nature and nurture in second-language acquisition. *J. Cogn. Neurosci.* 23, 2716–2730. doi: 10.1162/jocn.2011.21607

Zaccarella, E., and Friederici, A. D. (2015). Merge in the human brain: a sub-region based functional investigation in the seft pars opercularis. *Front. Psychol.* 6:1818. doi: 10.3389/fpsyg.2015.01818

Zaccarella, E., Meyer, L., Makuuchi, M., and Friederici, A. D. (2017). Building by syntax: the neural basis of minimal linguistic structures. *Cereb. Cortex* 27, 411–421. doi: 10.1093/cercor/bhv234

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., et al. (2009). Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *Neuroimage* 46, 226–240. doi: 10.1016/j.neuroimage.2009.01.028

# A-Book: A Feedback-Based Adaptive System to Enhance Meta-Cognitive Skills during Reading

*Ernesto Guerra[1,2]\* and Guido Mellado[1]*

[1] Experimental Psycholinguistics Lab, Pontificia Universidad Católica de Chile, Villarrica, Chile, [2] Center for Intercultural and Indigenous Research, Villarrica, Chile

In the digital era, tech devices (hardware and software) are increasingly within hand's reach. Yet, implementing information and communication technologies for educational contexts that have robust and long-lasting effects on student learning outcomes is still a challenge. We propose that any such system must a) be theoretically motivated and designed to tackle specific cognitive skills (e.g., inference making) supporting a given cognitive task (e.g., reading comprehension) and b) must be able to identify and adapt to the user's profile. In the present study, we implemented a feedback-based adaptive system called A-book (assisted-reading book) and tested it in a sample of 4th, 5th, and 6th graders. To assess our hypotheses, we contrasted three experimental assisted-reading conditions; one that supported meta-cognitive skills and adapted to the user profile (adaptive condition), one that supported meta-cognitive skills but did not adapt to the user profile (training condition) and a control condition. The results provide initial support for our proposal; participants in the adaptive condition improved their accuracy scores on inference making questions over time, outperforming both the training and control groups. There was no evidence, however, of significant improvements on other tested meta-cognitive skills (i.e., text structure knowledge, comprehension monitoring). We discussed the practical implications of using the A-book for the enhancement of meta-cognitive skills in school contexts, as well as its current limitations and future developments that could improve the system.

Keywords: adaptive ICTs, meta-cognitive skills, reading comprehension

## INTRODUCTION

The advent of increasingly accessible and cheaper digital information and communication technologies (ICTs) has raised the question about their role in the context of formal education. ICTs are seen by some as a natural, intuitive, and easy-to-use tool for mediated learning, and there are some examples of their successful application in education (e.g., Roschelle et al., 2000; Thibaut et al., 2015). There are, nevertheless, detractors and critics of indiscriminate use of ICTs in school contexts (e.g., Buckingham, 2007) and also less successful examples of their application to classrooms (see, e.g., Kramarski and Feldman, 2000). Currently, it is still unclear what kind of technology is most appropriate to support children's learning and development (see Hermans et al., 2008). What is clear, however, is that only making ICTs available to schools does not guarantee a significant impact in student performance and learning processes (Cuban et al., 2001; Selwyn et al., 2009). The answer to this question might depend partly on whether the technology in use is

specifically designed to tackle relevant cognitive processes supporting specific capabilities (see, e.g., Blok et al., 2002).

In the context of language education, ICTs have been predominately used as an aid for low-level language skills, such as decoding (e.g., Barker and Torgesen, 1995; Mathes et al., 2001; Bonacina et al., 2015) and less to help students to improve text comprehension skills (see National Reading Panel, 2000). It is well-established that word decoding is critical for reading comprehension during primary school years (e.g., Perfetti and Hogaboam, 1975; Kendeou et al., 2009). However, it is also known that meta-cognitive strategies such as inference making, comprehension monitoring, and text structure knowledge are relevant for understanding written stories, in particular when children transition from learning to read to learning by reading (see Paris et al., 1983; Oakhill et al., 2014). How could new technologies be used most effectively to foster and enhance these skills?

A number of reviews suggest that there is virtually no evidence of the benefits that ICT could provide to reading comprehension during school years (see Torgerson and Elbourne, 2002; See and Gorard, 2014; Paul and Clarke, 2016). An existing study examined the effects of the use of a software that focused on reading and spelling (Brooks et al., 2006). The authors report that the software allowed students to hear and correct themselves and work independently at their pace and had different difficulty levels to which students were assigned based on prior assessment. Pre- and post-treatment test were administered to the experimental and the control group. Children in the ICT group undertook sessions of 1 h a day for 10 consecutive school days. Statistical comparison showed an advantage in the reading test for the control group compared to the ICT group, suggesting a negative effect of the use of the software. Similarly, a study by Khan and Gorard (2012) assessed the effectiveness of a computer program designed to improve reading. The ICT consisted in a multi-sensory software that combined touch, vision and sound; it provided more than 100 texts, immediate feedback and the difficulty of items could be adapted. User's progress was also recorded by the software. In the study, participant's literacy skills were assessed pre- and post-treatment. Student in the ICT condition used the software for 10 weeks, period over which the control group did not use the software. Statistical analysis showed that both the control and the ICT group improved literacy skills after the 10 weeks period. Again, however, the control group performed significantly better than the ICT group.

Other studies have also failed to demonstrate the usability of ICT (see Rouse and Krueger, 2004; Lei and Zhao, 2007; Given et al., 2008; Borman et al., 2009). In contrast, interventions delivered to groups of students directly by teachers seem to be much more successful in improving participants reading skills (see Berkeley et al., 2011; Vaughn et al., 2011; McMaster et al., 2012). In this context, it might be tempting to argue that ICT are not a suitable tool to foster and improve reading comprehension. An alternative view is that ICT must be designed and grounded theoretically, must tackle specific cognitive skills supporting reading comprehension (rather than providing an "enhanced" reading experience, see Khan and Gorard, 2012), and in

addition, they should be able to adapt online to the individual characteristics and performance of the student (see McMaster et al., 2012).

In this article, we present the implementation and a preliminary assessment of an automatized feedback-based system we called A-book (assisted-reading book), designed to provide theoretically motivated user-based feedback during the process of reading. The A-book's aim is to offer an adequate context for primary school students to develop meta-cognitive strategies at an early stage. In a between-subject design we contrasted three experimental assisted-reading conditions; a training condition, an adaptive condition and a control condition. In all three experimental conditions, readers were presented with stories (one page at the time) and three types of *yes*-or-*no* questions. Each of these questions related to a critical meta-cognitive ability, namely, inference making, comprehension monitoring and text structure knowledge. Our working hypothesis was that pertinent feedback (on inference, monitoring, and structure) should prompt the young reader to begin to strategically apply these skills while reading (i.e., training and adaptive conditions) in particular when the system adapts to the users' profile purposefully focusing on meliorating her weaknesses (i.e., adaptive condition). Consequently, we predicted that both the training and adaptive conditions would in time produce better comprehension accuracy scores relative to the control condition. Furthermore, we hypothesized that the adaptive behavior of the system should benefit the user comprehension processes. Thus, children using the adaptive condition of the system should outperform the training condition group.

## MATERIALS AND METHODS

### Participants
Ninety primary school students from 4th, 5th, and 6th grade (aged between 9 and 12 years) from a local school, who participated voluntarily on a session basis, were recruited to take part of the study. All children were monolingual Spanish native speakers.

### Materials
#### Reading Materials
We selected 10 stories from "Un cuento al día – Antología" (Consejo Nacional del Libro y la Lectura, 2013), a book published by the Consejo Nacional de la Cultura y las Artes (National Board for Culture and the Arts), Government of Chile. This book was made freely available in 2013, in the context of the Plan Nacional de Fomento de la Lectura (National Plan for the Reinforcement of Reading) "Lee Chile Lee," and it was aimed to promote parental reading as a daily activity[1]. Each story was divided in a number of fragments ($m = 12.7$; range: 9–27) of about 100 words each ($m = 124$; range: 70–200) for their presentation. For each fragment, we wrote three questions, each of them

---

[1]Only one participant in the sample manifested he was familiar with one of the stories he read. No other student mentioned having previously read or heard the stories.

related to a specific meta-cognitive skill (i.e., inference making, comprehension monitoring and text structure knowledge; see the Supplementary Material for some examples). For each question, we then wrote two kinds of feedback (i.e., explanatory and control) and each of them in two equivalent versions (one for correct answers and one for incorrect answers).

Questions aimed to capture the readers' ability to make inferences were constructed to ask about information that was not explicitly given in the text. For instance, if the text said 'the old man mixed the content of jars to make rain. . .,' we asked whether 'the old man knew a recipe for rain.' The explanatory feedback alluded to such critical information, by stating for instance, 'if the old man mixed the content of jars to make rain, he most probably knew a recipe for rain.' Questions about the structure of the text directly asked the reader whether the story's characters had been already presented, whether they already knew the scenario or context in which the story was taken place, or whether the story was about to end, or only at the beginning. Such questions did not include any content of stories, in other words they were story independent. The corresponding feedback was also story independent insofar it just reminded the reader the linkage between characters, scenario, story conflict and conflict resolution, and the structure of the text (e.g., 'Exactly, if you are beginning to know the characters, the story is just starting'; 'Hmmm, I am not sure. . . if you are beginning to know the characters, the story is just starting').

Finally, questions intended to measure participants' comprehension monitoring, were always built as 'Did you realized that. . .' and the sentence was completed by a literal phrase of the text or one slightly modified for *No*-answers. Feedback for correct *Yes*-answers consisted on a reinforcement sentence that referred to the concentration of the reader ('Very good! It is clear that you are very concentrated'), while the feedback for correct answers *No*-answers always began by saying: 'Of course not, because that never happened.' and ended with the reinforcement sentence. When the readers responded incorrectly, the explanatory feedback consisted on a re-iteration of the text cited from the text plus the mentioning of the importance of paying attention to what one understands, as well as to what one does not clearly understand.

### Assisted-Reading Experimental Conditions

Our study contrasted three assisted-reading experimental conditions, two of them with explanatory feedback (i.e., training and adaptive experimental conditions) and one control condition. In the training and adaptive experimental conditions, the feedback the readers received after responding a question was aimed to encourage them to reflect and think over the question and the answer they chose, independently of whether the given answer was correct or not. Thus, the training/adaptive feedback provided the reader with an explanation for the answer of the question, while at the same time reassured readers when answering correctly (e.g., "Very good. . .," "Well done...") and inviting reconsideration (e.g., "Hmm, I am not so sure, perhaps. . . [. . .] Don't you think?"), when the response was incorrect (instead of penalizing it). The logic behind inviting reconsideration was to keep both training- and adaptive-feedback

explanatory in nature. In other words, the feedback should point to the relevant information necessary to answer the question accurately, independently of whether the actual answer of the participant was correct or incorrect.

The critical difference between the training and adaptive assisted-reading experimental conditions was the way in which the selection of the meta-cognitive skill, in other words the type of question (i.e., inference making, story structure, or comprehension monitoring), was presented to the reader in a particular moment. Presentation of questions in the training condition was counterbalanced: for each story, participants read the same number of questions on each meta-cognitive skill and their presentation was pseudo-randomized. The adaptive experimental condition, instead, selected the weakest meta-cognitive skill at the user level and prioritized its presentation. For this reason, the adaptive experimental condition required an individual profile as a starting point to work, and such data was obtained from participants' session using the training condition (see Design). It also joined text fragments, first two, then three, for readers with accuracy higher than 75% (two fragments) and 85% (three fragments) in all meta-cognitive skills. The control condition also presented counterbalanced question types but the feedback consisted only in the word "Correct" or "Incorrect," depending on whether the answer given by the reader was correct or not.

### Design

We constructed a website we used for presentation and management of the stories and data. All the stories, questions and feedback were presented one at the time on the screen. In such way, the readers could concentrate on a single task at the time (e.g., reading text, answering the question, reading feedback). Participants were given six 30-min reading sessions in two blocks (three sessions on each block) over a 2 weeks period (see **Table 1**). The first three sessions (week 1) formed the Exposure Block, while the subsequent three sessions (week 2) constituted the Testing Block. In the first block (week 1), the full sample of participants was randomly divided into two groups; one with approximately twice as many participants as the other ($n = 34$ and $n = 58$). The smaller group was assigned to the control condition, while the larger group was assigned to the training condition. During this block, participants were presented with five stories, each of which had nine fragments to be read, and 27 questions to be answer (nine on each experimental condition, i.e., inference making, text structure, and comprehension monitoring).

The exposure session was aimed to familiarized participants with the task before comparing the effects of different reading conditions. In this sense, participants should have an experience of continuity between the two blocks. Thus, we presented the control condition to a third of the participants and the training condition to the other two thirds; participants in the control condition worked on that condition across blocks. Moving from the training to the adaptive should be unnoticed; in fact, participants would just have more questions of that task that is difficult for them. In contrast, if we would have for instance presented everyone with the control condition in the exposure

**TABLE 1 | Design of the study.**

| Week 1 (Exposure) | | | Week 2 (Assessment) | | |
|---|---|---|---|---|---|
| Session 1 | Session 2 | Session 3 | Session 4 | Session 5 | Session 6 |
| 10–15 participants per session | | | 10–15 participants per session | | |
| Experimental conditions: Training \| Control | | | Experimental conditions: Training \| Control \| Adaptive | | |

block, the second block would have meant a different context for those participants in the training and adaptive conditions. In our view, this could have created a disadvantage for those conditions.

In the second block, 30 of the participants in the training condition stayed on that condition, while the other 28 were assigned to the adaptive condition. This design allowed the correct functioning of the adaptive condition, feeding readers' data in the second block from the data collected at the participant level during the first block in the training condition. In the second block, participants were presented with five new stories. The number of fragment per stories varied between 9 and 27 and for each fragment there were always three potential questions corresponding to each meta-cognitive skill.

## Procedure

On each reading session, the students were invited to participate voluntarily. No personal data from students was recorded and all of them participated voluntarily on a session basis. **Figure 1** shows a schematic presentation of the sequence that participants would see when performing the task. Every student was assigned a username (previously created) to enter the website and work individually on a computer. The number of participants per session varied between 10 and 15 students at the time. As participants logged in the website, they were first presented with written instructions. In these instructions, they were informed that they would read the stories fragment by fragment and receive questions about them. It also included words of encouragement such as 'We want to invite you to read… but we want your reading to be as fun as possible.' After reading the instructions, participants clicked on a 'Next'-button and were presented with five pictures, each of them representing one story.



**FIGURE 1 | A schematic representation of the reading task.**

Students could read the titles and select any out of five stories to read. When they clicked on a picture they were presented with the title and first fragment of the story they selected. There was no time constraint and participants could read the story fragments at their own pace. When participants clicked on a 'Next'-button, they were presented with a written question, plus a 'Yes'- and a 'No'-button. After they gave their response, the response feedback was presented on the screen plus a 'Next'-button. When they pressed 'Next' a new fragment was presented, and the loop of fragment-question-feedback continued until the end of the story. When the story ended, participants read a message stating that they had finished the story and encouraged them to read other stories or even the same one again if they wanted. However, the icon of the most recently read story temporally disappeared and did appear again only after a different story was fully read. When an already-read story was chosen by the participant, the system would assign a new question for each fragment until there were no more questions to be answered. This meant that each story could be fully read up to three times, after which the icon for the story disappeared from the story selection screen.

The procedure was approved by the Ethics Committee of the Campus Villarrica. All activities were performed in the school dependencies and during regular school hours as a complementary informatics and language activity. The participation of the students was approved by the Principal of the school and the head of the Technical Pedagogical Unit as the legally authorized representatives.

## Data analysis

Before inferential analysis, we examined individual participants' responses and decided to exclude four participants since they gave only 'Yes' responses. All other data were included in the analysis. Our basic dependent variable was participants' accuracy, but we were also interested in seeing how such accuracy developed in time and, particularly, as a function of the different assisted-reading experimental conditions. One clear candidate variable to evaluate such effect in time was the number of responses at the participants' level. The more questions they responded, the more experience they are supposed to gain. If there are any differences in the effect of such experience on participants' accuracy as a function of the assisted-reading experimental condition, we should observe them by contrasting the three experimental conditions across time (as reflected by the number of responses). However, due to the nature of the task, data were strongly unbalanced in the number of responses per participant, per condition and per meta-cognitive skill. The adaptive condition exhibits much
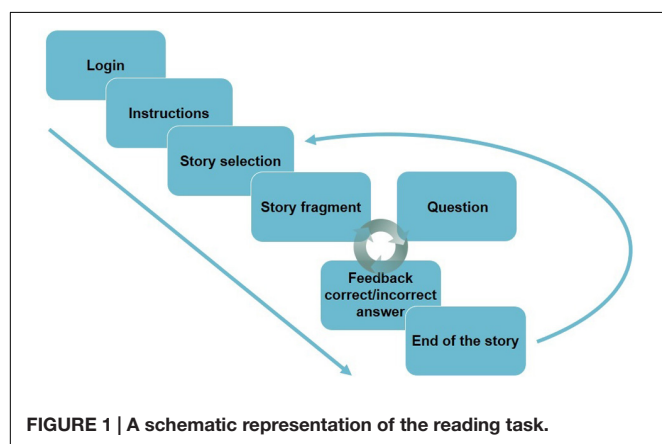
**TABLE 2 | Descriptive statistics per experimental condition and meta-cognitive skill.**
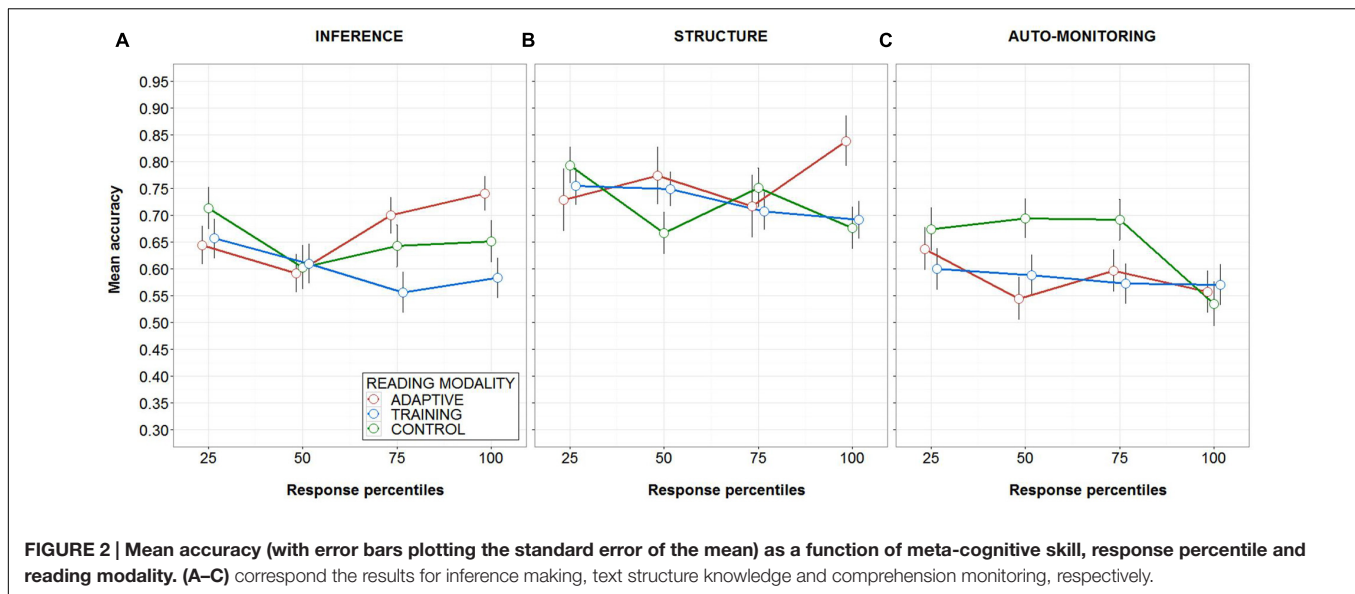
| Experimental condition | Meta-cognitive skill | Percentile | Cutting value | *n* | Cumulative % |
|---|---|---|---|---|---|
| Adaptive | Inference | 25 | 20 | 177 | 24 |
| | | 50 | 47.5 | 191 | 50 |
| | | 75 | 93 | 183 | 74.9 |
| | | 100 | | 185 | 100 |
| | | *Total* | | *736* | |
| | Comprehension monitoring | 25 | 12 | 146 | 23.6 |
| | | 50 | 28 | 158 | 49.2 |
| | | 75 | 48 | 156 | 74.4 |
| | | 100 | | 158 | 100 |
| | | *Total* | | *618* | |
| | Text structure | 25 | 21 | 59 | 24.3 |
| | | 50 | 51 | 62 | 49.8 |
| | | 75 | 81 | 60 | 74.5 |
| | | 100 | | 62 | 100 |
| | | *Total* | | *243* | |
| Control | Inference | 25 | 15 | 136 | 23.4 |
| | | 50 | 33 | 146 | 48.5 |
| | | 75 | 59 | 151 | 74.4 |
| | | 100 | | 149 | 100 |
| | | *Total* | | *582* | |
| | Comprehension monitoring | 25 | 15 | 135 | 23.1 |
| | | 50 | 34.5 | 157 | 50 |
| | | 75 | 59.75 | 146 | 75 |
| | | 100 | | 146 | 100 |
| | | *Total* | | *584* | |
| | Text structure | 25 | 15 | 145 | 24.9 |
| | | 50 | 34 | 147 | 50.2 |
| | | 75 | 59.25 | 145 | 75.1 |
| | | 100 | | 145 | 100 |
| | | *Total* | | *582* | |
| Training | Inference | 25 | 21 | 172 | 24.9 |
| | | 50 | 47.5 | 174 | 50 |
| | | 75 | 83 | 171 | 74.7 |
| | | 100 | | 175 | 100 |
| | | *Total* | | *692* | |
| | Comprehension monitoring | 25 | 22 | 170 | 24.8 |
| | | 50 | 47 | 170 | 49.6 |
| | | 75 | 82 | 173 | 74.9 |
| | | 100 | | 172 | 100 |
| | | *Total* | | *685* | |
| | Text structure | 25 | 21 | 163 | 23.6 |
| | | 50 | 47.5 | 183 | 50 |
| | | 75 | 83 | 171 | 74.7 |
| | | 100 | | 175 | 100 |
| | | *Total* | | *692* | |

less questions of text structure relative to inferences and comprehension monitoring, while the control condition exhibits overall much less questions per participant (maximum number of questions = 142, compared to 208 and 243 for training and adaptive conditions, respectively).

