Title
Learning with Conversational Agents

Author
Xu, Ying

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Learning with Conversational Agents

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Ying Xu

Dissertation Committee:
Professor Mark Warschauer, Chair
Professor Young-Suk Kim
Associate Professor Penelope Collins
Assistant Professor Stacy Branham
Professor Arthur Graesser

2020

TABLE OF CONTENTS

LIST OF TABLES iv
LIST OF FIGURES vi
ACKNOWLEDGEMENTS vii
VITA viii
ABSTRACT xii
CHAPTER 1 INTRODUCTION 1
Dissertation Overview 4
References 7
CHAPTER 2 LEARNING AND ENGAGEMENT WITH CONVERSATIONAL AGENTS 9
Study Abstract 9
Introduction 9
Literature Review 11
Method 17
Results 25
Discussion 31
Conclusion 36
References 37
CHAPTER 3 COMMUNICATION PATTERNS 51
Study Abstract 51
Introduction 51
Literature Review 54
Development of the CA Reading Partner 61
Method 63
Results 68
Robustness Check 74
Discussion 74
Conclusion 80
References 82
CHAPTER 4 VERBAL AND NON-VERBAL INTERACTIONS 106
Study Abstract 106

Introduction ... 106
CA Dialogue Flow ... 107
Method ... 108
Results ... 111
Discussion ... 119
Conclusion ... 125
References ... 126
CHAPTER 5 PERCEPTIONS ... 139
Study Abstract ... 139
Introduction ... 139
Literature Review ... 142
Method ... 146
Results ... 151
Discussion ... 156
Conclusion ... 162
References ... 163
CHAPTER 6 CONCLUSION ... 181
References ... 184
Appendix A ... 186
Appendix B ... 189

LIST OF TABLES

		Page
Table 1.1	Summary of Studies	3
Table 2.1	Background Information by Condition	45
Table 2.2	Descriptive Statistics of Outcome Measures for Full Sample	46
Table 2.3	Outcome Measures by Condition	47
Table 2.4	Regression Analysis of The Condition Effects on Story Comprehension	48
Table 2.5	Effects of Condition on Engagement	49
Table 3.1	Participant Background Information by Experimental Condition	98
Table 3.2	Intercorrelations Among Study Variables	99
Table 3.3	Descriptive Statistics of Story Comprehension and Verbal Engagement Variables by Experimental Condition	100
Table 3.4	Linear Regression Model on Story Comprehension Measures	101
Table 3.5	Multilevel Linear Models on Verbal Engagement Measures	102
Table 4.1	CA Performance in Intent Categorization	130
Table 4.2	Language Production	131
Table 4.3	Topic Relevance in Children's Responses	132
Table 4.4	Timing of Children's Responses	133
Table 4.5	Affect in Children's Responses to the CA	134

Table 4.6	Children's Affective Reactions to Receiving Positive Feedback from the CA	135
Table 4.7	Children's Affective Reactions to Receiving Negative Feedback from the CA	136
Table 4.8	Children's Affective Reactions to Receiving Vague, Neutral Feedback from the CA	137
Table 5.1	Children's Domain Membership Categorization of the CA in Interview and Drawing	172
Table 5.2	Children's Attribution of Cognitive, Psychological, and Behavioral Properties of the CA	173
Table 5.3	Children's Justifications of the CA's Properties. Bolded Numbers Indicate Salient Justification Patterns within Each Property	174

LIST OF FIGURES

		Page
Figure 2.1	SEM Analysis of Reading Condition, Vocalizations, and Story Comprehension	50
Figure 3.1	Dialogue Flow Design of the CA's Guided Conversation Module	103
Figure 3.2	Experiment Session Setup	104
Figure 3.3	Verbal Engagement by the Nature of Language Partner and Questions' Cognitive Demand Levels	105
Figure 4.1	Child-CA Dialogue Flow	138
Figure 5.1	Drawings That Illustrate Google Home Mini as Artifacts	175
Figure 5.2	Drawings That Illustrate Google Home Mini as Living Objects	176
Figure 5.3	Drawings That Illustrate Google Home Mini as a Combination of Artifacts and Living Objects or as Neither Artifacts nor Living Objects	177
Figure 5.4	Drawing Samples That Contain Cognitive Elements	178
Figure 5.5	Drawing Samples That Contain Psychological Elements	179
Figure 5.6	A Drawing Sample That Contains Behavioral Elements	180

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor and committee chair, Professor Mark Warschauer, who have been my biggest supporter in my PhD journey. I am grateful for his mentorship, which has helped me becoming a researcher and will always inspire me to pursue new heights. He trusts me, values my passions, and works side-by-side with me to bring my seemingly wild research ideas into fruition. He is the best advisor I can ever ask for.

I would also like to thank my other committee members, Professor Young-Suk Kim, Professor Penelope Collins, Professor Stacy Branham, and Professor Arthur Graesser. Their invaluable guidance has helped improve this dissertation work and will continue to benefit my future research.

Many thanks go to my friends, colleagues, and collaborators at UC Irvine and beyond. Your support has helped me get to the finish line of this journey and gives me the courage to embark my new adventure in the academic world.

Last but not the least, I want to thank my family who always accompany me and cheer for me.

This work is in part supported by UC Irvine’s Academic Senate Council on Research, Computing and Libraries (CORCL) funds, Public Impact Fellowship, and the Undergraduate Research Opportunity Fellowship.

VITA

Ying Xu

University of California

School of Education

EDUCATION

2020 ***Ph.D.* in Education** (specialization: Language, Literacy, and Technology)
University of California, Irvine
Dissertation: Conversational Agents as Social Learning Partners
Committee: Mark Warschauer (Chair, UCI Education), Stacy Branham (UCI Informatics), Arthur Graesser (U of Memphis), Young-suk Kim (UCI Education), Penelope Collins (UCI Education)

2010 ***B.A.* in Chinese Linguistics and Literature**
Sun Yat-Sen University, China

HONORS/AWARDS

2021 Best Paper Award Honorable Mention at the 2021 American Educational Research Association annual conferences (AERA)
2020 Best Paper Award Honorable Mention at the 2020 ACM Conference on Interaction Design and Children
2020 Research and Design Competition Honorable Mention at the 2020 ACM Conference on Interaction Design and Children
2020 Doctoral Consortium at the 2020 ACM Conference on Interaction Design and Children
2020 Grad Slam Finalist (Scientific Communication Competition) UC Irvine
2019 Public Impact Distinguished Fellow UC Irvine
2019 Best Short Paper Award at the 2019 ACM Conference on Interaction Design and Children
2018 Best Paper Award Honorable Mention at the 2018 American Educational Research Association annual conferences (AERA)

PUBLICATIONS

<u>Conference Proceedings</u>

CP8 **Xu, Y.,** & Warschauer, M. (2020). Using conversational agents to foster young children's science learning from screen media. In *Proceedings of the 19th ACM International Conference on Interaction Design and Children* (IDC '20). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3397617.3398031