Consequently, we decided to group the number of questions based on four quartiles per experimental condition and

meta-cognitive skill. In doing so, we found a principled way to obtain a more balanced data set for comparison. **Table 2** shows the cutting values that divided the percentile groups per condition and skill. It also shows the number of cases per group and the cumulative percentage this number meant for the total.

This grouping led to a more balanced data set for comparison, which we subsequently compared using a generalized linear

**FIGURE 2 | Mean accuracy (with error bars plotting the standard error of the mean) as a function of meta-cognitive skill, response percentile and reading modality. (A–C)** correspond the results for inference making, text structure knowledge and comprehension monitoring, respectively.

mixed model approach, henceforth GLMM (lmerTest Package in R; see Baayen et al., 2008). GLMM are particularly suitable for the analysis of binomial data since they offer a sufficiently conservative, yet balanced approach for accuracy analysis (see Hosmer and Lemeshow, 2000; Quené and van den Bergh, 2004, 2008). GLMM allows a multilevel analysis with crossed random factors (e.g., participants) while accommodating such intrinsic variation around the fixed factors and their interaction. These models have less assumption than classic ANOVAs (do not assume homoscedasticity or sphericity of the data), do not require data aggregation and are more robust against unbalanced data and missing values (Quené and van den Bergh, 2004, 2008; Baayen et al., 2008; Barr, 2008). Their output delivers estimates, standard errors, $z$- and $p$-values.

We contrasted the effects of assisted-reading experimental conditions for each meta-cognitive skill (i.e., inference making, text structure, and comprehension monitoring). To minimize collinearity between fixed factors, we centered the predictors' values on a mean of 0 before analysis, using a scale function (base Package in R). The models[2] included, as fixed effects, the assisted-reading condition and the response group (quartiles) and their interaction. They also included a random intercept for participants and fixed effects and interaction random slopes for the participant random intercept. To simplify the model and improve convergence, we did not include random correlations between predictors and the random intercept (see Barr et al., 2013 for such recommendation).

## RESULTS

The results from the GLMM for the inference experimental condition showed a main effect of condition ($\beta = -0.17$,

---

[2]R-code: glmer (accuracy ~ condition : group + condition + group + (0 + condition | participant) + (0 + group | participant) + ( 0 + condition: group | participant) + (1 | participant), data, family = binomial).

$t = -2.04$, $p < 0.05$) but no main effect of response percentile group. More importantly, it evidenced a reliable interaction effect between the condition and response group ($\beta = -0.15$, $t = -2.82$, $p < 0.01$). **Figure 2A**, shows a graphic representation of the observed interaction pattern. Accuracy remained relatively similar for all experimental conditions within the first two quartiles, yet from the third quartile on, a distinctive pattern for each condition emerged: accuracy in the adaptive condition increased to 0.69 [CI95% $\pm$ 0.7] and then to 0.74 [CI95% $\pm$ 0.6] in the third and fourth quartile respectively, compared to the other conditions that remained at 0.64 [CI95% $\pm$ 0.8] and 0.65 [CI95% $\pm$ 0.8] for control and 0.55 [CI95% $\pm$ 0.8] and 0.58 [CI95% $\pm$ 0.8] for the training condition in the same quartiles.

In contrast to the results for the inference experimental condition, the outcome of the GLMM for the text structure experimental condition showed no main effect of condition or group, neither showed interaction between predictors ($t$-values < |2|). **Figure 2B** illustrates the mean accuracy pattern across the response groups, showing no clear differences between conditions across time. Finally, the GLMM from the comprehension monitoring experimental condition detected a main effect of response percentile ($\beta = -0.16$, $t = -2.034$, $p < 0.05$). As it is shown in **Figure 2C**, there is a tendency of a decrease in accuracy, in particular for the control condition. Such effect, however, was not modified by the condition as reflected by the absence of the predicted interaction between assisted-reading condition and the response percentile groups. **Table 3** summarizes the results of the GLMMs analysis.

## DISCUSSION

The current study constitutes a proof of concept for the following hypothesis: the effective use of ICTs in learning contexts depend

on whether this technology is designed to enhance and support specific cognitive skills that underlie specific cognitive tasks. We proposed that any effective ICT system must be designed to provide a theoretically motivated context for learning and that such system must have the ability to adapt to the user's profile. We chose to investigate this hypothesis in the context of text comprehension in primary school students since most studies that used ICTs in the context of language instruction either focused on basic language skills (e.g., decoding), and those that concentrated in more high-level skills did so in older readers.

Consequently, we designed and implemented a web platform that presented 4th, 5th, and 6th graders with a set of A-books, questions about them, and corresponding response feedback. Critically, we contrasted three assisted-reading experimental conditions to investigate our hypothesis, namely (1) a condition that supported specific cognitive skills (i.e., meta-cognitive abilities) and that adapted to the users' profile (adaptive condition), (2) a condition that supported the same specific cognitive skills but did not adapt to the users' profile (training condition), and (3) a control condition that did not support cognitive skills nor adapted to the users' profile.

Participants read stories in one of these three different experimental conditions, while we measured their accuracy on each meta-cognitive ability question. We predicted that the adaptive condition would produce over time better accuracy scores compared to the training and the control experimental conditions, and that the training condition would surpass the results of the control condition. Indeed, the analysis of the inference experimental condition showed a reliable advantage for the adaptive condition relative to both the training and the control experimental conditions. However, the participants in the control condition performed better relative to those in the training condition (see **Figure 2A**). Moreover, analysis of the accuracy for the text structure and comprehension monitoring

did not reveal such advantage for the adaptive condition. In the text structure experimental condition, we observed a late advantage for the adaptive condition (see **Figure 2B**), which could be interpreted in favor of our predictions. This advantage was nevertheless not strong enough to bring about an interaction effect between condition and response percentile group that was statistically reliable. Finally, the results observed in the comprehension monitoring experimental condition are the most puzzling ones; we observed a reliable main effect that suggests that readers' performance deteriorated over time. The accuracy pattern, however, suggests that effect is carried by the control condition, which evidence a drop of around 15% in the fourth quartile relative to the first three quartiles.

Taken together, the present results provide support to our prediction: when readers were presented with adequate and personalized context for the support of specific cognitive skills (i.e., explanatory feedback) needed to performed specific cognitive tasks (i.e., inference making), their accuracy increased over time. Yet, there are a number of issues worth addressing with regards to the data pattern observed in the study. First, the results pattern suggests that for the inference experimental condition as for the comprehension monitoring experimental condition, readers in the control condition overcame the performance of the readers in the training condition. This unexpected result might find an explanation on the literature about students' self-regulation and behavior modulation (Lemos, 1999; Nilson, 2013). According to Lemos, self-regulated students are better in delaying the immediate reward after a task in order to achieve more important goals. Moreover, the author suggests that self-regulation capacities are based on the assimilation of values and incentives. On the other hand, it has been suggested that the Chilean educational system teaching style is predominantly oriented to results rather than the process of learning and reflection. This, in part, as a negative consequence of the use of school rankings based on standardized evaluations as a measure of quality of education. Ortiz (2012), for example, explains that the use of these measures and rankings do not provide significant guidance for teachers to implement specific pedagogical actions, and yet among teachers (and principals) there is an increasing tendency to consider deficient results as an indicator for the need of implementation of action to improve students learning. This blind-spot (i.e., knowing that something needs to be fixed, but not knowing exactly what and how to fix it), produces many unwanted practices in schools such as the exclusion and selection of students (Ortiz, 2012; see also Flórez Petour, 2015). Interestingly, Lemos (1999) proposes that the social context can lead children to believe they are not capable to achieve expected outcomes, and that those children tend to respond maladaptatively. Instead, children that see challenges as more achievable are more likely to act more constructively. Taken together, this evidence may explain a tendency of students to respond better to short and uninformative feedback compared to the explanatory feedback. An alternative explanation, however, might be that these differences arise from the weakness of the between-subject design, view that would weaken our results. Although readers were randomly assigned to the different experimental conditions, groups were not matched in any

**TABLE 3 | Main and interaction effects in the GLMM by meta-cognitive skill.**

| Fixed effects: | β | SE | z-value | Pr(>|z|) |
|---|---|---|---|---|
| **(A) Inference making** | | | | |
| (Intercept) | 0.628 | 0.080 | 7.871 | 0.000 |
| Condition | −0.171 | 0.083 | −2.048 | 0.041 |
| Group | −0.001 | 0.053 | −0.013 | 0.989 |
| Condition*group | −0.153 | 0.054 | −2.823 | 0.005 |
| **(B) Text structure** | | | | |
| (Intercept) | 1.052 | 0.092 | 11.418 | 0.000 |
| Condition | −0.037 | 0.103 | −0.362 | 0.718 |
| Group | −0.079 | 0.067 | −1.178 | 0.239 |
| Condition*group | −0.075 | 0.065 | −1.153 | 0.249 |
| **(C) Comprehension monitoring** | | | | |
| (Intercept) | 0.484 | 0.097 | 4.997 | 0.000 |
| Condition | 0.008 | 0.102 | 0.081 | 0.936 |
| Group | −0.165 | 0.081 | −2.034 | 0.042 |
| Condition*group | 0.073 | 0.080 | 0.921 | 0.357 |

parameter. Future research could address this issue by testing the platform in a within-subject study.

A second issue relates to the overall pattern observed for the comprehension monitoring experimental condition. The statistical analysis suggests a general tendency for a decrease in reader accuracy, in other words, the opposite to the intended effect. A potential explanation for the failure of this experimental condition could be that the questions intended to evaluate the comprehension monitoring of the comprehension process were not able to capture this meta-cognitive skill properly. Previous research has made used of the insertion of errors in the text. For instance, inconsistencies within the same paragraph were deliberately included to assess whether the reader was paying attention to the content of the story (e.g., Markman, 1979; Tunmer et al., 1983; Oakhill et al., 2005). We followed a similar logic by presenting error-free literal citation and citation with intended error. We moreover wrote the cueing phrase 'Did you realize that. . .' before each of this type of questions. However, to keep the text the same for inference making, text structure and comprehension monitoring questions, we inserted the errors after the fragment and within the question. In this context, the questions might have been too demanding and its answering logic hard to understand influencing negatively readers' performance overtime.

This discussion leads us to a final issue of the present study. The results (in particular from the comprehension monitoring skill) rise the question on whether our intervention can produce significant changes in different meta-cognitive skills, or such improvements are limited to a specific meta-cognitive skill, in this case, the capacity of the readers to make offline inferences about the text they are reading. Our findings speak against this possibility; however, there are some attenuating points that might prompt a more optimistic view. As we argue above, the comprehension monitoring questions were most probably not the most adequate ones. Moreover, recent research has also been unsuccessful in finding improvement in comprehension monitoring both in short and long term (see Potocki et al., 2013), perhaps because this skill is much harder to foster and enhance.

With regard to the other two assessed meta-cognitive skills, participants were from the beginning less accurate on inference questions, relative to their performance on questions about the structure of the text (see **Figure 2**). This meant that they received overall much more reinforcement in the inference condition relative to text structure condition (2010 vs. 1517 questions, respectively), particularly in the adaptive condition (736 vs. 243 questions). There is nevertheless a (non-significant) trend of improvement in the last quartile in the text structure skill (see **Figure 2B**). Considering this (namely, the amount of reinforcement received) and the time readers spent using the system (only 1 week), the significant improvement observed at least for the most reinforced ability seems promising.
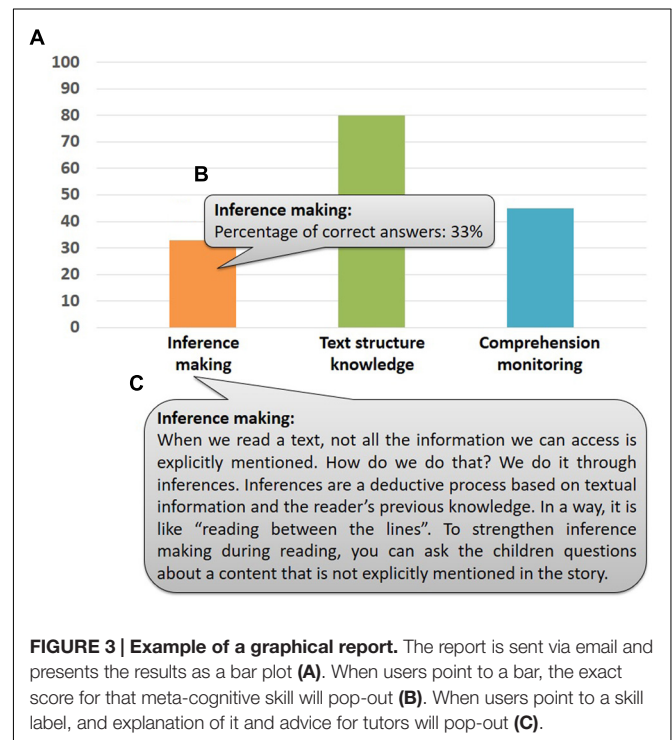
Without underestimating the caveats above-discussed, the results of the present study can be taken as evidence of the benefit of designing theoretically motivated (and empirically testing) ICTs interventions for educational contexts. They show that a system (and perhaps any kind of instruction) that can adapt to the user's profile is more effective compared to those that are less

flexible in the assignment of a task. Such principles are not new in the context of school teaching (e.g., Keller, 1968; Fuller, 1970), yet they have not permeated into the design and implementation of ICTs for school context (see c.f. Roschelle et al., 2000; Hammond, 2014).

## Practical Implications

One challenge for personalized teaching is avoiding the overt separation of students in different groups in the classroom or in different classrooms (see Ainscow and Miles, 2008). In this sense, the present tool allows the distinctive treatment of students in an implicit manner, that is, student do not need to be explicitly classified in groups of different achievement and being separated physically in the classroom or in different classrooms. The A-book can adapt to the user profiles even when apparently, all students are doing the same task.

Another practical implication is the potential use of the A-book as soft-assessment tool and guidance for teachers and parents. Students' accuracy data are recorded online at the individual level. Every time a user reads a full story, an updated graphical profile is send automatically to the email address with which the user was created. In the present study, as testing phase of the system, we created all profiles prior testing and thus, we received all graphical reports. However, the basic idea is that teachers or parents create the children's account and receive their progressive profiles. **Figure 3** shows an example of an individual report (**Figure 3A**) sent via email. This graphical information is accompanied by the exact score (by locating the mouse cursor on any bar, see **Figure 3B**) and explanations and advice (by locating



**FIGURE 3 | Example of a graphical report.** The report is sent via email and presents the results as a bar plot **(A)**. When users point to a bar, the exact score for that meta-cognitive skill will pop-out **(B)**. When users point to a skill label, and explanation of it and advice for tutors will pop-out **(C)**.

the mouse cursor on any of the three meta-cognitive skills icons, see **Figure 3C**) for teachers and parents.

## Limitations and Future Directions

Indeed, the present study has limitations. Among them, we identify four, which we think can be corrected relatively easily and would mean a significant improvement for the system: first, the type of questions; second, adding more texts and of more diverse genres; third, adding a reaction time measure and minimal time for continuation; finally, making the A-book a multimodal platform. We are aware that closed-questions are not the best way to assess reading comprehension (c.f. Oakhill et al., 2014). However, we also recognize that more multifaceted questions, such as open questions, present a more complex scenario for analysis, and might demand specific training for correction. Keeping in mind that our study presents a proof of concept of an adaptive assisted-reading book that could be easily made available online, we opted for the simplest version of the answers. Knowing this is a limitation, a next step in the development of the A-book is to implement richer questions (i.e., multiple choice and content answer, true or false, completion) that can better capture the skills at stake. In connection, it seems clear that adding other text genres would allow using a more varied set of questions. Adding a larger set of texts would also allow the use of the system for a more extensive period. We observed improvement after a week, which encourages the evaluation of the A-book's potential effect after a more prolonged usage.

Furthermore, in this first version of the A-book, we did not include a measure that could tell us how long students took to read each fragment (i.e., a reaction time measure), losing relevant data for exclusion of cases as well as for behavioral analysis. In this sense, including minimal time for continuation (e.g., calculated as 250 ms per word) would also mean an improvement. Finally, making the A-book a multimodal platform would make it not only more attractive for children and thus more likely to engage them in reading, but would also provide a much richer context, situating language within visual and auditory representations. Specifically, the insertion of illustrations accompanying text might allow the reader to construct a richer situation model of the narrative (see Arizpe and Styles, 2002 for a discussion on picturebooks), and the addition of audio-based feedback would guarantee that all students process the intended pointer and might also constitute a significant aid, in particular for less skilled readers (see Montali and Lewandowski, 1996).

## CONCLUSION

The present research started from the assumption that the interaction with the environment is of most relevance for the acquisition of language competencies (see Gee, 2004), without forgetting that individual differences are also critical for learning (Stanovich, 1986). Poor comprehension affects many children in primary school (Cornoldi and Oakhill, 1996) however; there is a variety of underlying reasons for such deficit (e.g., garden-variety, see, Nation and Snowling, 1998). Children might have strengths in one skill but deficits in others; they might be already skilled meta-cognitive readers and interruptions might disrupt their comprehension; they might as well have weaknesses in all three skills above described. We propose that any effective systems must be designed to provide a theoretically motivated context for learning and must have the ability to adapt to the user's profile. The results presented here are in coherence with our claims and future work should be able to clarify some of the open questions stated in the present paper.

## AUTHOR CONTRIBUTIONS

EG and GM developed the study concept, the experimental designed, prepared the materials and collected the data. GM implemented the website used for data collection. EG analyzed the data, interpreted the results and wrote the manuscript.

## ACKNOWLEDGMENTS

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2017.00098/full#supplementary-material

## REFERENCES

Ainscow, M., and Miles, S. (2008). Making education for all inclusive: where next? *Prospects* 38, 15–34. doi: 10.1007/s11125-008-9055-0

Arizpe, E., and Styles, M. (2002). *Children Reading Pictures: Interpreting Visual Texts*. Abingdon: Routledge.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Barker, T. A., and Torgesen, J. K. (1995). An evaluation of computer assisted instruction in phonological awareness with below average readers. *J. Educ. Comput. Res.* 13, 89–103. doi: 10.2190/TH3M-BTP7-JEJ5-JFNJ

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *J. Mem. Lang.* 59, 457–474. doi: 10.1016/j.jml.2007.09.002

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Berkeley, S., Mastropieri, M. A., and Scruggs, T. E. (2011). Reading comprehension strategy instruction and attribution retraining for secondary students with learning and other mild disabilities. *J. Learn. Disabil.* 44, 18–32. doi: 10.1177/0022219410371677

Blok, H., Oostdam, R., Otter, M. E., and Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: a review. *Rev. Educ. Res.* 72, 101–130. doi: 10.3102/00346543072001101

Bonacina, S., Cancer, A., Lanzi, P. L., Lorusso, M. L., and Antonietti, A. (2015). Improving reading skills in students with dyslexia: the efficacy of a sublexical training with rhythmic background. *Front. Psychol.* 6:1510. doi: 10.3389/fpsyg. 2015.01510

Borman, G. D., Benson, J. G., and Overman, L. (2009). A randomized field trial of the Fast ForWord Language computer-based training program. *Educ. Eval. Policy Anal.* 31, 82–106. doi: 10.3102/0162373708328519

Brooks, G., Miles, J. N. V., Torgerson, C. J., and Torgerson, D. J. (2006). Is an intervention using computer software effective in literacy learning? A randomised controlled trial. *Educ. Stud.* 32, 133–143. doi: 10.1080/03055690500416116

Buckingham, D. (2007). *Beyond Technology: Children's Learning in the Age of Digital Culture.* Cambridge: Polity Press.

Consejo Nacional del Libro y la Lectura (2013). *Un Cuento al día. de Soledad Camponovo Llanos (CNCA) Sitio.* Available at: http://www.cultura.gob.cl/wp-content/uploads/2014/01/un-cuento-al-dia-antologia.pdf [accessed May 4, 2016].

Cornoldi, C., and Oakhill, J. V. (1996). *Reading Comprehension Difficulties: Processes and Intervention.* Mahwah, NJ: Lawrence Erlbaum Inc.

Cuban, L., Kirkpatrick, H., and Peck, C. (2001). High access and low use of technologies in high school classrooms: explaining an apparent paradox. *Am. Educ. Res. J.* 38, 813–834. doi: 10.3102/00028312038004813

Flórez Petour, M. T. (2015). Validez y equidad en la medición escolar: El caso de SIMCE. *Psicoperspectivas* 14, 31–44.

Fuller, F. F. (1970). *Personalized Education for Teachers. An Introduction for Teacher Educators.* Austin, TX : Research and Development Center for Teacher Education.

Gee, J. P. (2004). *Situated Language and Learning: A Critique of Traditional Schooling.* Abingdon: Routledge.