CP7 **Xu, Y.,** & Warschauer, M. (2020). Wonder with Elinor: Designing a socially contingent video viewing experience. In *Proceedings of the 19th ACM International Conference on Interaction Design and Children* (IDC '20). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3397617.3398024 **(Research and Design Competition Honorable Mention)**

CP6 **Xu, Y.,** & Warschauer, M. (2020). Exploring young children's engagement in joint reading with a conversational agent. In *Proceedings of the 19th ACM International Conference on Interaction Design and Children* (IDC '20). Association for Computing Machinery, New York, NY, USA. https://doi.org/ 10.1145/3392063.3394417

CP5 **Xu, Y.,** & Warschauer, M. (2020). A content analysis of voice-based apps on the market for early literacy development. In *Proceedings of the 19th ACM International Conference on Interaction Design and Children* (IDC '20). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3392063.3394418 **(Best Paper Award Honorable Mention)**

CP4 **Xu, Y.,** & Warschauer, M. (2020). "Elinor is talking to me on the screen!" Integrating conversational agents into children's television programming. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. April 25-30, 2020, Honolulu, HI. ACM. https://doi.org/10.1145/3334480.3383000

CP3 **Xu, Y.,** & Warschauer, M. (2020). What are you talking to?: Understanding children's perceptions of conversational agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. April 25-30, 2020, Honolulu, HI. ACM. https://doi.org/10.1145/3313831.3376416

CP2 **Xu, Y.,** Yau, J.C., & Reich, S. (2019). The added challenge of digital reading: Exploring young children's page turning behaviors. In *Proceedings of the 2019 Conference on Interaction Design and Children.* June 12-15, 2019, Boise, ID. ACM. https://doi.org/10.1145/3311927.3323121 **(Best Short Paper Award)**

CP1 **Xu, Y.,** & Warschauer, M. (2019). Young children's reading and learning with conversational agents. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. May 4–9, 2019, Glasgow, Scotland, UK. ACM. https://doi.org/10.1145/3290607.3299035

<u>Journal Articles</u>

J8 **Xu, Y.,** Wang, D., Collins, P., Lee, H., & Warschauer, M. (2020). Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. *Computers & Education*. https://doi.org/10.1016/j.compedu.2020.104059

J7 Umarji, O., Day, S., **Xu, Y.**, Zargar, E., Connor, C., & Yu, R. (2020). Opening the black box: User-log analysis of children's e-book reading and association with word knowledge. *Reading and Writing.* https://doi.org/10.1007/s11145-020-10081-x

J6 **Xu, Y**., Yau, J., & Reich, S. (2020). Press, swipe, and read: Can interactive features in e-books facilitate engagement and learning?. *Journal of Computer Assisted Learning.* https://doi.org/10.1111/jcal.12480

J5 Reich, S., Yau, J., **Xu, Y**., Muskat, T., Campbell, J., & Cannata, D. (2019). Digital or print? A comparison of preschooler's comprehension, vocabulary, and engagement from a print book and an eBook. *AERA Open*. 5(3), 1-16. https://doi.org/10.1177/2332858419878389

J4 **Xu, Y.,** Xu. D., Simpkins, S., & Warschauer, M. (2019). Does it matter which parent is absent?: Labor migration, parenting, and adolescent development in China. *Journal of Child and Family Studies*. 28(6), 1635-1649. https://doi.org/10.1007/s10826-019-01382-z

J3 Tate, T., Collins, P., **Xu, Y**., Yau, J., Krishnan, J., Prado, Y., Farkas, G., & Warschauer, M. (2019). Visual-syntactic text format: Improving adolescent literacy. *Scientific Studies of Reading*. 23(4), 287-304. https://doi.org/10.1080/10888438.2018.1561700

J2 Park, Y., **Xu, Y**., Collins, P. Farkas, G., & Warschauer, M. (2018). Scaffolding learning of language structures with visual-syntactic text formatting. *British Journal of Educational Technology*. 50(4), 1896-1912. https://doi.org/10.1111/bjet.12689

J1 **Xu, Y.** Rethinking literacy education in digital era: Digital storybook as a learning tool for young children [J], *China Educational Technology*, 2014, 333(10): 29-35.

Book Chapters

B3 Xu D., & **Xu Y.** (2020) The ambivalence about distance learning in higher education. In: Perna L. (eds). *Higher Education: Handbook of Theory and Research*, vol 35. Springer, Cham

B2 Warschauer, M., & **Xu, Y**. (2018). Technology and Equity in Education. In J. Voogt, G. Knezek. R. Christensen, & K-W. Lai (Eds.) *International Handbook of Information Technology in Primary and Secondary Education* New York: Springer US. https://doi:10.1007/978-3-319-53803-7_76-1

B1 **Xu, Y.** & Warschauer, M. (2018). Language and digital divide. In B. Warf (Ed.) *The Sage Encyclopedia of the Internet*. Thousand Oaks, CA: SAGE.

Research Report

R1 Xu, D., & **Xu, Y.** (March 2019). The promises and limits of online higher education: Understanding how distance education affects access, cost, and quality. American Enterprise Institute. http://www.aei.org/publication/the-promises-and-limits-of-online-higher-education/

SELECTED CONFERENCE PRESENTATIONS

P16 Deng, X, Song, Y, & Xu, Y. (April 2021). Young children's reading with conversational agents: the role of age and language status. Poster to be presented at the Annual Meeting of the American Educational Research Association. **(Nominated for Best Paper Award in the Technology as an Agent of Change in Teaching and Learning <TACTL> SIG)**

P15 Warschauer, M., **& Xu, Y.** (September 2020). Can conversational agents support children's learning?. Conference on Educational Data Science, Stanford, CA (Virtual conference).

P14 **Xu, Y.** (June 2020). Using conversational agents to foster young children's science learning from screen media. Doctoral Consortium proposal to be presented at the at the ACM Conference on Interaction Design and Children, London, UK (Virtual conference).

P13 **Xu, Y.,** Hoang, T., Sun, B, & Warschauer, M. (April 2020). What are you talking to?: Understanding children's perceptions towards conversational agents. Paper accepted at the Annual Meeting of the American Educational Research Association, San Francisco, CA (Conference canceled).

P12 **Xu, Y.,** Lee, H., Bautista S., & Warschauer, M. (April 2020). Examining the effect of a conversational agent as a reading partner. Paper accepted at the Annual Meeting of the American Educational Research Association, San Francisco, CA (Conference canceled).

P11 **Xu, Y.,** & Warschauer, M. (October 2019). Conversational agents as educational video co-viewers for young children. Paper to be presented at the 2019 Connected Learning Summit, Irvine, CA.

P10 **Xu, Y.,** Wang, D.Q., Yau, J., Han, H., & Warschauer, M. (April 2019). Reading in a community: How social affordances of an eBook platform motivate children to read. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada.