Given, B. K., Wasserman, J. D., Chari, S. A., Beattie, K., and Eden, G. F. (2008). A randomized, controlled study of computer-based intervention in middle school struggling readers. *Brain Lang.* 106, 83–97. doi: 10.1016/j.bandl.2007.12.001

Hammond, M. (2014). Introducing ICT in schools in England: rationale and consequences. *Br. J. Educ. Technol.* 45, 191–201. doi: 10.1111/bjet.12033

Hermans, R., Tondeur, J., van Braak, J., and Valcke, M. (2008). The impact of primary school teachers' educational beliefs on the classroom use of computers. *Comput. Educ.* 51, 1499–1509. doi: 10.1016/j.compedu.2008.02.001

Hosmer, D. W., and Lemeshow, S. (2000). *Multiple Logistic Regression.* Hoboken, NJ: John Wiley & Sons, Inc. doi: 10.1002/0471722146.ch2

Keller, F. S. (1968). Good-bye, teacher. *J. Appl. Behav. Anal.* 1, 79–89. doi: 10.1901/jaba.1968.1-79

Kendeou, P., Van den Broek, P., White, M. J., and Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: the independent contributions of oral language and decoding skills. *J. Educ. Psychol.* 101, 765–778. doi: 10.1037/a0015956

Khan, M. A., and Gorard, S. (2012). A randomised controlled trial of the use of a piece of commercial software for the acquisition of reading skills. *Educ. Rev.* 64, 21–35. doi: 10.1080/00131911.2010.537315

Kramarski, B., and Feldman, Y. (2000). Internet in the classroom: effects on reading comprehension, motivation and meta-cognitive awareness. *Educ. Media Int.* 37, 149–155. doi: 10.1080/09523980050184709

Lei, J., and Zhao, Y. (2007). Technology uses and student achievement: a longitudinal study. *Comput. Educ.* 49, 284–296. doi: 10.1016/j.compedu.2005.06.013

Lemos, M. S. (1999). Students' goals and self-regulation in the classroom. *Int. J. Educ. Res.* 31, 471–485. doi: 10.1016/S0883-0355(99)00016-6

Markman, E. M. (1979). Realizing that you don't understand: elementary school children's awareness of inconsistencies. *Child Dev.* 50, 643–655. doi: 10.2307/1128929

Mathes, P. G., Torgesen, J. K., and Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *Am. Educ. Res. J.* 38, 371–410. doi: 10.3102/00028312038002371

McMaster, K. L., Van den Broek, P., Espin, C. A., White, M. J., Rapp, D. N., Kendeou, P., et al. (2012). Making the right connections: differential effects of

reading intervention for subgroups of comprehenders. *Learn. Individ. Differ.* 22, 100–111. doi: 10.1016/j.lindif.2011.11.017

Montali, J., and Lewandowski, L. (1996). Bimodal reading: benefits of a talking computer for average and less skilled readers. *J. Learn. Disabil.* 29, 271–279. doi: 10.1177/002221949602900305

Nation, K., and Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: evidence from children with reading comprehension difficulties. *J. Mem. Lang.* 39, 85–101. doi: 10.1006/jmla.1998.2564

National Reading Panel (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction.* Bethesda, MD: The National Institute of Child Health and Human Development.

Nilson, L. (2013). *Creating Self-Regulated Learners: Strategies to Strengthen Students? Self-Awareness and Learning Skills.* Herndon, VA: Stylus Publishing, LLC.

Oakhill, J., Cain, K., and Elbro, C. (2014). *Understanding and Teaching Reading Comprehension: A Handbook.* Abingdon: Routledge.

Oakhill, J., Hartt, J., and Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Read. Writ.* 18, 657–686. doi: 10.1007/s11145-005-3355-z

Ortiz, I. (2012). En torno a la validez del sistema de medición de la calidad de la educación en chile. *Estudios Pedagógicos* 38, 355–373. doi: 10.4067/S0718-07052012000200022

Paris, S. G., Lipson, M. Y., and Wixson, K. K. (1983). Becoming a strategic reader. *Contemp. Educ. Psychol.* 8, 293–316. doi: 10.1016/0361-476X(83)90018-8

Paul, S. A. S., and Clarke, P. J. (2016). A systematic review of reading interventions for secondary school students. *Int. J. Educ. Res.* 79, 116–127. doi: 10.1016/j.ijer.2016.05.011

Perfetti, C. A., and Hogaboam, T. (1975). Relationship between single word decoding and reading comprehension skill. *J. Educ. Psychol.* 67, 461–469. doi: 10.1037/h0077013

Potocki, A., Ecalle, J., and Magnan, A. (2013). Effects of computer-assisted comprehension training in less skilled comprehenders in second grade: a one-year follow-up study. *Comput. Educ.* 63, 131–140. doi: 10.1016/j.compedu.2012.12.011

Quené, H., and van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Commun.* 43, 103–121. doi: 10.1016/j.specom.2004.02.004

Quené, H., and van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *J. Mem. Lang.* 59, 413–425. doi: 10.1016/j.jml.2008.02.002

Roschelle, J. M., Pea, R. D., Hoadley, C. M., Gordin, D. N., and Means, B. M. (2000). Changing how and what children learn in school with computer-based technologies. *Future Child* 10, 76–101. doi: 10.2307/1602690

Rouse, C., and Krueger, A. (2004). Putting computerized instruction to the test: a randomized evaluation of a "scientifically-based" reading program. *Econ. Educ. Rev.* 23, 323–338. doi: 10.1016/j.econedurev.2003.10.005

See, B. H., and Gorard, S. A. (2014). Improving literacy in the transition period: a review of the existing evidence on what works. *Br. J. Educ. Soc. Behav. Sci. .* 4, 739–754. doi: 10.9734/BJESBS/2014/8417

Selwyn, N., Potter, J., and Cranmer, S. (2009). Primary pupils' use of information and communication technologies at school and home. *Br. J. Educ. Technol.* 40, 919–932. doi: 10.1111/j.1467-8535.2008.00876.x

Stanovich, K. E. (1986). Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Read. Res. Q.* 21, 360–407. doi: 10.1598/RRQ.21.4.1

Thibaut, P., Curwood, J. S., Carvalho, L., and Simpson, A. (2015). Moving across physical and online spaces: a case study in a blended primary classroom. *Learn. Media Technol.* 40, 458–479. doi: 10.1080/17439884.2014.959971

Torgerson, C. J., and Elbourne, D. (2002). A systematic review and meta-analysis of the effectiveness of information and communication technology (ICT) on the teaching of spelling. *J. Res. Read.* 25, 129–143. doi: 10.1111/1467-9817.00164

Tunmer, W. E., Nesdale, A. R., and Pratt, C. (1983). The development of young children's awareness of logical inconsistencies. *J. Exp. Child Psychol.* 36, 97–108. doi: 10.1016/0022-0965(83)90060-7

Vaughn, S., Klingner, J. K., Swanson, E. A., Boardman, A. G., Roberts, G., Mohammed, S. S., et al. (2011). Efficacy of collaborative strategic reading

with middle school students. *Am. Educ. Res. J.* 48, 938–964. doi: 10.3102/ 0002831211410305

**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.

Check for updates

# Child-Robot Interactions for Second Language Tutoring to Preschool Children

*Paul Vogt \*, Mirjam de Haas , Chiara de Jong , Peta Baxter and Emiel Krahmer*

*Tilburg Center for Cognition and Communication, Tilburg University, Tilburg, Netherlands*

In this digital age social robots will increasingly be used for educational purposes, such as second language tutoring. In this perspective article, we propose a number of design features to develop a child-friendly social robot that can effectively support children in second language learning, and we discuss some technical challenges for developing these. The features we propose include choices to develop the robot such that it can act as a peer to motivate the child during second language learning and build trust at the same time, while still being more knowledgeable than the child and scaffolding that knowledge in adult-like manner. We also believe that the first impressions children have about robots are crucial for them to build trust and common ground, which would support child-robot interactions in the long term. We therefore propose a strategy to introduce the robot in a safe way to toddlers. Other features relate to the ability to adapt to individual children's language proficiency, respond contingently, both temporally and semantically, establish joint attention, use meaningful gestures, provide effective feedback and monitor children's learning progress. Technical challenges we observe include automatic speech recognition (ASR) for children, reliable object recognition to facilitate semantic contingency and establishing joint attention, and developing human-like gestures with a robot that does not have the same morphology humans have. We briefly discuss an experiment in which we investigate how children respond to different forms of feedback the robot can give.

Keywords: social robots, second language tutoring, education, child-robot interaction, robot assisted language learning

## SOCIAL ROBOTS FOR SECOND LANGUAGE TUTORING

Given the globalization of our society, it is becoming increasingly important for people to speak multiple languages. For instance, the ability to speak foreign languages fosters people's mobility and increases their chances for employment. Moreover, immigrants to a country need to learn the official host language. Since young children are most flexible at learning languages, starting second language (L2) learning in preschool would provide them a good opportunity to acquire the second language more fluently at a later age (Hoff, 2013).

One trend in the digital age of the 21st century is that technologies are being developed for educational purposes, including technologies to support L2 tutoring. There exist many forms of digital technologies for PCs, laptops or tablet computers that support second language learning, although there is little evidence about their efficacy (Golonka et al., 2014; Hsin et al., 2014). While children can benefit from playing with such technologies, these systems lack the situated and

embodied interactions that young children naturally engage in and learn from (Glenberg, 2010; Leyzberg et al., 2012). Social robots represent an emerging technology that provides situatedness and embodiment, and thus have potential benefits for educational purposes. In essence, social robots are autonomous physical agents, often with human-like feature, that can interact socially with humans in a semi-natural way for prolonged periods of time (Dautenhahn, 2007). The use of social robots, in comparison to more traditional digital technologies, allows for the development of tutoring systems more akin to human tutors, especially with respect to the situated and embodied social interactions between child and robot. Thus, this offers the opportunity to design robots such that they interact in a way that optimizes the child's language learning.

Recently, an increasing interest has emerged to develop social robots to support children with learning a second language (Kanda et al., 2004; Belpaeme et al., 2015; Kennedy et al., 2016). While a social robot cannot provide tutoring to the level humans can, recent studies suggest that using social robots can result in an increased learning gain compared to digital learning environments for tablets or computers (Han et al., 2008; Leyzberg et al., 2012). It is, however, unclear why this is the case. Perhaps the physical presence of the robot draws the attention of children for longer periods of time, but the embodiment and situatedness of the learning environment perhaps also helps the children to ground the language more strongly than interactions with virtual objects do.

While there is a fair body of research on robot tutors, a comprehensive description of the design features for a second language robot tutor based on what is known about children's language acquisition is lacking. What are the design features of child-robot interactions that would support second language learning? And, to what extent can these interactions be implemented in today's social robot technologies? In this perspective article, we try to answer these questions based on theoretical accounts from the literature on children's language acquisition in combination with our own experiences in designing a tutor robot.

## DESIGNING CHILD-ROBOT INTERACTIONS

In our project, we aim to design a digital learning environment in which preschool children interact one-on-one with a social robot that supports either their learning of English as a foreign language, or the school language for those children who have a different native language (Belpaeme et al., 2015). In particular, the project aims to develop a series of tutoring sessions revolving around three increasingly complex domains (numbers, spatial relations and mental vocabulary). In each session, the child will engage with the robot (a Softbank Robotics NAO robot) in a game-like scenario focusing on learning a small number of target words. The contextual setting is generally displayed on a tablet computer that occasionally also provides some verbal support, however, the robot acts as the interactive tutor. Below we discuss the design features and considerations that we believe are crucial to design a successful tutoring system.

## Peer-Like Tutoring

One of the first questions that comes up when designing a robot tutor is whether the robot should take the role of a teacher or a peer. Research on children's language acquisition has demonstrated that children learn more effectively from an adult who can use well-defined pedagogical methods for teaching children using clear directions, explanations and positive feedback methods (Matthews et al., 2007). However, designing and framing the robot as an adult tutor has the disadvantage that children will form expectations about the robot's behavior and proficiency that cannot be met with current technology (Kennedy et al., 2015). Due to technological limitations of the robot and underlying software, communication breakdowns are more likely to occur than with a human. For a peer robot introduced as a fellow language learner, breakdowns in communication are more acceptable. Moreover, interacting with robots acting as peers is conceived as more fun (Kanda et al., 2004), allows for learning-by-teaching (Tanaka and Matsuzoe, 2012) and has a proven to be efficient in teaching children how to write (Hood et al., 2015). Furthermore, there is some evidence that children's learning can benefit from interacting with peers (Mashburn et al., 2009). Given these considerations, we believe it is desirable to frame or introduce the robot as a peer and friend, yet design its interactions insofar possible based on pedagogically well-established strategies to scaffold language learning.

## First Impressions

To implement effective tutoring, the robot needs to interact with children in multiple sessions, so they have to be motivated to engage in long-term interactions with the robot. Establishing common ground between child and robot can contribute to this (Kanda et al., 2004), but first impressions to establish trust and rapport are also crucial (Hancock et al., 2011).

Despite the wealth of studies regarding the introduction of entertainment robots as toys to children (e.g., Lund, 2003), surprisingly little research has been conducted on designing protocols on how to introduce a robot tutor to a group of preschool children. Fridin (2014) presents one exception, and found that introducing a robot tutor to children in group sessions improved subsequent interactions compared to introducing the robot to children in individual sessions. Another study by Westlund et al. (2016) found that the way a robot is framed, either as a machine or a social entity, affected the way children later engaged with the robot. They concluded that introducing the robot as a machine could create a more distant relation between child and robot, thus reducing acceptance. We therefore decided to frame the robot in our project as a social playmate for the children and introduced the robot in a group session. However, the NAO robot is slightly taller and more rigid than the fluffy huggable Tega robot, which Westlund et al. (2016) used, and we observed that some 3-year-old children were somewhat intimidated by the NAO robot on their first encounter. Such a first impression of the robot could reduce the trust that the child had for the robot, which could negatively affect their willingness to interact with the robot in the short-term, but also in the long-term. To develop a successful first encounter and to build

trust between the child and robot, we designed the following strategy for introducing the robot to 3-year-old children at their preschool.

Pilot studies revealed that some children got anxious when the robot was introduced and then suddenly started to move. To familiarize children prior to their first encounter with the robot, it is therefore advisable to prepare them well. For our study, we sent coloring pages of the robot to the preschools during recruitment and asked the pedagogical assistants to talk a little bit about the robots to the children. About 1 week before the experimental trials, the experimenters introduced the robot in class during their daily "circle time", as this provided a safe and familiar environment with the whole group in which the pedagogical assistants usually introduce new topics or new activities. One experimenter first introduced the robot by telling a story about Robin, the name of our robot, using a makeshift picture book. In this story we explained the similarities and dissimilarities between the robot and children to construct the type of common ground considered to have a positive effect on the learning outcome (Kanda et al., 2004). For example, we told that Robin enjoys dancing and wants to meet new friends, and even though he does not have a mouth and because of that cannot smile, he can smile using his eye LEDs.

After this story, another experimenter entered the room with the robot while it was actively looking at faces to provide an animate feeling. The robot introduced itself with a small story about itself and by performing a dance in which the children were encouraged to participate. The end of the circle time consisted of getting a blanket for the robot so it could "sleep". This introduction was repeated later on the days we conducted the experiment in one-on-one sessions. While by then most children were comfortable interacting with the robot, some were still timid and anxious. To encourage these children to feel comfortable, one of the experiment leaders would sit next to the child during the warm-up phase of the experiment and motivate the child to respond to the robot when necessary until the child was sufficiently comfortable to interact with the robot by herself/himself. We found that the younger 3-year olds required more support from the experimenters than the older 3-year olds (Baxter et al., 2017). Although we are still analyzing the experiments, preliminary findings suggest that our introduction helped children to build trust and common ground with the robot effectively.

## Temporal Contingency

Research has shown that it is crucial for children's language development that their communication bids are responded to in a temporally contingent manner (Bornstein et al., 2008; McGillion et al., 2013). This, however, faces a technological challenge. While adults tend to take over turns very rapidly, robots require relatively long processing time to produce a response. Nevertheless, in our first experiment (de Haas et al., 2016), we observed that children were at first surprised by the delayed responses, but quickly adapted to the robot and waited patiently for a response. Perhaps this is because children also require longer than adults to take turns

(Garvey and Berninger, 1981) and having framed the robot as a peer children made the delays more plausible or expected. Nevertheless, while a lag in temporal contingency may not harm the interaction with children, it may harm learning. One way to remedy this may be to have the robot start responding by providing a backchannel signal, such as "uhm" to indicate the robot is (still) taking his turn, but requires more time to process (Clark, 1996).

## Semantic Contingency

Robots should not only respond to children in a timely fashion, but also in a semantically contingent fashion (i.e., consistent with the child's focus of attention), as this too has a positive effect on children's language acquisition (Bornstein et al., 2008; McGillion et al., 2013). For instance, research has shown that by responding in a semantically contingent manner, either verbally or by following children's gaze, (joint) attention is sustained for a longer duration (Yu and Smith, 2016), allowing children to learn more about a situation. To achieve semantically contingent responses, the robot should be able to understand the child's communication bids, construct joint attention with the child, or at least identify what the child is attending to. Monitoring children's behavior and establishing joint attention are therefore considered crucial for designing a successful robot tutor.

## Monitoring Children's Behavior

To understand children's communication bids, as well as to test their pronunciation of the L2, it is important that the robot be equipped with well-functioning automatic speech recognition (ASR). However, the performance of state-of-the-art ASR for children is still suboptimal, especially for preschool-aged children (Fringi et al., 2015; Kennedy et al., 2017). Reasons for this include that children's pronunciation is often flawed and that their speech has a different pitch than adults. Moreover, relatively little research has been carried out in this domain and not much data exist to train ASR on. While it can be expected that the performance of ASR for children will improve in the not too distant future (Liao et al., 2015), until then alternative strategies need to be developed that do not (exclusively) rely on ASR.

In our project, we explore various strategies to achieve this, both based on monitoring non-verbal behaviors of the children and focusing on comprehending rather than producing L2. The first strategy relies on providing children tasks they have to perform in the learning environment, such as placing "a toy cow behind a tree" when teaching spatial language. This, however, requires the visual object recognition on the robot to work well, which is only the case when the scene contains a limited set of distinctively recognizable objects, such as distinctly colored objects (Nguyen et al., 2015). A potential solution explored in our project is to use objects with build-in RFID sensors that can be tracked automatically. The second solution we explore is to use a touch screen tablet that displays scenes the child can manipulate, which not only has the advantage of avoiding the problem of object recognition, but also allows us to control

the robot's responses and vary the scenes in real time. A downside, however, is that it takes away the 3-dimensional physical aspect of embodied cognition that would help the children to better entrench what they learn (Glenberg, 2010). Currently, experiments are underway to investigate the effect of using real vs. virtual objects. These solutions not only aid in understanding the child's communication bids, it also helps in identifying their attention and can thus contribute to establishing joint attention.

## Joint Attention and Gestures

Joint attention, where interlocutors attend on the same referent, is a form of social interaction that has been shown to support children's language learning (Tomasello and Farrar, 1986). One way to establish joint attention with a child is to guide their attention to a referent using gestures, such as pointing or iconic gestures. The ability to produce gestures in the real world is potentially one of the main advantages of using physical robots as opposed to virtual agents, who may have a harder time to establish joint attention. However, many robots' physical morphologies do not correspond one-to-one to the human body. Hence, many human gestures cannot be translated directly to robot gestures. For instance, the NAO robot that we use in our research has a hand with three fingers that cannot be controlled independently, so index finger pointing cannot be

achieved (see **Figure 1**). Will children still recognize NAO's arm extension as a pointing gesture? And if so, will they be able to identify the object the robot refers to? We are currently running an experiment to investigate how NAO's pointing gestures are perceived, and preliminary findings show that participants have difficulty identifying the referred object on a small tablet screen. Similar issues arise when developing other gestures. One of the other non-verbal behaviors we are using is the coloring of NAO's eye LEDSs to indicate the robot's happiness as a form of positive feedback, since the robot cannot smile with its mouth.

## Feedback

Feedback, too, is an interactional feature known to help language learning (Matthews et al., 2007; Ateş-Şen and Küntay, 2015). The question is how should the robot provide feedback, such that it is both pleasant and effective for learning? While adults provide positive feedback explicitly, they usually provide negative feedback implicitly by reformulating children's errors in the correct form. In child-child interactions, however, Long (2006) found that there was a clear advantage in learning from explicit negative feedback (e.g., by saying "no, that's wrong, you need to say 'he ran'") when compared to reformulating feedback (the learner says "he runned" and the teacher reacts with "he ran").



**FIGURE 1 | NAO pointing to a block with three fingers.** (Note that written, informed consent was obtained from the parents of the child for the publication of this image).

To investigate how children experience feedback from a peer robot, we carried out an experiment among 85 3-year-old Dutch-speaking children at preschools in Netherlands (de Haas et al., 2016, 2017). In this experiment, the children interacted with a NAO robot during which they received a short lesson on how to count from 1 to 4 in English. After a short training phase, in which the children were presented with the four counting words twice in relation to body parts and wooden blocks, they were given instructions by the robot to pick up a given number of blocks. While the instructions were given in their native language, the numbers were uttered in English. In response to the child's ability to achieve the task, the robot provided feedback. The experiment followed a between-subjects design with three conditions: adult-like feedback (explicit positive and implicit negative), peer-like feedback (no positive and explicit negative) and no feedback. We did not find significant differences in learning gain between the conditions, probably because the target words were insufficiently often repeated. However, we explored the way in which the children engaged with the robot after they received feedback and we found that children looked less often at the experimenter in the feedback conditions than in the no feedback condition. Further analyses are carried out to evaluate how the children responded to the various forms of feedback to find out what type of feedback would be most effective for achieving both acceptable and effective tutoring interactions.

## Zone of Proximity and Adaptivity

Finally, from a pedagogical point of view it is desirable that the interactions between child and robot be sufficiently challenging and varied so that the child has a target to learn from, but at the same time interactions should not be too difficult, because that may frustrate the child causing it to lose interest in the robot (Charisi et al., 2016). In other words, the robot should remain in Vygotsky's Zone of Proximity that supports an effective learning environment (Vygotsky, 1978). In order to achieve this, the robot should be able to keep track of the children's advancements in language learning and perhaps their emotional states during the tutoring sessions, and adapt to these. While the former can be monitored as discussed previously, it may be possible to detect emotional states known to influence learning (e.g., concentration, confusion, frustration and boredom) using methods from affective computing (D'Mello and Graesser, 2012). Using this type of information, it is possible to adapt the tutoring sessions by either reducing or increasing the number of repetitions, and/or change the subject (Schodde et al., 2017).

## CONCLUSION

This perspective article presented some design features that we consider crucial for developing a social robot as an effective second language tutor. We believe the robot is most effective when it is framed as a peer, i.e., as a fellow language learner and playmate, but that is designed to use adult-like interaction strategies to optimize learning efficacy. In order to establish common ground and trust to facilitate long-term interactions, we consider it essential that the robot be introduced with appropriate care on the first encounter. As an example, we outlined our strategy for introducing a robot to preschool children. Interactions between child and robot should be contingent and multimodal, and provide appropriate forms of feedback. We argued that the robot should remain within Vygotsky (1978) Zone of Proximal Development and thus should adapt to the individual level of the child.

We also discussed some technical challenges that need to be solved in order to implement contingent interactions; the most important of which we believe is ASR, which presently does not work well for children's speech. While various technical challenges still remain, we expect that social robots will provide effective digital technologies to support second language development in the years to come.

The present list of design features covers many aspects that need to be considered when developing a tutor robot, but it is not yet comprehensive. One aspect that has not been covered, for instance, concerns the design of robots for children from different cultures, which could require different design choices (Shahid et al., 2014). For example, in some cultures education is more teaching-centered (Hofstede, 1986) and thus designing the tutor as a peer robot may be less effective or acceptable (Tazhigaliyeva et al., 2016). Concluding, this perspective article offers only a first step towards a comprehensive list of design features for tutor robots and additional research is needed to complete and optimize the list.

## ETHICS STATEMENT

## AUTHOR CONTRIBUTIONS

PV, MH and EK designed the conceptual aspects of the article; PV, MH, CJ and PB carried out the literature review; PV, EK and MH designed the feedback study; MH, CJ and PB designed the introduction study; MH, CJ and PB carried out the studies; PV and MH wrote the article; CJ, PB and EK revised the article critically.

## FUNDING

## ACKNOWLEDGMENTS

# REFERENCES

Ateş-Şen, B. A., and Küntay, A. C. (2015). Children's sensitivity to caregiver cues and the role of adult feedback in the development of referential communication. *The Acquisition of Reference*, eds L. Serratrice and S. E. M. Allen (Amsterdam: John Benjamins), 241–262.

Baxter, P., De Jong, C., Aarts, A., de Haas, M., and Vogt, P. (2017). "The effect of age on engagement in preschoolers' child-robot interactions," in *Companion proceedings of the 12th Annual ACM International Conference on Human-Robot Interaction* (HRI'17), (Vienna, Austria).

Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E. J., Kopp, S., et al. (2015). "L2TOR-second language tutoring using social robots," in *First Workshop on Educational Robots* (WONDER), (Paris, France).

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C. S., and Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: longitudinal analysis of a multidimensional, modular and specific parenting construct. *Dev. Psychol.* 44, 867–874. doi: 10.1037/0012-1649.44.3.867

Charisi, V., Davison, D., Reidsma, D., and Evers, V. (2016). "Children and robots: a preliminary review of methodological approaches in learning settings," in *2nd Workshop on Evaluating Child Robot Interaction - HRI*, (Christchurch, New Zealand).

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human-robot interaction. *Philos. Trans. R. Soc. B Biol. Sci.* 1480, 679–704. doi: 10.1098/rstb.2006.2004

de Haas, M., Baxter, P., de Jong, C., Vogt, P., and Krahmer, E. (2017). "Exploring different types of feedback in preschooler and robot interaction," in *Companion proceedings of the 12th Annual ACM International Conference on Human-Robot Interaction* (HRI'17), (Vienna, Austria).

de Haas, M., Vogt, P., and Krahmer, E. J. (2016). "Enhancing child-robot tutoring interactions with appropriate feedback," in *Proceedings of First Workshop on Long-Term Child-Robot Interaction. IEEE Ro-Man*, (New York, NY).

D'Mello, S., and Graesser, A. (2012). Dynamics of affective states during complex learning. *Learn. Instr.* 22, 145–157. doi: 10.1016/j.learninstruc.2011.10.001

Fridin, M. (2014). Kindergarten social assistive robot: first meeting and ethical issues. *Comput. Hum. Behav.* 30, 262–272. doi: 10.1016/j.chb.2013.09.005

Fringi, E., Lehman, J., and Russell, M. J. (2015). "Evidence of phonological processes in automatic recognition of children's speech," in *16th Annual Conference of the International Speech Communication Association*, (Dresden, Germany), 1621–1624.

Garvey, C., and Berninger, G. (1981). Timing and turn taking in children's conversations. *Discourse Process.* 4, 27–57. doi: 10.1080/01638538109544505

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdiscip. Rev. Cogn. Sci.* 4, 586–596. doi: 10.1002/wcs.55

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., and Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Comput. Assist. Lang. Learn.* 27, 70–105. doi: 10.1080/09588221.2012.700315

Han, J. H., Jo, M. H., Jones, V., and Jo, J. H. (2008). Comparative study on the educational use of home robots for children. *J. Inf. Process. Syst.* 4, 159–168. doi: 10.3745/jips.2008.4.4.159

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors* 53, 517–527. doi: 10.1177/0018720811417254

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Dev. Psychol.* 49, 4–14. doi: 10.1037/a0027238

Hofstede, G. (1986). Cultural differences in teaching and learning. *Int. J. Intercult. Relat.* 10, 301–320. doi: 10.1016/0147-1767(86)90015-5

Hood, D., Lemaignan, S., and Dillenbourg, P. (2015). "When children teach a robot to write: an autonomous teachable humanoid which uses simulated handwriting," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, (New York, NY), 83–90.

Hsin, C.-T., Li, M.-C., and Tsai, C.-C. (2014). The influence of young children's use of technology on their learning: A. *Educ. Technol. Soc.* 17, 85–99. Available online at: https://eric.ed.gov/?id=EJ1045554

Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: a field trial. *Hum. Comput. Interact.* 19, 61–84. doi: 10.1207/s15327051hci1901&2_4

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2015). "Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions," in *International Conference on Social Robotics*, eds A. Tapus, E. André, J.-C. Martin, F. Ferland and M. Ammi (New York, NY: Springer International Publishing), 327–336.

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016). "Social robot tutoring for child second language learning," in *Proceedings of the 11th Annual ACM/IEEE International Conference on Human-Robot Interaction* (HRI'16), (Christchurch, New Zealand), 231–238.

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., et al. (2017). "Child speech recognition in human-robot interaction: evaluations and recommendations," in *Proceedings of the 12th Annual ACM International Conference on Human-Robot Interaction* (HRI'17), (Vienna, Austria).

Leyzberg, D., Spaulding, S., Toneva, M., and Scassellati, B. (2012). "The physical presence of a robot tutor increases cognitive learning gains," in *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, (Sapporo, Japan), 1882–1887.

Liao, H., Pundak, G., Siohan, O., Carroll, M. K., Coccaro, N., Jiang, Q. M., et al. (2015). "Large vocabulary automatic speech recognition for children," in *Interspeech*, (Dresden, Germany), 1611–1615.

Long, M. H. Ed. (2006). "Recasts in SLA: the story so far," in *Problems in SLA. Second Language Acquisition Research Series*, (Mahwah, NJ: Lawrence Erlbaum Associates), 75–116.

Lund, H. H. (2003). "Adaptive robotics in the entertainment industry," in *Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation (cira2003)*, (Vol. 2), (Kobe, Japan), 595–602.

Mashburn, A. J., Justice, L. M., Downer, J. T., and Pianta, R. C. (2009). Peer effects on children's language achievement during pre-kindergarten,. *Child Dev.* 80, 686–702. doi: 10.1111/j.1467-8624.2009.01291.x

Matthews, D., Lieven, E., and Tomasello, M. (2007). How toddlers and preschoolers learn to uniquely identify referents for others: a training study. *Child Dev.* 6, 1744–1759. doi: 10.1111/j.1467-8624.2007.01098.x

McGillion, M., Herbert, J., Pine, J., Keren-Portnoy, T., Vihman, M., and Matthews, D. (2013). Supporting early vocabulary development: what sort of responsiveness matters? *IEEE Trans. Auton. Ment. Dev.* 5, 240–248. doi: 10.1109/tamd.2013.2275949

Nguyen, T. L., Boukezzoula, R., Coquin, D., Benoit, E., and Perrin, S. (2015). "Interaction between humans, NAO robot and multiple cameras for colored objects recognition using information fusion," in *8th International Conference on Human System Interaction* (HSI), (Warsaw, Poland), 322–328.

Schodde, T., Bergmann, K., and Kopp, S. (2017). "Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making," in *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction* (HRI 2017), (Vienna, Austria).

Shahid, S., Krahmer, E., and Swerts, M. (2014). "Child–robot interaction across cultures: how does playing a game with a social robot compare to playing a game alone or with a friend? *Comput. Human Behav.* 40, 86–100. doi: 10.1016/j.chb.2014.07.043

Tanaka, F., and Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *J. Hum. Robot Interact.* 1, 78–95. doi: 10.5898/jhri.1.1.tanaka

Tazhigaliyeva, N., Diyas, Y., Brakk, D., Aimambetov, Y., and Sandygulova, A. (2016). "Learning with or from the robot: exploring robot roles in educational context with children," in *International Conference on Social Robotics*, eds A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs and H. He (New York, NY: Springer International Publishing), 327–336.

Tomasello, M., and Farrar, M. J. (1986). Joint attention and early language. *Child Dev.* 57, 1454–1463. doi: 10.2307/1130423

Vygotsky, L. (1978). *Mind in Society*. Harvard: Harvard University Press.

Westlund, J. M. K., Martinez, M., Archie, M., Das, M., and Breazeal, C. (2016). "Effects of framing a robot as a social agent or as a machine on children's social behavior," in *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, eds S. Y. Okita, T. Shibata and B. Mutlu (Washington, DC: IEEE), 688–693.

Yu, C., and Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Curr. Biol.* 26, 1235–1240. doi: 10.1016/j.cub.2016.03.026

**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.

# Language Learning Enhanced by Massive Multiple Online Role-Playing Games (MMORPGs) and the Underlying Behavioral and Neural Mechanisms

Yongjun Zhang[1,2]*, Hongwen Song[3], Xiaoming Liu[1,3], Dinghong Tang[1], Yue-e Chen[1,4] and Xiaochu Zhang[2,3,5,6,7]*

[1] School of Foreign Languages, Anhui Jianzhu University, Hefei, China, [2] Center for Biomedical Engineering, School of Information Science and Technology, University of Science and Technology of China, Hefei, China, [3] School of Humanities and Social Science, University of Science and Technology of China, Hefei, China, [4] School of Public Affairs, University of Science and Technology of China, Hefei, China, [5] CAS Key Laboratory of Brain Function and Disease, School of Life Science, University of Science and Technology of China, Hefei, China, [6] State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China, [7] Center of Medical Physics and Technology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, China

Massive Multiple Online Role-Playing Games (MMORPGs) have increased in popularity among children, juveniles, and adults since MMORPGs' appearance in this digital age. MMORPGs can be applied to enhancing language learning, which is drawing researchers' attention from different fields and many studies have validated MMORPGs' positive effect on language learning. However, there are few studies on the underlying behavioral or neural mechanism of such effect. This paper reviews the educational application of the MMORPGs based on relevant macroscopic and microscopic studies, showing that gamers' overall language proficiency or some specific language skills can be enhanced by real-time online interaction with peers and game narratives or instructions embedded in the MMORPGs. Mechanisms underlying the educational assistant role of MMORPGs in second language learning are discussed from both behavioral and neural perspectives. We suggest that attentional bias makes gamers/learners allocate more cognitive resources toward task-related stimuli in a controlled or an automatic way. Moreover, with a moderating role played by activation of reward circuit, playing the MMORPGs may strengthen or increase functional connectivity from seed regions such as left anterior insular/frontal operculum (AI/FO) and visual word form area to other language-related brain areas.

Keywords: Massive Multiple Online Role-Playing Games (MMORPGs), language learning, interaction, reward, behavioral mechanism, neural mechanism

## INTRODUCTION

Massive Multiplayer Online Role-Playing Games (MMORPGs) are gaining more and more popularity compared to other genres of commercial games. The main feature of MMORPGs is gamers' purposeful interaction with peers and game-embedded narratives elicited by the game design. The players' ultimate purpose is to get reward so as to progress through the game

hierarchy by undertaking game tasks known as quests, usually with the help of game-based organizations known as guilds. Guild membership offers novices opportunities to get their gaming skills promoted through interaction with more experienced players (Peterson, 2012). Notably, MMORPGs may bring about some negative effects such as excessive playing or gaming addiction (Petry and O'Brien, 2013), and psychiatric comorbidity (Han et al., 2015). However, MMORPGs can also provide players with benefits such as feelings of achievement and sense of community (Sublette and Mullan, 2012), and possibilities for educational use (González-González and Blanco-Izquierdo, 2012).

Applying MMORPGs to foreign language (FL) or second language (L2) learning has become a research focus in that, gamers/learners immersed in MMORPGs learning context are more relaxed and motivated to interact with peers or gaming instructions (Bytheway, 2014), and they outperform those attending traditional classrooms in terms of language skills (Rankin et al., 2009; Suh et al., 2010; Kim et al., 2013). The main affordances of MMORPGs for language learning are the immersive interactive environments and multiple options for players to engage in authentic communication through listening, speaking, reading, and writing in the target language with other interlocutors (Rama et al., 2012). Apart from the commercial MMORPGs, researchers develop some educational MMORPGs to facilitate FL/L2 learning. Such educational MMORPGs are also named serious games, which "include an identifiable teaching presence specifically for improving some aspect of language proficiency" (Thorne et al., 2012a). Serious games' main objectives are learning and behavior change (Connolly et al., 2012) and can also provide gamers with in-game rewards if they accomplish tasks (Nagle et al., 2014), leading to greater learning motivation and more effective learning relative to traditional tools or approaches (Iten and Petko, 2016). In this paper, we briefly review studies focusing on commercial or serious MMORPGs' benefit to learning FL/L2 and discuss the potential mechanisms underlying the educational assistant role of MMORPGs in language learning from behavioral and neural perspectives.

## METHODS

We searched for literatures on Google Scholar, Web of Science, and ScienceDirect with no date restrictions. Terms used were "massive multiplayer online role*" or "MMORPG*" in combination with "language learning" or "second language" or "FL" or "language teaching." Since some online games especially 3D online games bear features of MMORPGs, we also used terms like "online game" and "3D online game" in our search. We finally selected the most relevant papers for our review and some studies were identified through checking reference lists of the indexed papers. Available studies were organized in two groups based on their aims: macroscopic studies on MMORPGs' benefit for gamers/learners' overall FL/L2 learning and microscopic studies on MMORPGs' benefit for one or more specific FL/L2 abilities.

## MMORPGs' BENEFIT FOR GAMERS'/LEARNERS' FL/L2 LEARNING

In the studies conducted by Rankin et al. (2008), Zheng et al. (2009, 2012), and Rama et al. (2012), MMORPG's affordance of interaction was found to benefit FL/L2 acquisition or development. Interestingly, Zheng et al. (2009, 2012) found that gamers can realize their heterarchical values while learning English in MMORPG's interactive context. Peterson (2011, 2012) attached more importance to learners' attitudes exhibited in MMORPG-based interaction. The former study suggested that the MMORPG-based interaction can lead to learners' positive feedback, by which language development may be facilitated. The latter study showed that in online linguistic and social interaction, learners adopted polite expressions to build up collaborative relationships, used continuers and requests for assistance to maintain intersubjectivity, and became increasingly positive toward gaming and language learning emerged in gaming. Thus, such interaction can contribute to learners' sociocultural competence, positive attitudes toward FL learning, and coherence and appropriateness of target language production, all of which are beneficial for FL development. Considering that gamers are involved in both virtual spaces and real world settings, researchers are interested to ascertain whether the language learning-related resources and interactions in and out of the MMORPGs' context can influence each other or work together to promote the gamers' language development. In two successive studies, Kongmee et al. (2011, 2012) validated that linguistic knowledge and communicational skills can be transferable between the virtual spaces and real world. Scholz (2015) reached a similar conclusion that if learners are given the opportunity to communicate with other players and experience the game at their own pace, they can transfer linguistic constructions from MMORPGs' contexts to various non-gaming contexts, so that L2 learning can be developed more effectively. To dig it further, Thorne et al. (2012b) employed semiotic ecology theory to indicate that game-embedded texts, player-to-player interaction, and game-external websites' resources constitute gamers/learners' complex semiotic ecologies, which are significant for L2 development.

Comparatively, more studies have examined the effect of MMORPG on enhancing gamers' some specific FL/L2 abilities. In view of the central place of vocabulary in language learning, some studies have argued that vocabulary learning can be facilitated by gamers' interaction in playing MMORPGs (Bytheway, 2014; Shahriarpour and Kafi, 2014; Yudintseva, 2015; Zheng et al., 2015). Contrary to these studies, Milton et al. (2012) reached a relatively conservative conclusion that there is little opportunity for lexical growth without teacher's control in the MMORPG-based learning activities. Other studies have shown that vocabulary acquisition and other skills such as communicative competence (Peterson, 2010), sentence construction (Yang and Hsu, 2013), and reading skills (Dourda et al., 2014) can be developed simultaneously by gamers' interaction in MMORPG-based instruction. Besides, Huang and Yang (2014) investigated effects of English proficiency and gaming experience on incidental vocabulary acquisition in a

MMORPG and found that vocabulary was more noticed by learners with medium gaming experience in gaming requirement condition, and was more perceived by learners with higher English proficiency in flashcard condition. Apart from above-mentioned studies focusing on vocabulary development in playing MMORPGs, many studies have demonstrated the positive effects of MMORPGs on developing basic language skills such as FL listening ability (Hu and Chang, 2007), speaking ability (Lai and Wen, 2012), production of narratives (Colby and Colby, 2008; Neville, 2010, 2015), communicative competence (Wu and Richards, 2012; Berns et al., 2013), and communicative skills, together with learners' listening, reading, and writing skills (Suh et al., 2010). In addition, Hsu (2015) reported that the MMORPG has long-term effects on developing learners' incremental intelligence (i.e., accumulated intelligence through hard work) which was significantly related to their performance on standardized language test.

It is indicated that existing studies have mainly explored MMORPGs' benefit for FL/L2 learning based on MMORPGs' affordance of interactive function. Specifically, MMORPGs afford gamers opportunities to communicate with peers from the same guild. Such communication requires active negotiation of meaning in FL/L2 among gamers so that their language skills can be developed (Bytheway, 2011; Rama et al., 2012). Meanwhile, gamers also interact with game-embedded narratives or instructions and they may get positive feedback so as to move on if the embedded texts are properly understood. Notably, when comprehending those embedded texts, gamers may frequently ask for their peers' help (Dourda et al., 2014). Accordingly, some researchers (Thorne, 2008; Peterson, 2012; Sundqvist and Sylvén, 2012) have tried to explain MMORPGs' role in facilitating language learning from a sociocultural perspective that employs Vygotsky's zone of proximal development, which is "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978). They have suggested that FL/L2 learning can be promoted by in-game social interaction, during which less proficient gamers/learners can negotiate meaning with and learn from more capable gamers/learners. This explanation sheds light on the FL/L2 development process in gaming. However, the underlying behavioral and neural mechanisms of MMORPG-based FL/L2 development remain unexplored. Because learners/gamers are more motivated to interact with peers in MMORPGs' contexts than they are in traditional teaching settings (Peterson, 2011, 2012; Bytheway, 2014; Shahriarpour and Kafi, 2014; Zheng et al., 2015; Howard-Jones and Jay, 2016), to figure out the source of such stronger motivation appears to be fundamental for investigating the behavioral and neural mechanisms under discussion. Evidence has shown that rewarding the gamers for meeting progressively demanding performance levels increased gamers' intrinsic motivation (Cameron et al., 2001; Pierce et al., 2003). More recent studies have also shown that rewards such as virtual badges have positive effects on increasing learners' motivation and learning outcomes in serious games (Filsecker and Hickey, 2014), and that gamers may take meta-game reward

systems as intrinsically motivating in game contexts (Cruz et al., 2015). Therefore, reward is an essential factor in motivating gamers/learners to get involved in the in-game interaction and should be taken into account when the behavioral and neural mechanisms of MMORPGs' role in promoting FL/L2 learning are investigated.

# POSSIBLE BEHAVIORAL MECHANISM UNDERLYING MMORPGs' EDUCATIONAL ROLE IN LANGUAGE LEARNING

Recent studies have validated strong reward effects on the allocation of attention (Hickey et al., 2010; Anderson et al., 2011; Anderson and Yantis, 2013; Lucas et al., 2013), and have shown that stimuli associated with reward in both current and past contexts can bias attentional selection (Anderson et al., 2013; Bourgeois et al., 2015). Furthermore, social rewards such as positive expressions can also shape attentional bias (Anderson, 2015). An integrated review conducted by Le Pelley et al. (2016) concluded that reward influences attention to reward-relevant stimuli. These findings provide us with a deeper insight into the potential behavioral mechanism involved in MMORPG-based language learning. In MMORPGs, reward-associated stimuli can range from some certain gaming skills to interaction with game-embedded texts and peers, which can lead to accomplishment of quests and reward procurement. When gamers/learners are engaged in MMORPGs, they may procure both monetary-like reward such as badges or superior equipment and social reward such as compliments from peers, which prompt them to bias attention and allocate more cognitive resources toward all the reward-related cues emerged in either real-time gaming or past gaming behavior. We thus hypothesize that the potential behavioral mechanism may relate closely to learners' attentional bias toward both gaming process and gamers' interaction with embedded game texts and other gamers.