P9 **Xu, Y.**, Prado, Y., Collins, P., & Warschauer, M. (April 2019). Effective teaching in the digital classroom: A mixed-method study of middle school English language arts classes. Poster presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada.

P8 **Xu, Y.**, Yau, J., & Reich, S. (April 2019). Press, swipe, and read: Can interactive features in e-books facilitate engagement and learning? Poster presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada.

P7 **Xu, Y.,** Yau, J., & Reich, S. (March 2019). The added challenge of eBook reading: Exploring children's page-turning behaviors. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Baltimore, MD.

P6 **Xu. Y**., Yau, J., & Reich, S.(August 2018). Reading with hotspots: Exploring young children's haptic engagement with touchscreen stories. Poster presented at the American Psychological Association Annual Convention, San Francisco, CA.

P5 **Xu, Y.,** Yau, J., Warschauer, M., & Collins, P. (2018, April). Examining the role of teacher efficacy in the implementation of a large-scale technology-based intervention. Paper to be presented at the Annual Meeting of the American Educational Research Association, New York. **(Nominated for Best Paper Award in the Technology as an Agent of Change in Teaching and Learning <TACTL> SIG)**

P4 **Xu, Y.**, Yau, J., & Reich, S. (2018 April). Understanding preschool children's engagement and learning with interactive ebooks. Paper to be presented at the Annual Meeting of the American Educational Research Association, New York.

P3 **Xu, Y.,** Xu. D., Simpkins, S., & Warschauer, M. (2017 April). Does it matter which parent is absent?: Labor Migration, Parenting, and Adolescent Development in China. Paper to be presented at the Annual Meeting of the American Educational Research Association, San Antonio, TX.

P2 **Xu, Y.,** Wang, D.Q., Yau, J., & Warschauer, M. (2018 May). Reading in a community: how social affordances of an ebook platform motivate children to read. Paper to be presented at the Digital Learning in the Humanities and Beyond Symposium, Irvine, CA.

P1 **Xu. Y**., Yau, J., & Reich, S. (April 2018). Tablet for independent reading?: exploring young children's page navigation and story comprehension. Poster presented at the American Psychological Association: Technology, Mind, and Society, Washington. DC.

ABSTRACT

Learning with Conversational Agents

By

Ying Xu

Doctor of Philosophy in Education

University of California, Irvine, 2020

Professor Mark Warschauer, Chair

This dissertation consists of four studies that reveal how children learn from, respond to, interact with, and perceive conversational agent as their learning companion during shared storybook reading. The first study focuses on ***how children learn from*** the agent. I found that contingent, structured dialogue with the conversational agent led to children's enhanced story comprehension. Such benefit was largely driven by children's heightened level of vocalizations related to the story narratives and reduced off-topic vocalizations. The second study primarily focuses on how ***children respond to*** the conversational agent. I found that conversational agents promoted children's response intelligibility, while adults elicited longer, more lexically diverse, and more relevant responses. The differences in language productivity were amplified among the questions requiring high cognitive demand. The third study focuses on ***how children interact with*** the agent verbally and non-verbally. I found that children generally participated in the conversation with the agent smoothly: they generated on-topic responses and answered within the proper time frame. The result also confirmed the advantage of using a combination of open-ended questions as initial prompts to encourage children's free expression and multiple-choice questions as follow-up prompts to help ease the potential cognitive obstacles. Such scaffolding mechanisms appeared to benefit younger children more so than older ones. The fourth study

focuses ***how children perceive*** the agent. I found that children in general held positive perceptions in terms of conversational agents' cognitive and psychological capabilities. Children's such perceptions establish the feasibility of developing agents to socially engage children in learning activities. Overall, the four studies provide converging evidence on the promise of leveraging AI-powered conversational technologies to support young children's language development. The findings are intended to be generalized to designing socially interactive environments for different learning domains (e.g., science) and learning scenarios (e.g., television watching), with a goal of promoting children's long-term development.

CHAPTER 1 INTRODUCTION

Young children learn best when they socially interact with an engaging and knowledgeable adult. One important aspect of early childhood learning is children's development of language skills (Dickinson & Tabors, 2001), and storybook reading with adults has long been viewed as a prime context for stimulating such development (Bus, 2001). Moreover, the benefits of storybook reading are amplified if adults engage children in conversation that is temporally and topically contingent on children's verbal contributions. While storybook reading is a routinized activity engaged in by families across cultures, the quantity and quality of storybook reading children experience in their everyday lives differ depending on their caregivers' availability, skills, and inclinations. This unequal exposure to high quality shared reading experiences is believed to contribute to the language and literacy divide among children in the U.S (Farver et al., 2013; Phillips & Lonigan, 2009).

Researchers, educators, and policy makers have thus sought out innovative approaches to enriching children's early literacy experiences. The recent development of artificial intelligence (AI) has presented a promising means to this end. Conversational agents powered by AI have the affordances to understand natural speech language input, potentially allowing conversational agents to mimic interpersonal social interaction. Conversational agents, such as Apple Siri, Google Assistant, and Amazon Alexa, are currently prevalent in many homes, and children readily interact with them. The capability and accessibility of these technologies point to the feasibility of conversational agents standing in for a human language partner during children's learning processes (Sengupta & Garg, 2019). However, to my knowledge, research focusing on conversational agents' educational affordances appears to be at the beginning stage, with most research merely exploring how children interact with commercially available agents (e.g., Amazon Alexa) not specifically tailored for children's learning

This dissertation consists of four papers focusing on children's language learning with conversational agents designed to serve as a learning companion during storybook reading. Each of the papers tackles this issue from a different angle, and together they provide a comprehensive investigation of how children learn from, respond to, interact with, and perceive this conversational agent.

The data sources of this dissertation were collected from an experiment involving 117 children aged 3 to 6 years. This experiment utilized a two-by-two study design, with the two factors being whether children read with a conversational agent or an adult and whether children had engaged in "conversation" during the storybook reading activity (i.e., dialogic reading). This study design resulted in four conditions:

- Reading with an agent without conversation,
- Reading with an agent with conversation,
- Reading with an adult without conversation, and
- Reading with an adult with conversation.

Data was collected before, during, and after the storybook reading, and included: 1) a parent survey that contained children's demographic information, 2) an expressive vocabulary test (EWOPVT-4) that captured children's baseline language level, 3) video-taped reading sessions, which we used for detailed coding and analysis of children's engagement, communication, and interactions with the agent, 4) a story comprehension assessment after children's storybook reading as a measure of learning outcome, and 5) a semi-structured interview and a drawing task to probe children's perception of the agent they interacted with.

Table 1.1
Summary of Studies

	Data Sources	Analytic techniques	Sample
Study 1	Story comprehension Engagement coded from the Video-taped sessions - Global rating - Vocalizations - Affective expressions - Visual attention	OLS regression; Structural equation modelling	*n* = 117 Four conditions
Study 2	Story comprehension Verbal responses coded from the Video-taped sessions - Productivity - Lexical diversity - Topic relevance - Accuracy - Intelligibly	OLS regression; Multilevel linear regression	*n* = 90 Three conditions (“Agent without conversation” condition excluded
Study 3	Verbal and non-verbal interactions coded form the video-taped sessions - Productivity - Flow maintenance - Affect	ANOVA; Qualitative analysis	*n* = 33 One condition (Only “Agent with conversation” condition included)
Study 4	Perceptions collected from the semi-structured interview and drawing task - Domain membership - Human-like properties - Justification of property attribution	Qualitative analysis	*n* = 28 One condition (Only “Agent with conversation” condition included)

Each of the articles utilized a different sample of child participants and included one or more of the experimental conditions. Each article also relied on a different subset of the data sources and used a different analytic approach. Table 1 provides an overview of the studies, and information pertaining to each study is described within the corresponding chapter. Given that these articles targeted varying publication venues, different terms were sometimes used to refer to the same object, phenomena, or activity (e.g., “dialogic reading” in Study 1 versus “storybook reading with guided conversation” in Study 2).

Dissertation Overview

This dissertation has six chapters, including an introductory chapter, four chapters corresponding to each study, and a concluding chapter. There is no separate literature review or method chapter, because each study chapter can stand alone and includes its own literature review and method section, which provide specific information for the context of each study.

Study 1 focused on how dialogic reading with a conversational agent influenced children's story comprehension and reading engagement as measured by children's vocalization, affective expression, and visual attention as well as a global rating of their engagement. Specifically, I examined the effects of two factors–whether they read with a human or an agent and whether children had engaged in dialogue– as well as the interaction between these two factors. Additionally, I also explored engagement as a mediating factor between dialogic reading and story comprehension. Three research questions were asked:

1. What is the effect of dialogic reading with a conversational agent on children's story comprehension?
2. What is the effect of dialogic reading with a conversational agent on children's reading engagement?
3. Do changes in engagement serve as a mechanism through which dialogic reading with a conversational agent affects story comprehension?

Results revealed that a properly designed social agent can replicate the benefits of dialogic reading with an adult partner. Furthermore, dialogic reading with an agent promotes story comprehension through enhancing children's narrative-relevant vocalizations and reducing irrelevant vocalizations.

Study 2 compared children's verbal responses to questions asked by either an agent or an adult. In addition to the nature of the language partner, I also looked at whether the questions' cognitive demand levels play a role in how children respond. Two questions were asked in this study:

1. Do children's verbal responses with a human partner resemble or differ from their behaviors with a conversational agent?
2. Do the similarities or differences in verbal responses apply to both low- and high-cognitive-demand questions?

Overall, this study uncovered several differences in children's verbal engagement when interacting with a conversational agent rather than with an adult. Specifically, children who read with the conversational agent responded to questions with better intelligibility, whereas those who read with an adult responded to questions with higher productivity, lexical diversity, and topical relevance. Both groups responded to questions with a similar level of accuracy. In addition, questions requiring high cognitive demand amplified the differences in children's verbal responses to the two reading partners

Study 3 focused on the subsample of children who had conversation with the agent during storybook reading. This study examined children's verbal and non-verbal responses. Additionally, I examined the different patterns in children's responses to the agent's original open-ended questions and its follow-up multiple-choice questions. I also considered whether children's age played a role in their response patterns. Two questions were asked in this study:

1. How do children respond to a conversational agent reading partner, in terms of language production, flow maintenance, and affect?

2. Do younger children (3- to 4-year-olds) respond to the conversational agent reading partner differently than do older children (5- to 6-year-olds)?

Overall, I found that children actively participated in conversation with the CA and frequently generated on-topic responses. Children were generally able to respond to the CA within the proper time frame. Children also showed positive affect while speaking to the CA or listening to the CA's feedback. Moreover, younger children appeared to encounter more obstacles during their interactions with the conversational agents, yet such obstacles did not seem to impact their enthusiasm as these children also showed a higher level of affect towards their conversational agent reading partner than the older children.

Study 4 analyzed children's perceptions of the agent using the framework of animate-inanimate distinction. In particular, I examined children's perceptions of the agent's animate/ inanimate domain membership and properties, as well as their justifications for these perceptions. Three questions were asked in this study:

1. Which domain do children perceive conversational agents as belonging to (e.g., artifact, living object, or something else)?
2. Do children view conversational agents as possessing human-like cognitive, psychological, and behavioral properties?
3. How do children reason about whether conversational agents possess certain properties?

Overall, the findings to these three questions suggested that children sometimes take a more nuanced position and spontaneously attribute both artifact and animate properties to CAs. At least some children appeared unwilling to describe the conversational agents as either a living being or an artifact. Additionally, children appeared to consistently conceive of conversational agents as possessing a unique constellation of animate properties while lacking others. Examination of children's justifications for their perceptions further revealed nuanced reasoning. Taken together, these findings extend current research on children's perceptions of intelligent artifacts by adding conversational agents as a new genre of study and also provide some underlying knowledge that may guide the development of conversational agents to support young children's cognitive and social development.

In the conclusion chapter, I discuss the implications of the findings emerged across the four studies and future directions.

References

Bus, A. G. (2001). Joint caregiver-child storybook reading: A route to literacy development. *Handbook of early literacy research*, *1*, 179-191.

Dickinson, D. K., & Tabors, P. O. (2001). Beginning literacy with language: Young children learning at home and school. Paul H Brookes Publishing.

Farver, J. A. M., Xu, Y., Lonigan, C. J., & Eppe, S. (2013). The home literacy environment and Latino head start children's emergent literacy skills. *Developmental Psychology, 49*(4), 775.

Kats, R. (2018). *The Smart Speaker Series: Kids & Teens*. Retrieved from https://www.emarketer.com/content/the-smart-speaker-series-kids-teens-infographic

Phillips, B. M., & Lonigan, C. J. (2009). Variations in the home literacy environment of preschool children: A cluster analytic approach. *Scientific Studies of Reading*, *13*(2), 146-174.

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype Me! Socially Contingent

Interactions Help Toddlers Learn Language. *Child Development*, *85*(3), 956–970. https://doi.org/10.1111/cdev.12166

Sengupta, S., & Garg, R. (2019). Impact of voice-based interaction on learning practices and behavior of children. *CEUR Workshop Proceedings*, *2327*, 1–3.

CHAPTER 2 LEARNING AND ENGAGEMENT WITH CONVERSATIONAL AGENTS

Study Abstract

Dialogic reading, when children are read a storybook and engaged in relevant conversation, is a powerful strategy for fostering language development. With the development of artificial intelligence, conversational agents can engage children in some elements of dialogic reading. This study examined whether a conversational agent could improve children's story comprehension and engagement, as compared to an adult reading partner. Using a 2 (dialogic reading or non-dialogic reading) by 2 (agent or human) design, a total of 117 3- to 6-year-olds were randomly assigned into one of the four conditions. Results revealed that a conversational agent can replicate the benefits of dialogic reading with a human partner through enhancing children's narrative-relevant vocalizations and reducing irrelevant vocalizations.