Attentional bias has been validated as a behavioral tendency among excessive online gamers, who generally distribute more attention to game-related cues such as words or pictures and increase their emotional processing of those cues (for a review, see Zhang et al., 2016). Most studies reviewed here didn't filter participants, and thus included excessive gamers, casual gamers, and novice ones. As such it is worth discussing if the casual and novice gamers are also likely to exhibit attentional bias. An event-related potentials study conducted by Thalemann et al. (2007) revealed casual players also distributed more attention to game-related materials than to neutral cues and they might be highly emotionally involved in online gaming. Han et al. (2010) recruited healthy novices and asked them to play a novel online game for 10 days. Activity was elicited in the dorsolateral prefrontal cortex (DLPFC), parahippocampal gyrus, and thalamus by game cues in contrast to neutral cues for all participants. It is DLPFC that has been found to be related with attentional bias in some studies (Luijten et al., 2012; Jacob et al., 2014). Based on these findings, we may cautiously reach a
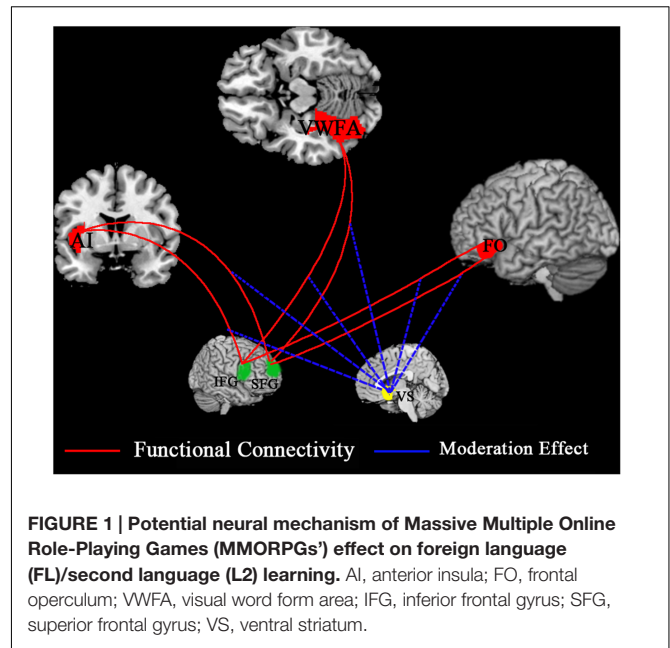
preliminary conclusion that attentional bias may also arise among casual gamers and novices after they are engaged in online games for a certain period.

Since attentional selection can be operated via a volitional top-down mode derived from task demands or an automatic bottom-up mode triggered by salient stimuli (Corbetta and Shulman, 2002; Buschman and Miller, 2007; Shomstein et al., 2010; Lee and Shomstein, 2014), how gamers/learners employ the two different modes to allocate their attentional resources is another issue warranting consideration. Le Pelley et al. (2016) raised the question whether attentional bias to task-relevant stimuli is a top-down (under participants' control) or a bottom-up (automatic) process, and they suggested that it was premature to define which one takes effect, because existing studies can be explained by either the former or the latter, or a combination of the two. As to the context of MMORPGs, we suppose that the two processes can be adopted in different ratios by different types of players. For novice players, they may more frequently use the top-down process in which they have to strategically control their own gaming behavior and allocate attentional resources to task-related cues in order to make less mistakes, while for the players with higher gaming proficiency, they tend to utilize more of the bottom-up process, because those task-related cues are psychologically more salient for them and their gaming experiences are rich enough to exert an automatic effect on attentional capture.

## POSSIBLE NEURAL MECHANISM UNDERLYING MMORPGs' EDUCATIONAL ROLE IN LANGUAGE LEARNING

Language processing depends on a widely distributed brain network, and specific first or second language abilities are proven to be positively related with various functional connectivities (FC) within this language network (Wei et al., 2012; Deng et al., 2015; Chai et al., 2016). Furthermore, similar brain areas can be activated in both language learning and online gaming (Khatibi and Cowie, 2013). We therefore suggest that gamers/learners' frequent in-game interaction may strengthen or increase their FC associated with language processing. Additionally, in view of the reward effect on gamers' motivation to interact in FL/L2 (Peterson, 2012; Howard-Jones and Jay, 2016), we further posit that brain reward circuit may play a moderating role in the increased FC within gamers' brain network.

To date, very few studies have explored the neural mechanism underlying MMORPG's educational role in language learning. Only one recent study using resting-state functional magnetic resonance imaging (fMRI) investigated an educational MMORPG's effect on increasing learners' brain FC responsible for language processing (Hong et al., 2016). This study did not include control groups, which might make its conclusion less robust. Thus, a cohort study design is needed to ensure more tenable results. Additionally, the specific seed regions identified for FC analysis are also worth further discussion. Studies covered



**FIGURE 1 | Potential neural mechanism of Massive Multiple Online Role-Playing Games (MMORPGs') effect on foreign language (FL)/second language (L2) learning.** AI, anterior insula; FO, frontal operculum; VWFA, visual word form area; IFG, inferior frontal gyrus; SFG, superior frontal gyrus; VS, ventral striatum.

in this review have revealed that MMORPGs' assistant role in FL/L2 learning is realized by games' affordance of interaction, in which gamers/learners should frequently retrieve appropriate vocabulary from their memory to fulfill their real-time in-game interaction; moreover, they have to continuously and rapidly, in most cases, read the game-embedded texts and peers' real-time speech scrolling down the screen to move on smoothly. Such opportunities to develop reading and vocabulary skills are favored by MMORPG players (Peterson, 2011). Therefore, lexical retrieval and reading speed, two central aspects of language processing (Chai et al., 2016), seem to be essential in language learning emerged in playing MMORPGs. Lexical retrieval is linked to left anterior insular/frontal operculum (AI/FO; Perani et al., 2003; Damasio et al., 2004; Baldo et al., 2006), and reading speed is associated with visual word form area (VWFA; Gaillard et al., 2006; Nakamura et al., 2012). Thus, left AI/FO and VWFA can be taken as seed regions in the underlying FC. As for the location of other language areas to which the FC is computed from the seed regions, language processing-related areas in the neural substrates of gamers' attentional bias should be included. The two above-mentioned modes of attentional bias are controlled by two segregated networks of brain areas. The top-down mode recruits superior frontal gyrus (SFG) and intraparietal cortex, while the bottom-up mode recruits inferior frontal gyrus (IFG) and temporoparietal cortex (Corbetta and Shulman, 2002; Lee and Shomstein, 2014). Both the SFG and the IFG are closely related with language processing. The SFG is associated with language organization (Kinoshita et al., 2012), syntactic sequencing (Chan et al., 2013), speech initiation and spontaneity (Fujii et al., 2015), while the IFG relates to sentence comprehension (Friederici et al., 2003), phonological processing (Nixon et al., 2004), and semantic processing (Simard et al., 2013). The IFG and the SFG can be involved in the increased FC within gamers' brain network.

Regarding the identification of areas in the reward circuit, two central nodes involved are ventral striatum (VS) related to reward anticipation and ventromedial prefrontal cortex (vmPFC) related to reward outcome and subjective value (Knutson et al., 2003; Levy and Glimcher, 2012). However, the VS may contribute more to the neural mechanism under discussion, because game behavior associated with reward anticipation processing always takes much more time than reward attainment accompanied by outcome processing does. The potential neural mechanism is shown in **Figure 1**.

## CONCLUSION AND FUTURE STUDY

To our knowledge, this is the first review centering on both the MMORPGs' benefits for language learning and discussion of the behavioral and neural mechanisms underlying such benefits. When gamers/learners are immersed in a MMORPG environment, their existing attentional bias or the bias developed in their gaming and learning processes would make them allocate more cognitive resources toward task-related stimuli. Moreover, this reward-guided effect can be realized in a controlled or an automatic way by different types of gamers/learners. Language learning enhancement in playing MMORPGs may be realized by strengthening or increasing the FC from seed regions including the left AI/FO and the VWFA to other language processing-related areas, mainly including the IFG and the SFG. Further, MMORPGs' effect on the FC can be moderated by the activity of the VS in the brain reward circuit, which warrants further systematic study.

In future studies stroop or dot-probe task can be adopted to examine the existence of attentional bias among gamers/learners whose FL/L2 proficiency get improved after playing MMORPGs. For validation of the proposed neural mechanism, either the resting-state fMRI or task-state fMRI can be considered for experimental design. Besides, functional near-infrared spectroscopy technology is also a good alternative in view of its portability, less cost, good temporal and spatial resolution (Scherer et al., 2012), and its feasibility in investigating resting-state or task-state FC in the human language network (Molavi et al., 2014; Huang et al., 2016). If the proposed behavioral and neural mechanisms are confirmed, new evidence will be provided for MMORPGs' educational effect on FL/L2 learning. These new findings may promote the development of educational MMORPGs, and more importantly, pedagogical innovations can thereby be expected in the field of FL/L2 teaching.

## AUTHOR CONTRIBUTIONS

YZ and XZ designed this study. YZ wrote this paper. HS provided suggestions on the structure of this paper. XL, DT, and Y-eC contributed to the data collection.

## ACKNOWLEDGMENT

## REFERENCES

Anderson, B. A. (2015). Social reward shapes attentional biases. *Cogn. Neurosci.* 7, 30–36. doi: 10.1080/17588928.2015.1047823

Anderson, B. A., Laurent, P. A., and Steven, Y. (2013). Reward predictions bias attentional selection. *Front. Hum. Neurosci.* 7:262. doi: 10.3389/fnhum.2013.00262

Anderson, B. A., Laurent, P. A., and Yantis, S. (2011). Learned value magnifies salience-based attentional capture. *PLoS ONE* 6:e27926. doi: 10.1371/journal.pone.0027926

Anderson, B. A., and Yantis, S. (2013). Value-driven attentional capture. *Proc. Natl. Acad. Sci. U.S.A.* 39, 6.

Baldo, J. V., Schwartz, S., Wilkins, D., and Dronkers, N. F. (2006). Role of frontal versus temporal cortex in verbal fluency as revealed by voxel-based lesion symptom mapping. *J. Int. Neuropsychol. Soc.* 12, 896–900. doi: 10.1017/S1355617706061078

Berns, A., Palomo-Duarte, M., Dodero, J. M., and Valero-Franco, C. (2013). "Using a 3D online game to assess students' foreign language acquisition and communicative competence," in *Scaling up Learning for Sustained Impact*, Vol. 8095, eds D. Hernández-Leo, T. Ley, R. Klamma, and A. Harrer (Berlin: Springer), 19–31.

Bourgeois, A., Neveu, R., Bayle, D. J., and Vuilleumier, P. (2015). How does reward compete with goal-directed and stimulus-driven shifts of attention? *Cogn. Emot.* 31, 109–118. doi: 10.1080/02699931.2015.1085366

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

Bytheway, J. (2011). *Vocabulay Learning Strategies in Massively Multiplayer Online Role-Playing Games*. MA thesis, Victoria University of Wellington, Wellington, 32–67.

Bytheway, J. (2014). In-game culture affects learners' use of vocabulary learning strategies in massively multiplayer online role-playing games. *Int. J. Comput. Assist. Lang. Learn. Teach.* 4, 1–13. doi: 10.4018/ijcallt.2014100101

Cameron, J., Banko, K. M., and Pierce, W. D. (2001). Pervasive negative effects of rewards on intrinsic motivation: the myth continues. *Behav. Anal.* 24, 1–44.

Chai, X. J., Berken, J. A., Barbeau, E. B., Soles, J., Callahan, M., Chen, J. K., et al. (2016). Intrinsic functional connectivity in the adult brain and success in second-language learning. *J. Neurosci.* 36, 755–761. doi: 10.1523/JNEUROSCI.2234-15.2016

Chan, S. H., Ryan, L., and Bever, T. G. (2013). Role of the striatum in language: syntactic and conceptual sequencing. *Brain Lang.* 125, 283–294. doi: 10.1016/j.bandl.2011.11.005

Colby, R. S., and Colby, R. (2008). A pedagogy of play: integrating computer games into the writing classroom. *Comput. Compos.* 25, 300–312. doi: 10.1016/j.compcom.2008.04.005

Connolly, T. M., Boyle, E. A., Macarthur, E., Hainey, T., and Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Comput. Educ.* 59, 661–686. doi: 10.1016/j.compedu.2012.03.004

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Cruz, C., Hanus, M. D., and Fox, J. (2015). The need to achieve: players' perceptions and uses of extrinsic meta-game reward systems for video game consoles. *Comput. Hum. Behav.* 1–9. doi: 10.1016/j.chb.2015.08.017

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., and Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition* 92, 179–229. doi: 10.1016/j.cognition.2002.07.001

Deng, Z., Chandrasekaran, B., Wang, S., and Wong, P. C. M. (2015). Resting-state low-frequency fluctuations reflect individual differences in spoken language learning. *Cortex* 76, 63–78. doi: 10.1016/j.cortex.2015.11.020

Dourda, K., Bratitsis, T., Griva, E., and Papadopoulou, P. (2014). Content and language integrated learning through an online game in primary school: a case study. *Electron. J. e-Learn.* 12, 243–258.

Filsecker, M., and Hickey, D. T. (2014). A multilevel analysis of the effects of external rewards on elementary students' motivation, engagement and learning in an educational game. *Comput. Educ.* 75, 136–148. doi: 10.1016/j.compedu.2014.02.008

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. *Cerebral Cortex* 13, 170–177. doi: 10.1093/cercor/13.2.170

Fujii, M., Maesawa, S., Motomura, K., Futamura, M., Hayashi, Y., Koba, I., et al. (2015). Intraoperative subcortical mapping of a language-associated deep frontal tract connecting the superior frontal gyrus to broca's area in the dominant hemisphere of patients with glioma. *J. Neurosurg.* 122, 1390–1396. doi: 10.3171/2014.10.JNS14945

Gaillard, R., Naccache, L., Pinel, P., Clémenceau, S., Volle, E., Hasboun, D., et al. (2006). Direct intracranial, fmri, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron* 50, 191. doi: 10.1016/j.neuron.2006.03.031

González-González, C., and Blanco-Izquierdo, F. (2012). Designing social videogames for educational uses. *Comput. Educ.* 58, 250–262. doi: 10.1016/j.compedu.2011.08.014

Han, D. H., Bolo, N., Daniels, M. A., Arenella, L., Lyoo, I. K., and Renshaw, P. F. (2010). Brain activity and desire for Internet video game play. *Compr. Psychiatry* 52, 88–95. doi: 10.1016/j.comppsych.2010.04.004

Han, D. H., Sun, M. K., Bae, S., Renshaw, P. F., and Anderson, J. S. (2015). Brain connectivity and psychiatric comorbidity in adolescents with Internet gaming disorder. *Addict. Biol.* doi: 10.1111/adb.12347 [Epub ahead of print].

Hickey, C., Chelazzi, L., and Theeuwes, J. (2010). Reward changes salience in human vision via the anterior cingulate. *J. Neurosci.* 30, 11096. doi: 10.1523/JNEUROSCI.1026-10.2010

Hong, J. S., Han, D. H., Kim, Y. I., Su, J. B., Sun, M. K., and Renshaw, P. (2016). English language education on-line game and brain connectivity. *ReCALL* 29, 3–21. doi: 10.1017/S0958344016000173

Howard-Jones, P. A., and Jay, T. (2016). Reward, learning and games. *Curr. Opin. Behav. Sci.* 10, 65–72. doi: 10.1016/j.cobeha.2016.04.015

Hsu, L. (2015). EFL learners' implicit theory of intelligence and the application of MMORPG in EFL learning. *Int. J. Comput. Assist. Lang. Learn. Teach. (IJCALLT)* 5, 58–71. doi: 10.4018/IJCALLT.2015040104

Hu, M. M., and Chang, B. (2007). "Massively multiplayer online game supported foreign language listening ability training," in *Proceedings of the First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL)*, Banff, AB, 176–178.

Huang, B. G., and Yang, J. C. (2014). "The effects of prior knowledge for incidental vocabulary acquisition on multiplayer online role-playing game," in *Proceedings of the 13th International Conference on Web-based Learning (ICW)*, Tallinn, 98–105.

Huang, C. J., Chou, P. H., Wei, H. L., and Sun, C. W. (2016). Functional connectivity during phonemic and semantic verbal fluency test: a multichannel near infrared spectroscopy study. *IEEE J. Sel. Top. Quantum Electron.* 22, 1. doi: 10.1109/jstqe.2015.2503318

Iten, N., and Petko, D. (2016). Learning with serious games: is fun playing the game a predictor of learning success? *Br. J. Educ. Technol.* 47, 151–163. doi: 10.1111/bjet.12226

Jacob, H., Brück, C., Domin, M., Lotze, M., and Wildgruber, D. (2014). I can't keep your face and voice out of my head: neural correlates of an attentional bias toward nonverbal emotional cues. *Cereb. Cortex* 24, 1460. doi: 10.1093/cercor/bhs417

Khatibi, E., and Cowie, E. (2013). *Language Learning through Interactive Games*. Malmo: Malmo Hogskola University 22–27.

Kim, P. W., Kim, S. Y., Shim, M., Im, C. H., and Shon, Y. M. (2013). The influence of an educational course on language expression and treatment of gaming addiction for massive multiplayer online role-playing game (MMORPG) players. *Comput. Educ.* 63, 208–217. doi: 10.1016/j.compedu.2012.12.008

Kinoshita, M., Shinohara, H., Hori, O., Ozaki, N., Ueda, F., Nakada, M., et al. (2012). Association fibers connecting the broca center and the lateral superior frontal gyrus: a microsurgical and tractographic anatomy. *J. Neurosurg.* 116, 323–330. doi: 10.3171/2011.10.JNS11434

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., and Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *Neuroimage* 18, 263–272. doi: 10.1016/S1053-8119(02)00057-5

Kongmee, I., Strachan, R., Pickard, A., and Montgomery, C. (2011). "Moving between virtual and real worlds: Second language learning through Massively Multiplayer Online Role Playing Games (MMORPGs)," in *Proceedings of the 3rd Computer Science and Electronic Engineering Conference*, Colchester, 13–18.

Kongmee, I., Strachan, R., Pickard, A., and Montgomery, C. (2012). A case study of using online communities and virtual environment in massively multiplayer role playing games (MMORPGs) as a learning and teaching tool for second language learners. *Int. J. Virtual Pers. Learn. Environ.* 3, 1–15. doi: 10.4018/jvple.2012100101

Lai, A. F., and Wen, S. S. (2012). "Evaluating an online role playing game for promoting the elementary school students' English speaking abilities," in *Proceedings of the 8th International Conference on Information Science and Digital Content Technology*, Daegu, 610–615.

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., and Wills, A. J. (2016). Attention and associative learning in humans: an integrative review. *Psychol. Bull.* 142, 1111–1140. doi: 10.1037/bul0000064

Lee, J., and Shomstein, S. (2014). Reward-based transfer from bottom-up to top-down search tasks. *Psychol. Sci.* 25, 466–475. doi: 10.1177/0956797613509284

Levy, D. J., and Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Curr. Opin. Neurobiol.* 22, 1027–1038. doi: 10.1016/j.conb.2012.06.001

Lucas, N., Schwartz, S., Leroy, R., Pavin, S., Diserens, K., and Vuilleumier, P. (2013). Gambling against neglect: unconscious spatial biases induced by reward reinforcement in healthy people and brain-damaged patients. *Cortex* 49, 2616–2627. doi: 10.1016/j.cortex.2013.06.004

Luijten, M., Veltman, D. J., Hester, R., Smits, M., Pepplinkhuizen, L., and Franken, I. H. (2012). Brain activation associated with attentional bias in smokers is modulated by a dopamine antagonist. *Neuropsychopharmacology* 37, 2772–2779. doi: 10.1038/npp.2012.143

Milton, J., Jonsen, S., Hirst, S., and Lindenburn, S. (2012). Foreign language vocabulary development through activities in an online 3D environment. *Lang. Learn. J.* 40, 99–112. doi: 10.1080/09571736.2012.658229

Molavi, B., May, L., Gervain, J., Carreiras, M., Werker, J. F., and Dumont, G. A. (2014). Analyzing the resting state functional connectivity in the human language system using near infrared spectroscopy. *Front. Hum. Neurosci.* 7:921. doi: 10.3389/fnhum.2013.00921

Nagle, A., Wolf, P., Riener, R., and Novak, D. (2014). The use of player-centered positive reinforcement to schedule in-game rewards increases enjoyment and performance in a serious game. *Int. J. Serious Games* 1, 35–47. doi: 10.17083/ijsg.v1i4.47

Nakamura, K., Kuo, W. J., Pegado, F., Cohen, L., Tzeng, O. J., and Dehaene, S. (2012). Universal brain systems for recognizing word shapes and handwriting gestures during reading. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20762–20767. doi: 10.1073/pnas.1217749109

Neville, D. O. (2010). Structuring narrative in 3D digital game–based learning environments to support second language acquisition. *Foreign Lang. Ann.* 43, 446–469. doi: 10.1111/j.1944-9720.2010.01092.x

Neville, D. O. (2015). The story in the mind: the effect of 3D gameplay on the structuring of written l2 narratives. *ReCALL* 27, 21–37. doi: 10.1017/S0958344014000160

Nixon, P., Lazarova, J., Hodinotthill, I., Gough, P., and Passingham, R. (2004). The inferior frontal gyrus and phonological processing: an investigation using rtms. *J. Cogn. Neurosci.* 16, 289–300. doi: 10.1162/089892904322984571

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., et al. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: an fmri study during verbal fluency. *Hum. Brain Mapp.* 19, 170–182. doi: 10.1002/hbm.10110

Peterson, M. (2010). Massively multiplayer online role-playing games as arenas for second language learning. *Comput. Assist. Lang. Learn.* 23, 429–439. doi: 10.1080/09588221.2010.520673

Peterson, M. (2011). Digital gaming and second language development: Japanese learners interactions in a MMORPG. *Digit. Cult. Educ.* 38, 289–299.

Peterson, M. (2012). Learner interaction in a massively multiplayer online role playing game (MMORPG): a sociocultural discourse analysis. *ReCALL* 24, 361–380. doi: 10.1017/S0958344012000195

Petry, N. M., and O'Brien, C. P. (2013). Internet gaming disorder and the DSM-5. *Addiction* 108, 1186–1187. doi: 10.1111/add.12162

Pierce, W. D., Cameron, J., Banko, K. M., and So, S. (2003). Positive effects of rewards and performance standards on intrinsic motivation. *Psychol. Rec.* 53, 561–578.

Rama, P. S., Black, R. W., Van Es, E., and Warschauer, M. (2012). Affordances for second language learning in world of Warcraft. *ReCALL* 24, 322–338. doi: 10.1017/S0958344012000171

Rankin, Y. A., Mcneal, M. K., Shute, M. W., and Gooch, B. (2008). "User centered game design: evaluating massive multiplayer online role playing games for second language acquisition," in *Proceedings of the 2008 ACM SIGGRAPH Symposium on Video Games*, Los Angeles, CA, 43–49.