Introduction

Preschool years are a critical time for developing language skills that are needed to succeed in school. Storybook reading with adults, typically caregivers or teachers, provides a prime context to bolster children's language development. In line with the Vygotskian principle of scaffolding (Berk & Winsler, 1995), the benefits of storybook reading are amplified by engaging children in contingent, structured interactions that revolve around story narratives and facilitate conversation about content that is just above the child's current level of understanding. This interactive reading style--termed dialogic reading--includes asking open-ended questions to stimulate children's thinking and providing feedback for child participation (Arnold et al., 1994; Whitehurst, 1992). Dialogic reading interventions with caregivers and teachers have confirmed the promise of using dialogue for enhancing children's engagement during reading and supporting children's vocabulary learning, comprehension, and expressive language (for reviews, see Flack et al., 2018; Mol et al., 2008; Noble et al., 2019; Towson et al., 2017). However, the quantity and quality of storybook reading children are exposed to largely depend on the training opportunities, availability, skills, and inclinations of their caregivers or teachers. Unequal access to high-quality reading experiences is believed to contribute to the language and literacy divide among children in the U.S. (Farver et al., 2013; Phillips & Lonigan, 2009).

With the rapid development of artificial intelligence, children are increasingly interacting with non-human intelligent agents through speech, gesture, or writing. Conversational agents that support natural speech interaction may be especially valuable for young children, whose lack of proficiency in literacy or fine motor skills causes them difficulty in navigating many digital contents (Lovato & Piper, 2019). Conversational agents comprehend speech, thus enabling complex dialogue that mimics human-to-human conversation. Several familiar examples of speech-based agents include Apple Siri, Google Assistant, and Amazon Alexa. Due to these products' growing prevalence, the developmental consequences of children interacting with speech-based agents has spurred much research interest (e.g., Garg & Sengupta, 2020; Sciuto et al., 2018; Yuan et al., 2019). Some researchers argue that machine-mediated communication afforded by agents could lead to "social interactions" akin to interpersonal communications, thus assuming the role of a "partner" or "guide" in children's language learning processes (Roseberry et al., 2014). However, there is little evidence as to whether and how interacting with conversational agents supports language development.

This experimental study provides a direct examination of this issue. We focus on the impacts and mechanisms of learning and engagement in storybook reading by young children when interacting with a conversational agent compared to a human partner. Evidence that conversational agents can emulate the benefits of an adult co-reader would offer a promising mechanism for supporting children's language development in daily life (Sengupta & Garg, 2019). In the following section, we discuss the theoretical perspectives underpinning this study and prior work that led to the formation of our research questions and hypotheses.

Literature Review

Sociocultural theory views language development as a socially mediated process in which children acquire their language skills through collaborative dialogue with more knowledgeable members of society in their everyday activities (John-Steiner & Mahn, 1996). Through back-and-forth conversation with more knowledgeable language partners who provide scaffolding and facilitate active participation, children internalize knowledge by focusing attention, expressing thoughts, and critically reflecting on the topic being discussed (Golinkoff et al., 2019). Moreover, sociocultural theory emphasizes that the experienced adult should purposely craft a language environment that is developmentally appropriate to the child (Bodrova & Leong, 2005). In other words, the adult should assume the role as a language guide, and scaffold children's participation in the conversation. A great deal of recent research has adopted this perspective and designed socially interactive learning experiences to support children's language development, either in face-to-face settings (e.g., dialogic reading with an adult) or computer-mediated environments (e.g., conversational agents). This relevant literature is reviewed in the following sections.

Dialogic Questioning during Reading

Whitehurst and colleagues established the interactive "dialogic reading" paradigm that involves adults using elaborative questioning and feedback techniques to encourage children's oral contributions (Whitehurst, 1992; Arnold & Whitehurst, 1994). Specifically, during dialogic reading sessions, the adult uses elaborative "wh-" and open-ended questions, repetition of good responses, and expansion of incomplete responses to model sentence formation. The benefits of dialogic reading are supported by a large volume of correlational, experimental, and intervention research (for reviews, see Flack et al., 2018; Mol et al., 2008; Noble et al., 2019; Towson et al., 2017). These studies look at a broad range of short-term outcomes pertaining to the specific books being read as well as long-term outcomes including expressive and receptive language skills and reading attitudes. For example, Lever and Sénéchal (2011) found that dialogic reading with parents improved children's story comprehension indicated by children's accuracy and linguistic complexity in oral retelling of the story. In another study, children were asked questions requiring them to label illustrations representing target vocabulary words during storybook reading, and they were able to comprehend and produce more of those words than children who simply listened to the same story without any prompted interactions (Sénéchal et al., 1995).

An important area of investigation within dialogic reading is understanding how it affects children with lower language proficiency who may, in theory, most benefit from language scaffolding provided by an adult. Hargrave and Sénéchal (2000) found that students with limited language proficiency who were exposed to dialogic reading made greater expressive vocabulary gains in both content specific vocabulary encountered in the books they read and on a standard expressive vocabulary assessment than students in a typical book reading session. Similar positive results were found with a sample of Head Start children from low-income households (Wasik et al., 2006). Overall, the extant literature shows that dialogic reading is an effective method to promote literacy and language development among children from under-resourced backgrounds (Lever & Sénéchal, 2011; McNeill & Flower, 1999; Zevenbergen, & Whitehurst, 2003).

Researchers have theorized that children's improved engagement during reading serves as a mechanism through which dialogic reading improves language learning. According to Guthrie and Klauda's (2014) well-cited framework, reading engagement consists of behavioral, emotional, and cognitive dimensions. Behavioral engagement refers to how attentive students are during the reading session; emotional engagement refers to students' feelings, interest, and enthusiasm about what they are reading; and cognitive engagement refers to a child's activated thinking in order to comprehend the story and participate in discussion. Using this framework, several studies have examined the mediating effect of engagement on language learning outcomes. For example, Zhou and Yadav (2017) found that children who engaged in dialogic reading of a storybook showed higher levels of engagement and developed a better understanding of the vocabulary and story plot. Neuman and colleagues (2019) used an eye-tracking approach and found that dialogic co-viewing, where an adult prompted children's attention using techniques such as repeating words, pointing to objects, or providing brief recaps of certain plot points, enhanced children's visual attention to the narrative content and also resulted in enhanced word learning. Taken together, these studies have established theoretical and empirical models to examine engagement and its mediating role during storybook reading.

Social Learning with Artificially Intelligent Agents

Artificial intelligence has powered agents that allow for communication using natural spoken language. These conversational agents possess different properties, including those with embodiment (e.g., robots, avatars) or not (e.g., phone-based voice assistants, smart speakers; Lee et al., 2006). Researchers have explored how children talk with and perceive conversational agents to investigate whether and how conversational agents evoke meaningful interaction. A number of studies have found that children engage in natural conversation with such agents. For example, through analyzing audio recordings of children talking with the smart speakers deployed in their home, Beneteau and colleagues (2020) identified three common purposes of children's interaction, namely entertainment, assistance, and information seeking. Children's interactions with conversational agents have corroborated findings that children attributed many human properties to the agents. For example, Xu and Warschauer (2020) found the majority of preschool-aged children perceived conversational agents to possess cognitive abilities, which they believed enabled the agents to comprehend speech. Together, these studies point to the feasibility of conversational agents as social partners (Roseberry et al., 2014).

Along these lines, studies have specifically explored the use of conversational agents to accompany children during learning processes. For example, Kory and Breazeal (2014) studied how a robot, operated by a human experimenter behind the scenes, could support children's story creation by prompting children to draw attention to the main elements of stories (e.g., what, where, who). This robot taught children the story structures and facilitated children's telling of more complex stories. Targeting slightly older children, Michaelis and Mutlu (2019) developed a robot companion to promote elementary school students' reading interest, designed to make preprogrammed comments intermittently as children read aloud and to provide non-verbal cues (e.g., eye gaze, semi-randomized idle movements) to demonstrate good listening. This in-home study found the robot motivated children to read and elicited children's social response (i.e., affliction). These studies demonstrate the role artificial intelligence may play in enrichingchildren's early literacy experience. Xu and Warschauer (2020) studied the use of an intelligent media character to engage children in science-related talk during an animated video and found that it helped children learn scientific vocabulary.

Further, several studies suggest that properly designed agents can be equally effective as human language partners. Almost all of these studies involved conversational agents that were embodied, such as robots or on-screen intelligent avatars. For example, Westlund and colleagues (2017) found that children learned unfamiliar words equally well with a robot or a human interlocutor. Hong and colleagues (2016) also demonstrated that incorporating a robot teaching assistant in a classroom led to similar levels of student reading and writing improvement as compared to having a human assistant. To our knowledge, there is only one study focusing on the comparison between a disembodied conversational agent and a human partner (Aeschlimann et al., 2020). Children collaborated with either a voice assistant (i.e., a smart speaker) or an adult experimenter in a treasure hunt game, which required children to provide necessary information to their respective collaborator. Children supplied more information to the adult experimenter than to the voice assistant. However, this study was carried out in a gameplay setting and thus was not able to answer the questions of specific language learning benefits resulting from interaction with a disembodied conversational agent during book reading.

The Present Study

This study is the first to focus on preschool-aged children's engagement with and learning from a disembodied conversational agent compared to the engagement and learning from reading with an adult. Three questions were asked:

1. What is the effect of dialogic reading with a conversational agent on children's story comprehension? (RQ1)
2. What is the effect of dialogic reading with a conversational agent on children's reading engagement? (RQ2)

3. Do changes in engagement serve as a mechanism through which dialogic reading with a conversational agent affects story comprehension? (RQ3)

To answer these questions, we conducted a two-by-two factorial experiment, with the two factors being whether children had dialogic reading or non-dialogic reading and whether children were partnered with a conversational agent or an adult during the reading. One hundred and seventeen children aged 3 to 6 were randomly assigned into one of the four conditions. Children's story comprehension was measured after reading, and their engagement was analyzed from the video recording of the reading sessions.

For RQ1 and RQ2, we hypothesized that children in the dialogic reading groups would be more engaged in the reading and comprehend the story better than those in the non-dialogic reading groups. This is expected given the advantages documented by dialogic interactions (Hargrave & Sénéchal, 2000; Şimşek & Işıkoğlu Erdoğan, 2015). However, no clear hypothesis was formed regarding the effects of reading partners. While an in-person partner has long been viewed as more natural than artificially intelligent agents (Aeschlimann et al., 2020), studies have repeatedly shown that properly designed agents enhance engagement and learning (Tewari & Canny, 2014).

For RQ3, we hypothesized that engagement would be a significant mechanism through which conversational agents enhance learning. Engagement has been posited as a key factor in enhancing reading comprehension, and engaged children are more often motivated to understand the story content with higher level of cognitive efforts (Guthrie & Klauda, 2014).

Method

Participants

One hundred and twenty-two children aged 3 to 6 years were recruited from five childcare centers serving middle-class communities and participated in the experiment (data collection: 2/2019-8/2019). To recruit these children, we reached out to the directors of the childcare centers, and with their approval, we set up a recruitment booth at each site during pick-up times to gather parent signatures and also answer any questions they may have had. Parents or guardians also completed a brief survey on demographic characteristics and information related to their child's prior experiences with conversational technologies. Five children were excluded due to data loss resulting from technological problems with the recording device, which resulted in an analytic sample consisting of 117 children (age range = 37 months to 81 months, M = 58.10 months, SD = 9.53 months). Fifty percent of the children were girls, and approximately one third were identified as White. Almost 80% of these children predominantly spoke English at home. Table 1 presents participants' background information.

Study Design

This study used a 2 (conversational agent vs. adult) x 2 (dialogic reading vs. non-dialogic reading) factorial design, where participants are randomly assigned into one of four conditions. Specifically, we utilized a randomized block design, in which participants in each school site were randomly assigned into their experimental condition. The purpose of such a design is to increase the homogeneity of experimental units, thus reducing experimental errors and increasing the power for detecting treatment factor effects. The four conditions were as follows:

- **Agent Dialogic Reading (Agent DR)** where the agent narrated the story to a child and engaged the child in dialogue by asking questions and providing feedback

- **Agent Non-Dialogic Reading (Agent Non-DR)** where the agent merely narrated the same story to a child but did not ask any questions to engage the child in dialogue
- **Human Dialogic Reading (Human DR)** where an adult narrated the story to a child and engaged the child in dialogue by asking questions and providing feedback
- **Human Non-Dialogic Reading (Human Non-DR)** where an adult merely narrated the same story to a child but did not ask any questions to engage the child in dialogue

In the "Human DR" condition, the human experimenter followed the dialogue script designed for the agent. Adherence to the script ensured that the verbal exposure in the two dialogic reading conditions (Agent DR and Human DR) was comparable, thus increasing the internal validity of the study findings.

Experimental Stimuli

The story reading materials were adopted from a commercially available picture book, "Three Bears in a Boat," authored by David Soman. The story is about three little bears who accidentally break their mother's precious seashell and then embark on an adventure to search for a new seashell. The story was chosen based on length, potential story interest, the low likelihood that participants would have read the book previously, and appropriate level of narrative complexity. The print book was 16 pages long, with each page consisting of about 6 sentences (an average of 11 words per sentence) accompanied by illustrations. We analyzed the book's narrative complexity using the rubric developed in Petersen et al. (2008) and determined that the book is appropriate for preschool children because it contains i) main characters with names, ii) specific places and times where the story took place, and iii) a clear story sequence with causes and consequences.