Rankin, Y. A., Morrison, D., McNeal, M., Gooch, B., and Shute, M. W. (2009). "Time will tell: in-game social interactions that facilitate second language acquisition," in *Proceedings of the 4th International Conference on Foundations of Digital Games*, ed. R. Young (New York, NY: ACM), 161–168. doi: 10.1145/1536513.1536546

Scherer, L. C., Fonseca, R. P., Amiri, M., Adrover-Roig, D., Marcotte, K., Giroux, F., et al. (2012). Syntactic processing in bilinguals: an fNIRS study. *Brain Lang.* 121, 144–151. doi: 10.1016/j.bandl.2011.09.009

Scholz, K. W. (2015). *Online Digital Game-Based Language Learning Environments: Opportunities for Second Language Development.* Ph.D. thesis, University of Waterloo, Waterloo, ON, 52–68.

Shahriarpour, N., and Kafi, Z. (2014). On the effect of playing digital games on Iranian intermediate efl learners' motivation toward learning English vocabularies. *Procedia Soc. Behav. Sci.* 98, 1738–1743. doi: 10.1016/j.sbspro.2014.03.601

Shomstein, S., Lee, J., and Behrmann, M. (2010). Top-down and bottom-up attentional guidance: investigating the role of the dorsal and ventral parietal cortices. *Exp. Brain Res.* 206, 197–208. doi: 10.1007/s00221-010-2326-z

Simard, F., Monetta, L., Nagano-Saito, A., and Monchi, O. (2013). A new lexical card-sorting task for studying fronto-striatal contribution to processing language rules. *Brain Lang.* 125, 295–306. doi: 10.1016/j.bandl.2011.08.002

Sublette, V. A., and Mullan, B. (2012). Consequences of play: a systematic review of the effects of online gaming. *Int. J. Ment. Health Addict.* 10, 3–23. doi: 10.1007/s11469-010-9304-3

Suh, S., Kim, S. W., and Kim, N. J. (2010). Effectiveness of MMORPG-based instruction in elementary English education in Korea. *J. Comput. Assist. Learn.* 26, 370–378. doi: 10.1177/1059840512449653

Sundqvist, P., and Sylvén, L. K. (2012). "World of VocCraft: computer games and Swedish learners' L2 vocabulary," in *Computer Games in Language Learning and Teaching*, ed. H. Reinders (Basingstoke: Palgrave Macmillan), 189–208. doi: 10.1057/9781137005267.0016

Thalemann, R., Wolfling, K., and Grusser, S. (2007). Specific cue reactivity on computer game-related cues in excessive gamers. *Behav. Neurosci.* 121, 614–618. doi: 10.1037/0735-7044.121.3.614

Thorne, S. L. (2008). "Transcultural communication in open internet environments and massively multiplayer online games," in *Mediating Discourse Online*, ed. S. Magnan (Amsterdam: John Benjamins), 305–327. doi: 10.1075/aals.3.17tho

Thorne, S. L., Cornillie, F., and Desmet, P. (2012a). ReCALL special issue: digital games for language learning: challenges and opportunities. *ReCALL* 24, 243–256. doi: 10.1017/S0958344012000134

Thorne, S. L., Fischer, I., and Lu, X. (2012b). The semiotic ecology and linguistic complexity of an online game world. *ReCALL* 24, 279–301. doi: 10.1017/S0958344012000158

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes.* Cambridge, MA: Harvard University Press, 86.

Wei, T., Liang, X., He, Y., Zang, Y., Han, Z., Caramazza, A., et al. (2012). Predicting conceptual processing capacity from spontaneous neuronal activity of the left middle temporal gyrus. *J. Neurosci.* 32, 481–489. doi: 10.1523/JNEUROSCI.1953-11.2012

Wu, M. L., and Richards, K. (2012). "Massively multiplayer online role-playing games as digital game-based english learning platforms: a study of the effects of digital game play on ESL students' english use," in *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, Chesapeake, VA, 1370–1373.

Yang, J. C., and Hsu, H. F. (2013). "Effects of online gaming experience on English achievement in an MMORPG learning environment," in *Proceedings of the WorldCALL 2013 Conference*, Glasgow, 379–381.

Yudintseva, A. (2015). Game-enhanced second language vocabulary acquisition strategies: a systematic review. *Open J. Soc. Sci.* 3, 101–109. doi: 10.4236/jss.2015.310019

Zhang, Y. J., Yamikani, N., Hou, J., Chen, J. W., Yang, L. Z., Wang, Y., et al. (2016). Cue-induced behavioral and neural changes among excessive Internet gamers and possible application of cue exposure therapy to Internet gaming disorder. *Front. Psychol.* 7:675. doi: 10.3389/fpsyg.2016.00675

Zheng, D., Bischoff, M., and Gilliland, B. (2015). Vocabulary learning in massively multiplayer online games: context and action before words. *Educ. Technol. Res. Dev.* 63, 771–790. doi: 10.1007/s11423-015-9387-4

Zheng, D., Newgarden, K., and Young, M. F. (2012). Multimodal analysis of language learning in World of Warcraft play: languaging as values-realizing. *ReCALL* 24, 339–360. doi: 10.1017/S0958344012000183

Zheng, D., Young, M. F., Wagner, M., and Brewer, B. (2009). Negotiation for action: english language learning in game-based virtual worlds. *Mod. Lang. J.* 93, 489–511. doi: 10.1111/j.1540-4781.2009.00927.x

# A Cultural Evolution Approach to Digital Media

*Alberto Acerbi* *

*School of Innovation Science, Eindhoven University of Technology, Eindhoven, Netherlands*

Digital media have today an enormous diffusion, and their influence on the behavior of a vast part of the human population can hardly be underestimated. In this review I propose that cultural evolution theory, including both a sophisticated view of human behavior and a methodological attitude to modeling and quantitative analysis, provides a useful framework to study the effects and the developments of media in the digital age. I will first give a general presentation of the cultural evolution framework, and I will then introduce this more specific research program with two illustrative topics. The first topic concerns how cultural transmission biases, that is, simple heuristics such as "copy prestigious individuals" or "copy the majority," operate in the novel context of digital media. The existence of transmission biases is generally justified with their adaptivity in small-scale societies. How do they operate in an environment where, for example, prestigious individuals possess not-relevant skills, or popularity is explicitly quantified and advertised? The second aspect relates to fidelity of cultural transmission. Digitally-mediated interactions support cheap and immediate high-fidelity transmission, in opposition, for example, to oral traditions. How does this change the content that is more likely to spread? Overall, I suggest the usefulness of a "long view" to our contemporary digital environment, contextualized in cognitive science and cultural evolution theory, and I discuss how this perspective could help us to understand what is genuinely new and what is not.

Keywords: cultural evolution, cultural transmission, transmission biases, cultural attraction, digital media, social media

## 1. INTRODUCTION

Digital media are media encoded in digital format, typically to be transmitted and consumed on electronic devices, such as computers and smartphones. Digital media of wide diffusion includes emails, digital audio and video recordings, ebooks, blogs, instant messaging, and more recently social media. Although, digital media started to be developed with the creation of digital computers in the 1940s, their wide cultural impact can be traced back only to two or three decades, with the widespread diffusion of personal computers and especially the internet (Briggs and Burke, 2009).

Social media and ubiquitous connectivity (e.g., allowed by portable digital devices) are even more recent developments. Facebook, in its early stage limited to university or high-school students and employees of a handful of companies, was open to the public 10 years ago, in September 2006 (Boyd and Ellison, 2007). The first version of the iPhone, which gave the initial momentum to the worldwide diffusion of smartphones, was launched shortly after, at the beginning of 2007 (West and Mace, 2010).

Despite that, digital media, and social media in particular, have today an enormous reach. Facebook for example counts, as of June 2016, more than 1.7 billion monthly active users[1]. The influence of digital media on the behavior of a vast part of the human population is unanimously recognized. As a consequence, academic interest for digital media has grown rapidly in different disciplines. Here, I will not attempt a review of the existing literature, but I will propose that a specific scientific field, cultural evolution, could provide a suitable framework to analyse how the massive diffusion of digital media influences human cultural behavior.

The article is structured as follows. In the next section I will provide a brief and general introduction to the field of cultural evolution, focusing on the aspects I consider more relevant for the study of contemporary digital media. These aspects are cultural evolution's naturalistic and quantitative approach and its commitment to develop hypotheses informed by cognitive science and evolutionary theory. I will then explore more in depth two areas of research where cultural evolution could give an original contribution. First, I will discuss how cultural transmission biases, i.e., simple rules such as "copy the majority" or "copy prestigious individuals," a central topic in cultural evolutionary research, might influence cultural transmission in the digital age, and conversely how digitally-supported cultural transmission might disrupt these biases. I will explore at some length two of these biases, related to prestige and popularity. Second, I will examine how cultural evolutionary dynamics could be influenced by the fact that digitally-supported cultural transmission allows virtually error-free propagation of cultural traits. I will conclude suggesting that the cultural evolution framework places the digital age in a broader context, and I will discuss how this theoretical and historical "long view" could help us to better understand the changes we are confronted with in our society.

## 2. CULTURAL EVOLUTION

Cultural evolution is a relatively recent scientific field that studies human and, partly, non-human cultural behavior (see Mesoudi, 2015, for a recent review). Cultural behavior is generally defined as behavior transmitted through social learning, as opposed to individual learning or genetic inheritance (Henrich and McElreath, 2003). The distinction between cultural and non-cultural behavior is not a sharp one (Morin, 2015) but it works quite well for practical purposes. Cultural evolutionists study things such as the evolution of uniquely human forms of cooperation (Boyd and Richerson, 2009; Turchin et al., 2013), indigenous knowledge of plants' properties (Reyes-Garcia et al., 2008), the cultural evolution of language (Tamariz et al., 2014; Kirby et al., 2015), the spread of fashions in contemporary culture, using cases like baby names (Bentley et al., 2004) or dog breeds (Ghirlanda et al., 2013, 2014), or how ineffective medical treatments can nonetheless be successful (Tanaka et al., 2009; de Barra et al., 2014; Miton et al., 2015), just to give a few examples. Similarly, a wide range of methodologies are

---

[1]https://newsroom.fb.com/company-info/

used, including simulation and mathematical models (Acerbi et al., 2009; Kempe et al., 2014; Smaldino and McElreath, 2016), laboratory experiments (Caldwell and Smith, 2012; Derex and Boyd, 2015; Muthukrishna et al., 2015; Schillinger et al., 2016), phylogenetic analysis (Fortunato and Jordan, 2010; Tehrani, 2013; Watts et al., 2015), ethnographic research (Mathew and Boyd, 2014; Colleran and Mace, 2015), and comparative studies of social learning in humans and other animals (Whiten et al., 2009; Dean et al., 2012; Reindl et al., 2016).

What brings together all these researches is, more than a unitary view about how culture should be considered an evolutionary process (see Claidière et al., 2014; Acerbi and Mesoudi, 2015; Lewens, 2015, for a general discussion), a strong commitment to provide explanations that are naturalistic and quantitative, as well as grounded in cognitive science and evolutionary theory. At the minimum, all cultural evolutionists share the idea that a cultural phenomenon is a population-level aggregate of individual-level interactions and that, to explain the former, one needs to take seriously the latter. Accordingly, the works of Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985) are considered as establishing modern cultural evolution. These works consisted in mathematical models, inspired by population genetics, developing formalisms to link micro-processes of transmission—like different "directions" of transmission, e.g., from parents to offsprings, between peers, etc. or different transmission biases, see below—to macro-processes of cultural change—like the diffusion dynamics of cultural traits. In parallel, cognitive anthropologists such as Sperber (1985, 1996) started to consider in depth the role of individual cognition in the explanation of cultural patterns, focusing on the fact that the success of some widespread beliefs may depend on them being generally attractive to human minds (I will discuss some examples in the next sections).

The psychology of digital media, in particular online activities (sometimes described as "cyberpsychology" Attrill, 2015) is a growing field (see e.g., Wallace, 2001; Suler, 2015). A cultural evolution approach adds, as mentioned, an explicit interest for the micro-macro link, in other words, for how individual-level properties (e.g., psychological) influence population-level dynamics and vice versa. In addition, the naturalistic and quantitative framework provided by cultural evolution seems perfectly suited for the study of contemporary digital media. One of the opportunities that the widespread diffusion of digital media offers to social sciences is the availability of vast amounts of data on human behavior (Lazer et al., 2009). While the understanding offered by ethnographic (e.g., Boyd, 2014) or critical-theory-inspired (e.g., Fuchs, 2014) perspectives remain clearly important, the cultural evolution approach is in a better position to make sense also of the quantitative data that digital media usage quasi-automatically produces. On the other side, computer scientists and physicists had promptly made use of these data to study the diffusion of information in digital social networks (see Weng et al., 2012; Adamic et al., 2014; Cooney et al., 2016; Del Vicario et al., 2016, for few recent examples). These works importantly include quantitative analysis and models, and they can offer valuable insights on online activity. However, the perspective of cultural evolution can

complement this thread of research by providing a refined view of the micro-processes of transmission and of the psychological motivations underpinning them.

To sum up, cultural evolution may offer a privileged perspective to look at digital media, including both a sophisticated view of human behavior and a methodological attitude to modeling and quantitative analysis. In the next sections I will try to substantiate this claim with some examples of investigations that a cultural evolution approach suggests.

## 3. TRANSMISSION BIASES IN THE DIGITAL AGE

For the majority of cultural evolutionists the widespread utilization of social learning is the reason of the ecological success of the human species (Henrich, 2016). Social learning provides a shortcut to long and potentially dangerous individual learning and a fast and flexible alternative to genetic evolution. However, simply copying from others can be risky: to be effective, social learning needs to be selective (Laland, 2004). According to this view, social learning is made possible by domain-general heuristics—often referred to as "transmission biases" or "social-learning strategies"—helping us to choose what, when, and from whom to learn (Boyd and Richerson, 1985). To use a mundane example, imagine you find yourself in a new and unknown town, searching a restaurant for dinner. You may first decide that is worth to look to what others do, instead of trying to figure it out by yourself ("copy when asocial learning is costly"), and then that it does not make much sense to follow the first person you see in the street, but look for restaurants that seem full of customers ("copy the majority"). After few days, you might have found your favorite place, and you can stop to check where other people go ("copy when uncertain").

Transmission biases are a good place to start as much research has been developed in cultural evolution on this topic. Theoretical models and simulations have explored the adaptive value of different biases, and predictions from the models have been tested in empirical settings (see Rendell et al., 2011, for a review). In parallel, various works have attempted to detect the presence of transmission biases in real-life cultural dynamics (e.g., Reyes-Garcia et al., 2008; Henrich and Broesch, 2011; Kandler and Shennan, 2013; Acerbi and Bentley, 2014). Importantly, for our focus on digital media, transmission biases are considered a suite of psychological adaptations shaped by natural selection (Henrich, 2016), hence generally effective in the social and physical environment of small-scale societies. A question only partially explored in cultural evolution is how these biases scale in contemporary, complex, societies, and especially in the novel digital environment.

### 3.1. Prestige
Various heuristics are available when choosing from whom to copy. From an evolutionary point of view, for example, kin share a common genetic interest, so they will be willing to circulate useful information. Copying from parents and from other close members of the family makes thus perfect sense. Elders, especially

in small-scale and slow-changing societies, have two important qualities. First, they had time to learn themselves a substantial part of the cultural repertoire of the society, and, second, they must have done it effectively, exactly because they arrived to old age. Age-biased social learning is thus another evolutionary expected strategy (Henrich, 2016).

However, for specialized expertises (i.e., only few people possess them), or for expertises that exhibit variability in a population (i.e., some people are very good at them and others are not), kin- and age- based strategies are not particularly effective. In these cases, an alternative is to try to assess directly the ability of others. Copying skilled or successful individuals is then another of the heuristics suggested by cultural evolutionists (see e.g., Mesoudi, 2011, for an experimental approach). This strategy presents, in turn, another problem. Skills can be opaque, difficult to recognize, and this is especially true when one does not possess the expertise in question, which is exactly the case when there is the need to learn it. Similarly, success can be volatile, or due to luck. How many successful hunts an apprentice hunter should assess before deciding to copy from a particular individual and not from another?

A possible solution is prestige-biased social learning. Cultural evolutionist Joe Henrich defines prestige cues as a "second-order cultural learning" (Henrich, 2016, p. 45): one can make use of signs of deference, respect, or simply check from whom other people are learning, and choose those individuals as cultural models. The risk, with prestige-biased social learning, is that prestige and skills may not correlate. What if an individual is prestigious because of his hunting abilities, but I am attempting to learn how to build harpoons? What if an individual is prestigious because he belongs to an influential family, but he does not possess any particular skill? The answer is that in small-scale societies this is a minor problem. Specialization and inequality are limited, so that respected individuals will indeed be, on average, generally skilled.

Of course, the situation is different today. Our reliance on celebrities, for example in advertisement, is generally considered a good candidate for a cultural evolutionary mismatch (Henrich, 2016). The acting abilities of George Clooney are unlikely to correlate with his expertise in coffee-tasting, still, the story goes, the success of a Nestlé brand of coffee depends on the presence of the actor in the advertisements. Internet and in particular social media would possibly push things even further, because the rapidity of communications and of the extension and the number of the virtual communities. The real risk for the society is not much that we end up to parrot the—alleged—favorite coffee brand of celebrities, but that social media users will attempt to copy skills that are not existent at all (such as Clooney's coffee tasting ability) or existent, but not relevant in the local environment (such as Clooney's acting ability). More worryingly, extremist groups could make use, consciously or not, of prestige-biased influence mechanisms for on-line proselytism (Barkow et al., 2012)[2]. These ideas could be tested empirically, but, to my knowledge, not much research has been done yet.

---

[2]See also: http://www.cato-unbound.org/2016/02/08/jerome-h-barkow/how-internet-subverts-cultural-transmission

One could examine whether usage of internet and social media correlates with higher preferential attention to "global" cues of prestige (as opposed to "local" ones), possibly taking into account confounding factors such as the exposition to traditional mass-media, like television or cinema. In addition, attention to global cues of prestige does not need to be harmful, especially in a fast-changing and deeply interconnected society. Although, it might be argued that acting abilities are not necessarily relevant, the same digital media allow to access also to prestigious surgeons, programmers, or philanthropists in a way that would not be possible in a local environment.

Research on social media "influencers" is in its infancy, and results are not conclusive (see Bakshy et al., 2011; Aral and Walker, 2012, and the studies reviewed therein). Bakshy et al. (2011), for example, measured how links to webpages posted in Twitter spread in the social media itself, and found that, indeed, users with more followers and who have been already influential in the past tended to produce larger "cascades." However, it is not clear how to distinguish the fact that the number of followers is a sign of prestige, in the cultural evolution meaning, from the fact that, at the same time, it indicates how many individuals are exposed to the link. In this sense, the effect could be simply due to a larger number of possible events of transmission. Even not considering this confounding, Bakshy et al. (2011) comment that, given that cascades-sizes are power-law distributed (i.e., there are very few large cascades, while the majority of links are never reposted), "individual-level predictions of influence nevertheless remain relatively unreliable." They thus proceed to analyse the contribute of the actual content of the links tweeted, showing that content independently rated as more interesting and positive generated larger cascades. These findings resonate with theoretical results showing that wide-ranging events of diffusion of traits in networks are favored less by influencers than by the presence of large masses of easily influenceable individuals (Watts and Dodds, 2007).

The same celebrity influence is, at least in cultural evolution literature, mainly anecdotal, and marketing studies show that the effect of celebrities in advertisements is mediated by various cues, such as their relationship with the product advertised (see e.g., Kelting and Rice, 2013). We do not know, for every George Clooney, how many advertisements with celebrities did not succeed (Stephen Hawking, for example, was featured in the early 2000s in a high-profile campaign for an online fund platform that closed in 2004), and how many campaigns succeed without the presence of a celebrity. Moreover, as the results from Bakshy et al. (2011) suggest, there is an interaction between content and prestige. An interesting possibility is that relatively low-cost alternatives, like which coffee brand to choose or which haircut, could be celebrity-biased, but the effect would be less important for high-cost choices. This would mean that prestige-biased epidemics of extremism might not be such a realistic danger. On the other side, Clooney would not be probably able to persuade smokers to quit, for example.

In sum, although we have some convincing evidences of the effect of prestige-biased social learning in small-scale societies (Henrich and Broesch, 2011) and from laboratory experiments (Atkisson et al., 2012; Chudek et al., 2012), the question of how automatic is the influence of digital media's "influencers" in contemporary society remains open. Morin (2015) writes of "flexible imitators" that selectively use social—such as prestige—or asocial cues, depending on various factors, e.g., the above mentioned cost of the alternatives. Others (Heyes, 2016b) suggest that, at least in some circumstances, human social learning strategies are explicitly metacognitive. This means that these strategies include adjustable learning targets, changing from situation to situation, such as "copy digital natives," referring to copying knowledgeable young persons in the specific domain of technology, instead of a general rule "copy young individuals" (Heyes, 2016a).

In this case, like in the others we will explore in the next sections, the cultural evolutionary approach suggests a perspective from which to look to digital media and a series of questions that might be addressed in further research. What is the difference between the usage of prestige cues in small-scale societies and in our contemporary digital environment? What are the differences between local prestige, as in the case of small-scale societies or in contemporary circles of friends, and global prestige, as in the case of celebrities? Is prestige modulated by content? We already mentioned a possible difference between high-cost and low-cost choices; another could be related to the presence or absence of previous knowledge: real coffee connoisseurs might be less impressed by Clooney's approval.

## 3.2. Popularity

A similar way of reasoning can be applied to frequency-dependent biases. In the idiom of cultural evolution, frequency-dependent biases are heuristics that make use of the estimated frequency of a cultural trait to help deciding whether to copy it or not. The usefulness of positive, i.e., preferences for popular traits, frequency-dependent biases is easy to understand. When in a new environment, or when confronted with a new technology, it makes sense to take advantage of the cumulative experience of other individuals.

When cultural evolutionists talk about positive frequency-dependent biases, they generally refers to "conformity" in a precise and quite restrictive sense, meaning a disproportionate tendency to copy from the majority (Boyd and Richerson, 1985). This means that, returning to our restaurants example, if 60 people are eating in restaurant A and 40 people in restaurant B, the probability to choose A should be *higher than 60%* in conformist-biased social learning. In fact, it has been noted that, in almost all cases, social learning imply to "follow the majority" in a loose sense (Boyd and Richerson, 1985). In the above case, for example, one individual would still be more likely to go to restaurant A without any particular bias, i.e., copying randomly (imagine to ask to a random person where she was for dinner and follow her advice: your probability to go to restaurant A will be 60%).