Both human and agent dialogic reading conditions followed the exact same dialogue script (i.e., asking the same questions and providing responsive feedback in the same manner). Nine open-ended questions were asked throughout the storytelling. Based on Blewitt and colleagues' (2009) suggestion, we incorporated a combination of 6 low-cognitive demand questions and 3 high-cognitive demand questions. For example, the following is a sentence from the story: "One day, when their mother was out, the three bears did something they really shouldn't have, and with a crash, their mother's beautiful blue seashell lay scattered in pieces across the floor." A low-cognitive demand question asked, "What did the bears break?" And the answer to that question was "seashell", which was found directly in the text. A high-cognitive demand question asked children to make an inference based on the given information in the story or to summarize the information (e.g., "How did the bears search for the seashell?"). Both the human and agent provided elaborative feedback to children's responses in a way that acknowledged what the children had said and explained the question to solidify children's understanding or clarify any confusion. For example, after children responded to the question of why the bears stopped at an island, the agent first assessed the children's answer, and then explained the reason the bears stopped there as follows, "The bears stopped at this island because they think they can find a blue seashell here. The old salty bear said the blue seashell is on the island shaped like a lumpy hat."

Children in all conditions looked at a hard copy of the storybook. A Google Home Mini device was utilized in the two agent conditions. In the dialogic reading condition, the Google Home Mini device "narrated" the story and "conversed" with the children, while in the non-dialogic reading condition, the device merely narrated the story without asking questions.

Procedure

Children met individually with a trained adult experimenter in a designated quiet area at their school for two sessions. In the first session, the participants received an Expressive One Word Picture Vocabulary Test as a pretest, which was used as the baseline measure of their expressive vocabulary skills.

In the second session, children engaged in the storybook reading activity and answered post-reading assessment questions. Prior to the reading, children interacted with the experimenter or the conversational agent through a structured dialogue, depending on their assigned condition. The dialogue involved the conversational agent or experimenter asking the children their age, favorite color, as well as simple animal fact questions and then repeating the children's responses (Agent/human: "What is your favorite color?"; child: "I like red the best."; agent/human: "Great choice! My favorite color is also red."). The purpose of including this pre-reading interaction is to build rapport between the child and the reading partner, as well as to provide children in the Agent DR condition with opportunities to practice conversing with the Google Home device.

During the reading session, children were encouraged to take responsibility for turning pages when the narration of a page was finished. An experimenter was present in the room but interfered only when/if technical issues interrupted the reading. Any time a child asked a question or initiated conversation, the experimenter simply addressed the question or replied "okay," but avoided elaborating or extending the conversation. The reading session lasted approximately 15 minutes. The reading sessions were videotaped.

Following the reading session, children's story comprehension was assessed using a battery developed by the research team. The experimenter asked questions orally, and children responded orally to the questions or identified images presented on laminated cards. Children's answers were recorded on a paper-based checklist.

Measures

Demographic Information

A parent survey was used to collect demographic information such as children's date of birth (month and year), gender, race/ethnicity, and predominant home language (i.e., English, English as a second language). This survey also asked about children's prior experience with voice technologies because this factor has been found to influence children's interactions with voice technologies (Bartneck et al., 2007). If parents indicated that their child used voice technologies at least monthly, the child was categorized as a regular user of voice technologies.

Expressive Vocabulary

Children's oral language skills are positively associated with children's comprehension of storybook reading activities (Kendeou et al., 2009). Children's baseline oral language skills were measured by the Expressive One Word Picture Vocabulary Test Fourth Edition (EOWPVT-4), which is an experimenter-administered, norm-referenced picture-naming assessment. Each child was asked to name objects, actions, and concepts that were depicted graphically. The test lasted on average 15-20 minutes. The internal reliability (Cronbach's coefficient alpha) of EOWPVT-4 for 3- to 6-year-olds is 0.95 (Martin & Brownell, 2011).

Story Comprehension

Children's comprehension of the storybook was measured as a proximal learning outcome, similar to the research approach in Zhou and Yadav (2017). A 10-item comprehension measure was developed. These 10 questions, which were different from the ten questions asked during the dialogic reading activity, assessed children's ability to 1) memorize main story events and make inferences, 2) sort narrative sequences, and 3) retell part of the story. There were eight items on memorization and inferences. For these items, an open-ended question was first asked, then if children could not recall the answer correctly, the researcher provided three multiple-choice options to choose from. Two points were given to each item that was answered correctly through free recall and one point was given if answered correctly with multiple-choice options. There was one narrative sequence sorting item, where children were asked to place images from the book in the order they occurred in the story. Children earned two points for the correct order and one point for a partially correct order. Finally, there was one item to prompt children to retell a part of the story. For this item, children earned one point each for mentioning each of the four key elements of a specific portion of the story (i.e., the four places the bears searched to find a new seashell) for a maximum of four.

An overall story comprehension score was calculated by summing the number of points across all the items; this score was used as a dependent variable for the analysis. The range is from 0 to 22 points (16 points maximum for the 8 memorization and inference-making items, 2 points maximum for the single sequence sorting item, and 4 points maximum for the story-retelling item). Cronbach's coefficient alpha was 0.87 for this story comprehension assessment.

Engagement

Children's engagement during story listening was coded from the video-taped reading sessions. Videos were divided into 5-second segments and each segment was coded by trained researchers (Willoughby, Evans, & Nowak, 2015; Zhou & Yadav, 2017).

We used a *global* and *itemized* coding system to capture children's engagement. The global coding offers a holistic view of child engagement, while the itemized coding provides fine-grained indexing of children's specific behaviors. Using two approaches captures children's engagement from different angles and establishes concurrent validity. A total of five research assistants were involved in the coding process, and the interrater-reliability (IRR) for all items was above a satisfactory level (see the detailed report below).

Global scale of child engagement. The global scale was adapted from Children's Orientation to Book Reading (Kaderavek, 2014). For each time segment, we provided a 5-point rating based on a child's posture, facial expression, eye gaze, distractibility, verbal and nonverbal comments, and responsiveness to the adult or agent's direction (e.g., turning papes). A score of 5 indicated the highest level of engagement (e.g., showing clear signs of excitement that stems from the reading, making large movements with hands to illustrate a point). A score of 3 indicated a medium level of engagement where a child did the minimum work required to follow protocols (e.g., listening, remaining seated). A score of 1 was the lowest level of engagement where a child was clearly distracted and had little interest in the reading. An average global engagement rating was calculated by the mean of the ratings across all time segments in each child's reading session. The Interrater reliability (IRR, calculated by Cohen's kappa) was 0.80 for this global coding.

The itemized coding system. We coded four items on three dimensions of engagement: vocalizations (two items), affective expressions (one item), and visual attention (one item). This three-dimensional itemized system is based on Unrau and Quirk's (2014) framework that is widely used in other similar studies concerning young children's reading engagement (e.g., Xu, Yau, & Reich, 2020). For each time segment, we coded whether each item was present (score of 1 if present and 0 if not present). To calculate the proportion of time segments each item was present, we divided the total number of time segments an item was present by the total number of time segments in the reading session.