This over-response to frequency information (Efferson et al., 2008) has a special importance for cultural evolution. First, it has been shown to contribute to maintain culturally homogenous groups, despite certain levels of migrations and individual variations (see e.g., Boyd and Richerson, 2009). Second, it allows to directly "jump" to the best alternative in presence of noisy

information (Henrich, 2016). In what follows I will thus use the more generic term "popularity bias" to indicate that the perception of something as popular makes it preferable to other—less popular—cultural traits, and I will reserve the usage of the term "conformity" for the technical sense described above. Finally, "social influence" simply means that people copy, without any bias, the choices of others.

As in the case of prestige, it is important to draw an explicit comparison between the conditions in which a psychological bias implementing a preference for popular cultural traits could have evolved and today's digital age. The first interesting aspect is that, in a small-scale and perhaps illiterate society, popularity needs to be estimated from various cues. The situation with digital media appears clearly different. Popularity is quantified and explicitly made public—the number of Facebook "likes" or "share," the number of Twitter "retweets," etc.—in practically all digital platforms. While one could speculate whether the success of this practice might be due to a universal sensitivity to this kind of information, as a cultural evolution perspective would suggest, it is not clear what kind of effect this could have on cultural transmission patterns. One possibility is that such low-cost availability of popularity signals would discourage individual exploration, prompting people to follow cheap social cues (Derex and Boyd, 2015), with digital media amplifying the effect of popularity-biased cultural transmission.

For example, success in digital media, especially regarding internet websites, has been repeatedly described as following a power-law distribution (as mentioned in the previous section for the links posted on Twitter). Power-law distributions are typical of winner-take-all markets, with very few websites monopolizing visitors whereas the vast majority remains relatively unsuccessful (Adamic and Huberman, 2000). However, it is useful to remind that power-law distributions are not necessarily generated by popularity-biased dynamics, as defined above. Power-law distributions naturally arise with unbiased social influence, because simply copying at random amplifies small initial differences. In fact, cultural evolutionary studies have shown that power-law distributions are present in many domains where social influence is important, such as baby names, dog breeds, scientific citations (Bentley et al., 2004), or even decoration styles in neolithic pottery (Neimann, 1995), where one can safely exclude the influence of digital media. The tell-tale of a positive-frequency-dependent bias is a distribution that is even more skewed in favor of successful items than power-laws (Mesoudi and Lycett, 2009).

In addition it is difficult, when not impossible without additional data, to set apart the effect of social influence and the effect of the intrinsic quality of the items in creating these skewed distributions (Aral and Walker, 2012; Muchnik et al., 2013; Morin, 2015). Ghirlanda et al. (2013), trying to deal with this problem, examined the case of dog breeds popularity. They showed that desired characteristics of breeds, such as trainability or good health, were not correlated with their success. This suggests that, in this specific domain, the role of popularity, or simply social influence, is more important than the

intrinsic characteristics of the cultural traits, i.e., the dog breeds themselves.

Some studies manipulated directly the perceived popularity of items in digital media, trying to detect the effect on their subsequent success. In a recent experiment, Muchnik et al. (2013) assigned randomly more than 100,000 comments submitted to a website with a structure similar to Reddit to three treatment groups: up-treated (comments were artificially given a +1 rating at their creation), down-treated (comments were artificially given a −1 rating at their creation), and control. Up-treated comments were indeed more likely to be subsequently up-voted than control. Down-treated comments were, as expected, more likely to be subsequently down-voted than comments in the control group. However, they were up-voted to a greater extent, so that the net effect was slightly positive, even if not significant with respect to the control group, as if users of the website tended to counterbalance negative comments. Muchnik et al. (2013) explain their results as due to an increasing turnout (i.e., up- or down-treated comments generated overall more ratings than comments in the control group) coupled with a common preference for positive ratings.

In a previous large-scale experiment, Salganik et al. (2006) created a digital "artificial market" where subjects could listen to and download unknown songs. Participants in the social influence condition could see how many times a song was downloaded previously, and they were randomly assigned to one of eight "worlds" where the counts of download were evolving independently. Salganik et al. (2006) showed that the social influence condition created more inequality (defined as difference between successful and unsuccessful songs) and unpredictability (defined as the difference between songs' results in the different worlds) with respect to the independent condition, where participants did not have information on previous download. Interestingly, two forms of visual presentation were proposed to participants in the social influence condition: in the first, the songs were presented in the same configuration of the independent condition, simply adding the number of previous downloads, and in the second they were presented as an ordered list, with the most downloaded on the top. Social influence was noticeably stronger in the latter case (more on this below).

Unpredictability, however, was not complete: there was a significant correlation between the perceived quality of the songs, as measured in the independent condition, and their success in the social influence condition or as Salganik et al. (2006) put it: "in general, the "best" songs never do very badly, and the "worst" songs never do extremely well." Given that choices (downloading or not a song) were extremely low-cost for the participants and the fact that the songs were previously unknown, the effect of popularity seems relatively limited in this experiment (Lewens, 2015; Morin, 2015, argument more thoroughly for a similar interpretation of these results). In a follow up study, the manipulations were stronger, such as completely reversing the perceived popularity order of the songs, i.e., presenting as the most popular the "worst" song of the independent condition, and so on (Salganik and Watts, 2008). Again, however, the best songs

tended to recover their popularity in the long run. Moreover, strong distortions of the correlation between intrinsic appeal and popularity were intuitively perceived by the participants, as showed by the fact that they resulted in fewer downloads overall. As above, the effect of popularity seems to be more nuanced that what an intuitive, clear-cut, understanding would suggest.

A more extreme version of the explicit advertisement of popularity cues is the proliferation of "top-N" lists. The spreading of top-lists predates digital media, and it is almost an hallmark of the broadcast era (in the United Kingdom the first introduction of a top-chat program in BBC radio dates back to 1957[3]), but it reached enormous diffusion in the recent years, with on-line top-lists of virtually everything. From a cultural evolution perspective, top-lists are not only sources of cheap estimates of popularity, but they also supply a direct way to implement a variant of the above mentioned conformist-bias, giving disproportionate publicity to already popular items (Acerbi and Bentley, 2014). The presentation of alternatives in form of top-lists, or ranked tables, do seem to enhance popularity influence (Salganik et al., 2006).

Another, more elaborate, variant of popularity displays is represented by the spreading of information in form of consumer—as opposed to "expert"—reviews, whether as a part of commercial websites (such as Amazon), or through websites specifically dedicated to reviews (such as Tripadvisor, Yelp, etc.). The positive economic effect of favorable reviews has been shown in several domains, including books (Chevalier and Mayzlin, 2006), restaurants (Luca, 2011), or hotels (Ye et al., 2009). The where-to-go-to-dinner example I used to illustrate cultural transmission biases looks rather outdated nowadays, when people can glance at their smartphones and obtain cheap, real-time, information on all restaurants in their surroundings. Finally, a large number of websites and, in particular, almost all social media and commercial websites, provide direct personalized recommendations, e.g., "inspired by your browser history" in Amazon, "who to follow" in Twitter, etc.

Consumer reviews and recommendation systems have complex effects on users' preferences (Duan et al., 2008; Fleder and Hosanagar, 2009) that is not possible to explore in this article. Moreover, the contemporary trend might even be to replace these explicit systems with more subtle presentation cues, embedded in the layout of the user interface, or simply deciding the informations that are presented and the informations that are not, as in Facebook News Feed (Vanderbilt, 2016). These recent and less recent (such as top-lists diffusion) developments are stimulating material for future cultural evolutionary studies, and looking at them through the perspective of cultural transmission biases seems a promising direction.

In conclusion, the details of how popularity influences the spreading of cultural traits need further investigation. The quantitative data resulting from digital media usage may be of great significance for this endeavor. At the same time, new ways to signal and perceive popularity in the digital

environment represent an important new area of research for cultural evolutionary studies.

# 4. PRESERVATIVE AND RECONSTRUCTIVE CULTURAL TRANSMISSION

How faithful is cultural transmission? While, in the popular image, cultural "evolution" implies that ideas and behaviors spread by replicating gene-like from individual to individual, practitioners tend to be more cautious about the analogy genes-cultural traits, in particular regarding fidelity of transmission. The term "meme," invented by Richard Dawkins, is dismissed by the majority of cultural evolutionists, even though sometimes used in social-media literature (e.g., Weng et al., 2012; Adamic et al., 2014).

The oral transmission of stories provides a case in point. Transmission chain experiments, where individuals are asked to iteratively listen to and repeat short narratives (starting from Bartlett, 1932), have shown that, because of memory and attention limits, or biases from previous knowledge, the original material is quickly disrupted (more on transmission chain experiments below). In fact, what is surprising is on the contrary how some orally transmitted folktales have remained relatively stable through centuries or even millennia (Graça da Silva and Tehrani, 2016).

There are various options to explain cultural macro-stability. Some (see e.g., Sperber, 1996; Sperber and Hirschfeld, 2004; Morin, 2015) prefer to concentrate on universal, or slow-changing, factors of attraction that make some cultural traits, or some features of them, particularly memorable, or more likely to be reproduced individually. The stability of a long, oral, transmission chain of a story—say Cinderella—does not depend on a series of faithful acts of copying, but on the fact that some features of the story are particularly likely to be remembered and reconstructed in successive retellings (the example of Cinderella is used in Acerbi and Mesoudi, 2015). The Pumpkin Coach might be one cultural attractor, as an example of a minimally counterintuitive concept (a concept that mainly fits our intuitive cognitive expectations but with few exceptions; for an analysis of the success of folktales due to the presence of minimally counterintuitive concepts see Norenzayan et al., 2006); another might be the relationship between Cinderella and the wicked stepmother (stepparents are considered a serious threat for stepchildren from the point of view of kin selection theory, see Daly and Wilson, 1999).

Others links instead macro-stability to precision of transmission at individual level (micro-stability). Some focus on the fact that, compared to other species that make nevertheless use of social learning, such as other great apes, humans are faithful copiers (Tennie et al., 2009; Dean et al., 2012). Another possibility is that the above mentioned transmission biases provide a way to repeatedly encounter the same behavior, supplying redundancy to the process of cultural transmission (Boyd and Richerson, 1985). Finally, another option yet is provided by epistemic technologies (Sterelny, 2006), i.e., modifications of the external environment that improve

---

[3]https://en.wikipedia.org/wiki/Pick_of_the_Pops

individuals' cognitive abilities, in this case specifically related to facilitate transmission, including extensive apprenticeship or practice.

Acerbi and Mesoudi (2015) argued that these explanations are not mutually exclusive, and that their importance varies depending, among other things, on the domain being studied. Some cultural domains, such as orally transmitted stories, can be considered mainly based on reconstructive cultural transmission, i.e., they derive their stability from the presence of features that are likely to be reconstructed each time by individuals, no matter how faithful is the process of transmission itself. Other domains, for example complex technologies, are characterized by preservative cultural transmission, implemented through faithful copying and external epistemic technologies. As might be expected, reconstruction and preservation, or attraction and faithful copying, are important, in various degrees, in all cultural domains. Rhymes are epistemic tools that make attractive stories even more transmissible (Rubin, 1995); recipes books contain scripts that make universally palatable dishes easier to prepare (Acerbi and Mesoudi, 2015).

Digital media can therefore be considered as a technology that makes cultural transmission more preservative. Cinderella does not need to be listened to, remembered, and retold, but can be "shared" in social media, and practically replicated with extremely low mutation rate. In this sense, the usage of the term "meme" for content that spreads in digital media could be possibly reconciled with its meaning in cultural evolution. An interesting question, from a cultural evolution perspective, is whether the degree of fidelity of transmission influences the kind of content that is more likely to spread.

Cultural evolutionists have investigated content effects experimentally mainly using the above mentioned transmission chain methodology. Transmission chain experiments show that the distortion of the content are consistent, that is, some kinds of content tend to survive along the chains, and others do not. A growing, if somehow unsystematic, catalog of so-called content biases is being built, including among others: a bias for social information (or gossip), involving peoples' relationships and interactions (e.g., Mesoudi et al., 2006); a bias for survival-relevant information, such as location of resources or predators (e.g., Stubbersfield et al., 2015); a bias for content that elicits emotional reactions, especially related to disgust (e.g., Eriksson and Coultas, 2014); a bias for the above mentioned minimally counterintuitive concepts (e.g., Barrett and Nyhof, 2001); a negativity bias, where negatively valenced information is preferred to positively valenced one (Bebbington et al., 2017); a bias for simplicity in linguistic structure (balanced by informativeness, e.g., Kirby et al., 2015), and so on.

However, what if information can be easily reproduced with high-fidelity, as it happens in preservative digital transmission? Promising steps in this direction have recently been made, for example, by experiments from Eriksson and Coultas (2014) and Stubbersfield et al. (2015), which considered each passage in the transmission chain as composed by three distinct phases: choose-to-receive, encode-and-retrieve, and choose-to-transmit. The choose-to-receive and the choose-to-transmit phases indicate respectively the willingness to receive and to circulate cultural information. They are comparable to social media "share," as they do not require the memorization and the repetition of the material, which are required only in the encode-and-retrieve phase. Eriksson and Coultas (2014) found that the bias favoring disgust-related information was operating in the same way in all phases of the transmission. Stubbersfield et al. (2015) compared social and survival information biases, and they found that social information bias had an advantage on survival information bias only in the encode-and-retrieve phase (i.e., the "standard" transmission chain methodology), but not in the choose-to-receive and choose-to-transmit.

Berger and Milkman (2012), with a different approach, examined directly what people share in a 3-month "field study" conducted on New York Times articles. Among other findings, they report that the most shared articles were characterized by a preponderance of positive emotion-valenced terms with respect to negative emotion-valenced ones. This might appear surprising when compared with transmission chain studies that found, on the contrary, that a story with negative content had an advantage in terms of probability to spread and to not be distorted (Bebbington et al., 2017). This negative bias, in terms of favoring attention and memorization, has been confirmed in several experiments, and there are evolutionary reasons to think that negative information should be more salient than positive one (Fessler et al., 2014). One way to reconcile these findings with the results of Berger and Milkman (2012) might be indeed to consider that they studied a paradigmatic case of digitally-mediated preservative transmission, whereas the findings supporting the importance of a negative bias come from cases of reconstructive transmission, or simply related to recall. In this particular case, digital media would favor—because memory and reconstruction are less important than, perhaps, self-presentation motifs, and desire to share positive content with familiars and friends— different content with respect to traditional oral transmission. Other features, for example simplicity and repetitiveness, which have been shown important for the maintenance of oral traditions (Rubin, 1995), seem to contribute in the same way to the success of digital content (Shifman, 2012).

Interestingly, some social media texts, in particular Facebook updates, come with the explicit instruction to "copy-and-paste"— as opposed to share—them. It is not entirely clear why this is the case[4], but, from the point of view we are discussing, copy-and-paste reintroduce variation in highly preservative digital transmission, allowing for modifications that could make the messages more successful (Acerbi and Mesoudi, 2015). Adamic et al. (2014) estimated a "mutation rate" of $\mu = 0.11$ for Facebook

---

[4]One reason might be that shared malicious messages or hoaxes, if reported as such by users, can be easily traced back to the original, and in case all the thread can be deleted by administrators of the social media. Each copy-and-pasted status, by contrast, is an independent piece of content, and can not be immediately linked to the others.

status updates asked to be copy-and-pasted, i.e., 11, every 100 copies, were different from the original, which is extremely high considering the fidelity provided by the digital support.

In fact, some researchers (for example Shifman, 2013) have proposed that one of the main features of internet "memes" is to provide templates that individuals use to introduce personal innovations. Whereas, in oral transmission reconstruction is practically unavoidable, in digitally-supported transmission the content is actively modified by individuals. Shifman (2013) distinguishes two major ways individuals use to modify content: "remix," involving the digital editing of pre-existent material, and "mimicry," involving the actual creation of a new content, inspired by the source. A well-known example of remix is the "Hitler Reacts" meme[5], where fake subtitles, often related to contemporary popular culture topics, are added to a scene of the 2004 movie "Downfall," where an angry Hitler addresses his strict collaborators in his bunker few days before committing suicide. An example of mimicry is instead "Harlem Shake." In the first 2 weeks of February 2013 around 40,000 videos, in which groups of people dance on the music of the song "Harlem Shake," were uploaded to YouTube[6]. The videos are all based on the same concept: they usually start with a single person dancing, surrounded by other people apparently indifferent to the event. Suddenly, the entire group starts to dance, generally with exaggerated and spasmodic-looking movements, often using props and costumes.

More studies are needed to clarify whether there is a specific effect of digital media on the content that is transmitted, but, again, cultural evolution may provide a favorable perspective to investigate this problem. In addition, the distinction preservative/reconstructive is only one of the possible ways to look at the effects of supporting cultural transmission digitally. It has been argued, for example, that universal factors of attraction, or stable content-biases, are especially important with respect to context-based transmission biases (such as popularity and prestige, examined in the previous sections) when cultural transmission chains have two properties. First, they extend through long time-scales, and, second, they are "narrow," that is, the connections between individuals are sparse (Morin, 2015). Digital media seem exactly to be the opposite case, providing fast spreading and high connectivity between individuals (Doer et al., 2012). On the other side, successful cultural traits that spread through digital media can reach enormous diffusion—the well known Gangnam style music video has, as of September 2016, more than two and half billions views on YouTube[7]—which may imply they can reach a very diverse audience, possibly by tapping common psychological preferences.

As above, this review of the cultural evolution literature suggests a way to frame possible questions, more than providing answers. Does the fact that digital media support cheap and high fidelity transmission have an influence on the kind of content that is more likely to spread? What is the role of

mechanisms that introduce variation in digital transmission? Are universal cognitive biases more, less, or equally important in the digital age?

# 5. TAKING THE LONG VIEW

Overall, very few studies in cultural evolution have dealt with these subjects. As a consequence, this review is only proposing some possible directions, and, mainly, suggesting that cultural evolution can provide a "long view" to the contemporary digital environment. When put into perspective, the new phenomena that characterize our digital age appear to have their roots in deeper psychological and historical dynamics, and, to understand what is genuinely new and what is not, we may need to take seriously these dynamics.

The spread of *massive digital misinformation*, for example, is considered one of the most worrying contemporary global risks by the World Economic Forum[8]. Models that explicitly address the spread of misinformation in social networks (Acemoglu et al., 2009; Del Vicario et al., 2016) could greatly benefit of the inclusion of the knowledge developed in cultural evolution. The transmission chain experiments mentioned in the previous section show that certain kinds of information, related for example to gossip or disgust, are more likely to spread than others. How these, and others, predispositions to be influenced in cultural transmission interact with the novel characteristics of digital media (such as high fidelity of transmission, speed, etc.) is material for future studies.

A similar reasoning can be applied to another allegedly worrying phenomenon associated to digital, in particular social, media, that is, the formation of echo chambers. The term "echo chambers" describes the fact that individuals tend, in social media, to associate in communities of like-minded people, and they are thus repeatedly exposed to the same kind of information (e.g., a political ideology) and, especially, they are not exposed to information that could counterbalance it. More concerning, it has been suggested that groups of like-minded people tend to produce opinions that are not an "average" of the opinions of the members of the groups, but their radical version, according to a phenomenon called "group polarization" (Sunstein, 2002).

The empirical evidence for the existence of echo chambers in social media is, however, mixed. Studies showing their existence considered explicitly separated communities of individuals (e.g., Facebook users associated to groups coded as "science news" and "conspiracy theories" in Del Vicario et al., 2016), whereas other researches gave a more nuanced image. Barberá (2014), in a study of Twitter accounts from Germany, Spain, and the United States, found that the usage of social media decreases political polarization, arguing that social media contains more weak ties (i.e., acquaintances or occasional contacts as opposed to close friends or family) with respect to offline networks. In another example, Shore et al. (2016) found that Twitter users post links that are, on average, more moderate than the links they receive in

---

[5]http://knowyourmeme.com/memes/downfall-hitler-reacts
[6]http://knowyourmeme.com/memes/harlem-shake
[7]https://www.youtube.com/watch?v=9bZkp7q19f0

[8]http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/

their feed, and that the perception of polarization at global level is due to the activity of a core of few, but more active, extremist users.

As above, a cultural evolution approach suggests to look at polarization, and echo chambers formation, from a broader perspective. Cultural evolutionists have identified, among the cultural transmission biases discussed in the previous sections, one that refers to "self-similarity," i.e., to the fact that individuals preferentially copy from others similar to them. This has been particularly studied for the arbitrary signals that mark ethnic groups membership. As in the case of prestige bias, or popularity bias, there are reasons to think that a self-similarity bias is an adaptive strategy. The logic is that people of the same group are more likely to live in similar situations, and thus to share the same challenges (Henrich, 2016). One may thus wonder whether or not social media are amplifying the effects of the similarity bias with respect to offline interactions. How polarized are groups of offline friends or coworkers? And what about traditional, broadcast, media?

The broad perspective suggested by cultural evolution does not imply, of course, that the recent modifications produced by digital media are not important, or that media are neutral, and they do not influence what is transmitted. On the contrary, the long view proposed here might be necessary to bring out clearly the novelties. An example toward this direction concerns the incredible amount of user-generated content that has been developed and published with the advent of the so-called Web 2.0, such as blogs, videos, or wiki platforms (van Dijck, 2009). If the motivations of producing some of this content, for example in the case of blogs or video sharing, are likely to be self-promotional, other collaborative enterprises (e.g., Wikipedia, or the WikiHow platform) are more puzzling from a cultural evolutionary point of view. It is common, in cultural evolution (starting from Rogers, 1988), to consider social learners as "information scroungers," that do not pay the cost—and avoid the risk—of individual trial-and-error, relying on the effort of individual learners (Rogers' model shows that populations composed by only, or a great majority of, social learners can not track environmental variation). However, digital media made obvious that, if they have the possibility, individuals seem to be happy to provide, for free, information to unknown "scroungers." How, and to what degree, this may provide a return in terms of reputation or within-group advantage is an interesting question for cultural evolutionary studies of digital media.

Finally, digital media interactions involve substantial changes in the form in which information is transmitted. On one side, digital media favored a surge of text-based, as opposed to oral, communication. For example, the majority of day-to-day conversations between US teenagers happen through text messaging. Non-digital, in person, contacts are in fourth position, preceded by instant messaging and interactions through social media websites[9]. Arguably, previous works on the differences between oral cultures and cultures where writing is widely diffused (e.g., Ong, 1982) are an intriguing starting point to shed light on this phenomenon. Ong (1982), for instance, classified (his) contemporary culture as characterized by a "secondary" orality, i.e., the orality promoted by traditional-broadcast media, profoundly influenced by writing and thus different from the primary orality. One could use the term "secondary literacy" to describe the current situation. Secondary literacy provide, as primary literacy, a way to improve micro-stability of transmission, making it highly preservative, as discussed at length in the previous section. However, it also differs from primary literacy in several respects, including, among others, a more widespread utilization, informal tone, and instantaneity of transmission. In parallel, transmission based on digital media is characterized by the facility of including non-written content, such as images and videos. A significant proportion of the content successfully spreading in the digital environment is in fact characterized by a combination of visual and textual features (think, for example, to image-macro "memes" such as *LOLcats*, or "demotivational" posters[10]).

## 6. CONCLUSION

In the previous sections I highlighted few of the possible investigations that a cultural evolution approach to digital media suggests. One is to look to how traits spread in digital media through the lens of cultural transmission biases. Transmission biases, such as preferentially paying attention to prestigious individuals, or to items that are already popular, are considered adaptations. As such, they are tuned to the conditions of small-scale, slow-changing, and orally-based, societies. How these transmission biases operate in contemporary culture, in which cultural transmission heavily relies on the support of digital media, is an important, and so far unanswered, question. In the same time, I endorsed an elastic view of these biases. Popularity and prestige are not—or, at least, not always—blind forces that push people to copy compulsively. Fears of internet epidemics of extremism, harassment, or similar, driven by influentials or informational cascades, should be considered in a broader context. The quantitative data produced by digital media, together with dedicated experiments, may help us to understand when and how social cues, such as prestige and popularity, interact with the individual evaluation of the content of cultural traits and with other tendencies.

Next, I examined how digital media can be seen as a technology that makes cultural transmission preservative, by providing, practically for free, high fidelity of transmission. This is quite a departure from the conditions usually examined in cultural evolutionary experiments, where items are generally transformed when passing from an individual to another. In addition, digitally-mediated cultural transmission is characterized by other features such as speed, dense connections among individuals, heavy utilization of writing and, in the same time, facility of combining written and audio-visual content. How the interactions of these features influence what

---

[9]http://www.pewinternet.org/2015/08/06/teens-technology-and-friendships/

[10]http://knowyourmeme.com/memes/image-macros

kind of content is more likely to spread is another important investigation.

Cultural evolution is a mature field that could give its contribution to the exam of contemporary cultural phenomena. The digitalization of many instances of cultural transmission seems both relevant for our society and suitable for the theoretical and methodological tools that cultural evolutionists have developed. More empirical and modeling works are needed for this task, and possibly the suggestions sketched here may provide some guidance.

## AUTHOR CONTRIBUTIONS

AA wrote the article, conceived the work, searched and studied the literature, and elaborated the viewpoint that the article expresses.

## REFERENCES

Acemoglu, D., Ozdaglar, A., and ParandehGheibi, A. (2009). Spread of misinformation in social networks. *arXiv*:0906.5007.

Acerbi, A., and Bentley, R. A. (2014). Biases in cultural transmission shape the turnover of popular traits. *Evol. Hum. Behav.* 35, 228–236. doi: 10.1016/j.evolhumbehav.2014.02.003

Acerbi, A., Enquist, M., and Ghirlanda, S. (2009). Cultural evolution and individual development of openness and conservatism. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18931–18935. doi: 10.1073/pnas.0908889106

Acerbi, A., and Mesoudi, A. (2015). If we are all cultural Darwinians what's the fuss about? Clarifying recent disagreements in the field of cultural evolution. *Biol. Philos.* 30, 481–503. doi: 10.1007/s10539-015-9490-2

Adamic, L. A., and Huberman, B. A. (2000). The nature of markets in the World Wide Web. *Q. J. Electron. Comm.* 1, 5–12. doi: 10.2139/ssrn.166108

Adamic, L. A., Lento, T. M., Adar, E., and Ng, P. C. (2014). Information evolution in social networks. *arXiv*:1402.6792.

Aral, S., and Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science* 337, 337–341. doi: 10.1126/science.1215842

Atkisson, C., O'Brien, M. J., and Mesoudi, A. (2012). Adult learners in a novel environment use prestige-biased social learning. *Evol. Psychol.* 10, 519–537. doi: 10.1177/147470491201000309

Attrill, A. (2015). *Cyberpsychology*. Oxford: Oxford University Press.

Bakshy, E., Hofman, J. M., Mason, W. A., and Watts, D. J. (2011). "Everyone's an influencer," in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining - WSDM '11*, ACM Press, New York, NY.

Barberá, P. (2014). *How Social Media Reduces Mass Political Polarization*. Evidence from Germany, Spain, and the US. Job Market Paper, New York University.

Barkow, J. H., O'Gorman, R., and Rendell, L. (2012). Are the new mass media subverting cultural transmission? *Rev. Gen. Psychol.* 16, 121–133. doi: 10.1037/a0027907

Barrett, J., and Nyhof, M. (2001). Spreading non-natural concepts: the role of intuitive conceptual structures in memory and transmission of cultural materials. *J. Cogn. Cult.* 1, 69–100. doi: 10.1163/156853701300063589

Bartlett, F. C. (1932). *Remembering : A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.

Bebbington, K., MacLeod, C., Ellison, T. M., and Fay, N. (2017). The sky is falling: evidence of a negativity bias in the social transmission of information. *Evol. Hum. Behav.* 38, 92–101. doi: 10.1016/j.evolhumbehav.2016.07.004

Bentley, R. A., Hahn, M. W., and Shennan, S. J. (2004). Random drift and culture change. *Proc. Biol. Sci.* 271, 1443–1450. doi: 10.1098/rspb.2004.2746

Berger, J., and Milkman, K. L. (2012). What makes online content viral? *J. Market. Res.* 49, 192–205. doi: 10.1509/jmr.10.0353

Boyd, D. (2014). *It's Complicated: The Social Lives of Networked Teens*. New Haven, CT; London: Yale University Press.

Boyd, D., and Ellison, N. B. (2007). Social network sites: definition, history, and scholarship. *J. Comp. Med. Commun.* 13, 210–230. doi: 10.1111/j.1083-6101.2007.00393.x

Boyd, R., and Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago, IL: University of Chicago Press.

Boyd, R., and Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philos. Trans. R. Soc. B* 364, 3281–3288. doi: 10.1098/rstb.2009.0134

Briggs, A., and Burke, P. (2009). *A Social History of the Media : From Gutenberg to the Internet*. Cambridge: Polity.

Caldwell, C. A., and Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PLoS ONE* 7:e43807. doi: 10.1371/journal.pone.0043807

Cavalli-Sforza, L. L., and Feldman, M. W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, NJ: Princeton University Press.

Chevalier, J. A., and Mayzlin, D. (2006). The effect of word of mouth on sales: online book reviews. *J. Market. Res.* 43, 345–354. doi: 10.1509/jmkr.43.3.345

Chudek, M., Heller, S., Birch, S., and Henrich, J. (2012). Prestige-biased cultural learning: bystander's differential attention to potential models influences children's learning. *Evol. Hum. Behav.* 33, 46–56. doi: 10.1016/j.evolhumbehav.2011.05.005

Claidière, N., Scott-Phillips, T. C., and Sperber, D. (2014). How Darwinian is cultural evolution? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130368. doi: 10.1098/rstb.2013.0368

Colleran, H., and Mace, R. (2015). Social network- and community-level influences on contraceptive use: evidence from rural Poland. *Proc. R. Soc. B* 282:20150398. doi: 10.1098/rspb.2015.0398

Cooney, D., Prieto-Castrillo, F., and Bar-Yam, Y. (2016). Analysis of infectious-recovery epidemic models for membership dynamics of online social Networks. *arXiv*:1608.07870.

Daly, M., and Wilson, M. (1999). *The Truth about Cinderella : A Darwinian View of Parental Love*. New Haven, CT; London: Yale University Press.

de Barra, M., Eriksson, K., and Strimling, P. (2014). How feedback biases give ineffective medical treatments a good reputation. *J. Med. Internet Res.* 16:e193. doi: 10.2196/jmir.3214

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., and Laland, K. N. (2012). Identification of the social and cognitive processes underlying human cumulative culture. *Science* 335, 1114–1118. doi: 10.1126/science.1213969

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., et al. (2016). The spreading of misinformation online. *Proc. Natl. Acad. Sci. U.S.A.* 113, 554–559. doi: 10.1073/pnas.1517441113

Derex, M., and Boyd, R. (2015). The foundations of the human cultural niche. *Nat. Commun.* 6:8398. doi: 10.1038/ncomms9398

Doer, B., Fouz, M., and Friedrich, T. (2012). Why rumors spread so quickly in social networks. *Commun. ACM* 55, 70–75. doi: 10.1145/2184319.2184338

Duan, W., Gu, B., and Whinston, A. B. (2008). Do online reviews matter? An empirical investigation of panel data. *Decis. Support Syst.* 45, 1007–1016.

Efferson, C., Lalive, R., Richerson, P., McElreath, R., and Label, M. (2008). Conformists and mavericks: the empirics of frequency-dependent cultural transmission. *Evol. Hum. Behav.* 29, 56–64. doi: 10.1016/j.evolhumbehav.2007.08.003

Eriksson, K., and Coultas, J. C. (2014). Corpses, maggots, poodles and rats: emotional selection operating in three phases of cultural transmission of urban Legends. *J. Cogn. Cult.* 14, 1–26. doi: 10.1163/15685373-12342107

Fessler, D. M. T., Pisor, A. C., and Navarrete, C. D. (2014). Negatively-biased credulity and the cultural evolution of beliefs. *PLoS ONE* 9:e95167. doi: 10.1371/journal.pone.0095167

Fleder, D., and Hosanagar, K. (2009). Blockbuster culture's next rise or fall: the impact of recommender systems on sales diversity. *Manage. Sci.* 55, 697–712. doi: 10.1287/mnsc.1080.0974

Fortunato, L., and Jordan, F. (2010). Your place or mine? A phylogenetic comparative analysis of marital residence in Indo-European and Austronesian societies. *Philos. Trans. R. Soc. B* 365, 3913–3922. doi: 10.1098/rstb.2010.0017

Fuchs, C. C. (2014). *Social Media : A Critical Introduction*. London: SAGE.

Ghirlanda, S., Acerbi, A., and Herzog, H. (2014). Dog movie stars and dog breed popularity: a case study in media influence on choice. *PLoS ONE* 9:e106565. doi: 10.1371/journal.pone.0106565

Ghirlanda, S., Acerbi, A., Herzog, H., and Serpell, J. A. (2013). Fashion vs. function in cultural evolution: the case of dog breed popularity. *PLoS ONE* 8:e74770. doi: 10.1371/journal.pone.0074770

Graça da Silva, S., and Tehrani, J. J. (2016). Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *R. Soc. Open Sci.* 3:150645. doi: 10.1098/rsos.150645

Henrich, J. (2016). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton, NJ: Princeton University Press.

Henrich, J., and Broesch, J. (2011). On the nature of cultural transmission networks: evidence from Fijian villages for adaptive learning biases. *Philos. Trans. R. Soc. B* 366, 1139–1148. doi: 10.1098/rstb.2010.0323

Henrich, J., and McElreath, R. (2003). The evolution of cultural evolution. *Evol. Anthropol.* 12, 123–135. doi: 10.1002/evan.10110

Heyes, C. (2016a). Blackboxing: social learning strategies and cultural evolution. *Philos. Trans. R. Soc. B* 371, 70–76. doi: 10.1098/rstb.2015.0369

Heyes, C. (2016b). Who knows? Metacognitive social learning strategies. *Trends Cogn. Sci.* 20, 204–213. doi: 10.1016/j.tics.2015.12.007

Kandler, A., and Shennan, S. (2013). A non-equilibrium neutral model for analysing cultural change. *J. Theor. Biol.* 330, 18–25. doi: 10.1016/j.jtbi.2013.03.006

Kelting, K., and Rice, D. H. (2013). Should we hire David Beckham to endorse our brand? Contextual interference and consumer memory for brands in a celebrity's endorsement portfolio. *Psychol. Market.* 30, 602–613. doi: 10.1002/mar.20631

Kempe, M., Lycett, S. J., and Mesoudi, A. (2014). From cultural traditions to cumulative culture: parameterizing the differences between human and nonhuman culture. *J. Theor. Biol.* 359, 29–36. doi: 10.1016/j.jtbi.2014.05.046

Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141, 87–102. doi: 10.1016/j.cognition.2015.03.016

Laland, K. N. (2004). Social learning strategies. *Learn. Behav.* 32, 4–14. doi: 10.3758/BF03196002

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., et al. (2009). Computational social science. *Science* 323, 721–723. doi: 10.1126/science.1167742

Lewens, T. (2015). *Cultural Evolution: Conceptual Challenges*. Oxford: Oxford University Press.

Luca, M. (2011). Reviews, reputation, and revenue: the case of Yelp.Com. *SSRN Electron. J.* doi: 10.2139/ssrn.1928601

Mathew, S., and Boyd, R. (2014). The cost of cowardice: punitive sentiments towards free riders in Turkana raids. *Evol. Hum. Behav.* 35, 58–64. doi: 10.1016/j.evolhumbehav.2013.10.001

Mesoudi, A. (2011). An experimental comparison of human social learning strategies: payoff-biased social learning is adaptive but underused. *Evol. Hum. Behav.* 32, 334–342. doi: 10.1016/j.evolhumbehav.2010.12.001

Mesoudi, A. (2015). Cultural evolution: a review of theory, findings and controversies. *Evol. Biol.* 43, 481–497. doi: 10.1007/s11692-015-9320-0

Mesoudi, A., and Lycett, S. J. (2009). Random copying, frequency-dependent copying and culture change. *Evol. Hum. Behav.* 30, 41–48. doi: 10.1016/j.evolhumbehav.2008.07.005

Mesoudi, A., Whiten, A., and Dunbar, R. (2006). A bias for social information in human cultural transmission. *Br. J. Psychol.* 97, 405–423. doi: 10.1348/000712605X85871

Miton, H., Claidière, N., and Mercier, H. (2015). Universal cognitive mechanisms explain the cultural success of bloodletting. *Evol. Hum. Behav.* 36, 303–312. doi: 10.1016/j.evolhumbehav.2015.01.003

Morin, O. (2015). *How Traditions Live and Die*. Oxford: Oxford University Press.

Muchnik, L., Aral, S., and Taylor, S. J. (2013). Social influence bias: a randomized experiment. *Science*, 341, 647–651. doi: 10.1126/science.1240466

Muthukrishna, M., Morgan, T., and Henrich, J. (2015). The when and who of social learning and conformist transmission. *Evol. Hum. Behav.* 37, 10–20. doi: 10.1016/j.evolhumbehav.2015.05.004

Neimann, F. D. (1995). Stylistic variation in evolutionary perspective: inferences from decorative diversity and interassemblage distance in Illinois Woodland ceramic assemblages. *Am. Antiq.* 60, 7–36. doi: 10.2307/282074

Norenzayan, A., Atran, S., Faulkner, J., and Schaller, M. (2006). Memory and mystery: the cultural selection of minimally counterintuitive narratives. *Cogn. Sci.* 30, 531–553. doi: 10.1207/s15516709cog0000_68

Ong, W. (1982). *Orality and Literacy. The Technologizing of the World*. London: Methuen & Co.

Reindl, E., Beck, S. R., Apperly, I. A., and Tennie, C. (2016). Young children spontaneously invent wild great apes' tool-use behaviours. *Proc. R. Soc. B* 283:20152402. doi: 10.1098/rspb.2015.2402

Rendell, L., Fogarty, L., Hoppitt, W. J., Morgan, T. J., Webster, M. M., and Laland, K. N. (2011). Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends Cogn. Sci.* 15, 68–76. doi: 10.1016/j.tics.2010.12.002

Reyes-Garcia, V., Molina, J. L., Broesch, J., Calvet, L., Huanca, T., Saus, J., et al. (2008). Do the aged and knowledgeable men enjoy more prestige? A test of predictions from the prestige-bias model of cultural transmission. *Evol. Hum. Behav.* 29, 275–281. doi: 10.1016/j.evolhumbehav.2008.02.002

Rogers, A. R. (1988). Does biology constrain culture. *Am. Anthropol.* 90, 819–831. doi: 10.1525/aa.1988.90.4.02a00030

Rubin, D. C. (1995). *Memory in Oral Traditions : The Cognitive Psychology of Epic, Ballads, and Counting-Out Rhymes*. Oxford: Oxford University Press.

Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311, 854–856. doi: 10.1126/science.1121066

Salganik, M. J., and Watts, D. J. (2008). Leading the herd astray: an experimental study of self-fulfilling prophecies in an artificial cultural market. *Soc. Psychol. Q.* 71, 338–355. doi: 10.1177/019027250807100404

Schillinger, K., Mesoudi, A., and Lycett, S. J. (2016). Copying error, evolution, and phylogenetic signal in artifactual traditions: an experimental approach using model artifacts. *J. Archaeol. Sci.* 70, 23–34. doi: 10.1016/j.jas.2016.04.013

Shifman, L. (2012). An anatomy of a YouTube meme. *New Media Soc.* 14, 187–203. doi: 10.1177/1461444811412160

Shifman, L. (2013). *Memes in Digital Culture*. Cambridge: MIT Press.

Shore, J., Baek, J., and Dellarocas, C. (2016). Network structure and patterns of information diversity on Twitter. *arXiv*:1607.06795.

Smaldino, P. E., and McElreath, R. (2016). The natural selection of bad science. *R. Soc. Open Sci.* 3:160384. doi: 10.1098/rsos.160384

Sperber, D. (1985). *On Anthropological Knowledge: Three Essays*. Cambridge: Cambridge University Press.

Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach*. Oxford: Wiley.

Sperber, D., and Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends Cogn. Sci.* 8, 40–46. doi: 10.1016/j.tics.2003.11.002

Sterelny, K. (2006). Memes revisited. *Br. J. Philos. Sci.* 57, 145–165. doi: 10.1093/bjps/axi157

Stubbersfield, J. M., Tehrani, J. J., and Flynn, E. G. (2015). Serial killers, spiders and cybersex: social and survival information bias in the transmission of urban legends. *Br. J. Psychol.* 106, 288–307. doi: 10.1111/bjop.12073

Suler, J. R. (2015). *Psychology of the Digital Age*. Cambridge: Cambridge University Press.

Sunstein, C. R. (2002). The law of group polarization. *J. Polit. Philos.* 10, 175–195. doi: 10.1111/1467-9760.00148

Tamariz, M., Ellison, T. M., Barr, D. J., and Fay, N. (2014). Cultural selection drives the evolution of human communication systems. *Proc. R. Soc. Lond. B Biol. Sci.* 281:20140488. doi: 10.1098/rspb.2014.0488

Tanaka, M. M., Kendal, J. R., and Laland, K. N. (2009). From traditional medicine to witchcraft: why medical treatments are not always efficacious. *PLoS ONE* 4:e5192. doi: 10.1371/journal.pone.0005192

Tehrani, J. J. (2013). The phylogeny of little red riding hood. *PLoS ONE* 8:e78871. doi: 10.1371/journal.pone.0078871

Tennie, C., Call, J., and Tomasello, M. (2009). Ratcheting up the ratchet: on the evolution of cumulative culture. *Philos. Trans. R. Soc. B* 364, 2405–2415. doi: 10.1098/rstb.2009.0052

Turchin, P., Currie, T. E., Turner, E. A. L., and Gavrilets, S. (2013). War, space, and the evolution of Old World complex societies. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16384–16389. doi: 10.1073/pnas.1308825110

van Dijck, J. (2009). Users like you? Theorizing agency in user-generated content. *Media Cult. Soc.* 31, 41–58. doi: 10.1177/0163443708098245

Vanderbilt, T. (2016). *You May Also Like. Taste in An Age of Endless Choice*. London: Simon and Schuster.

Wallace, P. (2001). *The Psychology of the Internet*. Cambridge: Cambridge University Press.

Watts, D. J., and Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *J. Consum. Res.* 34, 441–458. doi: 10.1086/518527

Watts, J., Greenhill, S. J., Atkinson, Q. D., Currie, T. E., Bulbulia, J., and Gray, R. D. (2015). Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proc. R. Soc. B* 282:20142556. doi: 10.1098/rspb.2014.2556

Weng, L., Flammini, A., Vespignani, A., and Menczer, F. (2012). Competition among memes in a world with limited attention. *Sci. Rep.* 2, 37–52. doi: 10.1038/srep00335

West, J., and Mace, M. (2010). Browsing as the killer app: explaining the rapid success of Apple's iPhone. *Telecommun. Policy* 34, 270–286. doi: 10.1016/j.telpol.2009.12.002

Whiten, A., McGuigan, N., Marshall-Pescini, S., and Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philos. Trans. R. Soc. B* 364, 2417–2428. doi: 10.1098/rstb.2009.0069

Ye, Q., Law, R., and Gu, B. (2009). The impact of online user reviews on hotel room sales. *Int. J. Hospital. Manage.* 28, 180–182. doi: 10.1016/j.ijhm.2008.06.011

# Advantages of publishing in Frontiers

**OPEN ACCESS**
Articles are free to read, for greatest visibility

**COLLABORATIVE PEER-REVIEW**
Designed to be rigorous – yet also collaborative, fair and constructive

**FAST PUBLICATION**
Average 85 days from submission to publication (across all journals)

**COPYRIGHT TO AUTHORS**
No limit to article distribution and re-use

**TRANSPARENT**
Editors and reviewers acknowledged by name on published articles

**SUPPORT**
By our Swiss-based editorial team

**IMPACT METRICS**
Advanced metrics track your article's impact

**GLOBAL SPREAD**
5'100'000+ monthly article views and downloads

**LOOP RESEARCH NETWORK**
Our network increases readership for your article

**Find us on**