Vocalizations. Children's vocalizations during each 5-second time segment of the reading episode were transcribed and coded as (1) relevant to the story content, which we call narrative-relevant (e.g., "I had lots of beautiful seashells."), and (2) irrelevant to the story content (e.g., "I want to have a snack"). Note that these vocalizations may be spontaneous or prompted by the agent or the human experimenter. For each type of comment, segments received a score of 1 if the comment type was present and a score of 0 if it was absent. Every time segment was coded for both types of vocalizations, but the frequency of each type of vocalization in the segment was not coded (e.g., a score of 1 was given for narrative-relevant vocalization whether the child made one narrative-relevant comment during the segment or if they made three). The IRR (Cohen's kappa) was 0.89 for narrative-relevant vocalization and 0.87 for irrelevant vocalization.

Affective expressions**.** Affective expressions were indicated by the presence or absence of children's positive expressions during each 5-second segment. Positive expression was scored (score of 1) if the child showed at least one of the following 16 expressive displays during the segment: smiling, cheering, clapping, dancing, jumping in excitement, laughing audibly, singing, showing eagerness, giggling, raising cheeks, pulling up lip corners, crinkling eyes, showing affection, smirking, speaking in a warm emotional tone, and using terms of endearment (Bai, Repetti, & Sperling, 2016). The IRR (Cohen's kappa) was 0.73 for positive expression.

Visual attention**.** Attention was coded as children's complete *visual* attention to the book during the 5-second segment. If children maintained orientation to the book during the entire time segment, their visual attention was coded as present (score of 1). If children shifted their orientation away from the book at any point, their visual attention was coded as absent (score of 0). The IRR (Cohen's kappa) for this item was 0.86.

Results

The results section first presents the descriptive statistics of outcome measures for the full sample. The data analyses for the two research questions (i.e., effects of dialogic reading with a conversational agent and the mechanisms of story comprehension from dialogic reading with a conversational agent) are then presented sequentially.

Descriptive Statistics of Outcome Measures for Full Sample

The descriptive statistics for the full sample are presented in Table 2. Children's mean story comprehension score was 10.6, indicating that these children on average correctly answered about half of the post-test questions. In terms of attention, children on average were visually attentive to the print book about 60.0% of the time. In terms of emotion, children showed obvious positive expression about 11% of the time. In terms of vocalization, children across the four conditions were observed to make narrative-relevant comments in 11.0% of the time segments, while the frequency of irrelevant comments (1.9%) was much lower. In terms of the global engagement rating, the average score was 3.0 across the four conditions, which represented a medium level of engagement in our coding system (1-5).

We also looked at the correlations between the outcome variables (Table 2). Children's story comprehension was positively correlated with their frequency of narrative-relevant comments ($r\ (115) = 0.23$, $p < 0.05$) and was negatively correlated with the frequency of irrelevant comments ($r\ (115) = -0.21$, $p < 0.05$). In terms of the relations between the global engagement rating and itemized coding (i.e., vocalizations, positive expression, visual attention), global engagement was positively correlated with narrative-relevant vocalization ($r(115) = 0.44$, $p < 0.001$), positive expression ($r\ (115) = 0.53$, $p < 0.001$), and visual attention (r(115) = 0.36, $p < 0.001$), and negatively correlated with irrelevant vocalization (r(115) = -0.20, $p < 0.05$).

Effects of Reading Condition on Story Comprehension (RQ1)

As shown in Table 3, children in the two dialogic reading (DR) groups, who interacted with either the agent or an adult, scored higher in story comprehension than the two non-dialogic reading (Non-DR) groups. In particular, children in the Agent DR condition averaged 3.2 points higher than those in the Agent non-DR condition and 2.1 points higher than those in the Human non-DR condition. There was only 0.1 point difference in story comprehension score between the Agent-DR and Human-DR groups. Analysis of variance (ANOVA) revealed a significant condition effect on story comprehension ($F(3, 116) = 2.89$, $p < 0.05$). Bonferroni-adjusted post hoc tests suggested that the two dialogic reading groups performed equally well in the story comprehension assessment.

Regression analysis was subsequently performed to identify the magnitude of the effects of dialogic reading and conversational agents (Table 4). Model 1 included the dialogic reading and agent factors, and Model 2 included the interaction between these two factors (DR*Agent) to examine whether the effects of dialogic reading varied by the type of reading partner. In both models, we controlled for children's baseline language proficiency, age, and whether children were regular users of conversational technologies to increase the precision of the estimates.

Results of Model 1 suggested that dialogic reading significantly increased children's story comprehension by 0.60 SD ($\beta = 0.60$, $p < 0.001$), while the type of reading partner (adult or agent) did not have a significant impact ($\beta = -0.17$, $p = 0.15$). Model 2 further suggested that the agent induced a comparable level of positive effect on children's story comprehension as an adult reader ($\beta = 0.04$, $p = 0.87$). In both models, children's age in months and expressive vocabulary levels were significant covariates.

Because previous literature posited that children with lower language ability may particularly benefit from dialogic scaffolding (Lever & Sénéchal, 2011), we examined whether the effects of dialogic reading varied based on children's language proficiency. In our regression model, we included a three-way-interaction among the dialogic reading factor, the agent factor, and children's expressive vocabulary score assessed by EOWPVT-4 (Model 3, Table 4). The results indicated that children's expressive vocabulary score was not a significant moderator in the linkages between children's reading conditions and their story comprehension outcomes.

Effects of Reading Condition on Engagement (RQ2)

The descriptive statistics of children's reading engagement are displayed in Table 3. Narrative-relevant vocalization was more frequently observed among children in the dialogic reading conditions, while irrelevant vocalization only rarely occurred across the four conditions. ANOVA revealed significant differences among reading conditions on narrative-relevant vocalizations ($F(3, 116) = 26.15$, $p < 0.001$) and irrelevant vocalizations ($F(3, 116) = 3.93$, $p = 0.05$). Highest instances of positive expressions were found among the Agent DR reading condition, and visual attention (i.e., children's eyes fixating on the book) was generally lower for the two dialogic reading groups. Although the global engagement rating was not found to differ significantly across conditions with ANOVA, Bonferroni-adjusted post hoc tests indicated that the Agent DR group had a significantly higher global engagement rating than the Agent non-DR group ($p < 0.05$).

The regression results on the effects of condition on reading engagement are presented in Table 5. For all engagement variables, we first tested the individual effects of the dialogic reading and agent factors, and then included an interaction term between dialogic reading and agent. Children's age, expressive vocabulary, and prior use of conversational technologies were included as covariates.

Examining the global engagement rating, Model 1 suggested that dialogic reading did not affect children's overall engagement significantly ($\beta = 0.27$, $p = 0.17$), nor did whether children read with the agent or adult ($\beta = -0.11$, $p = 0.54$). However, Model 2 revealed that the effect of agent partner on engagement was, in fact, dependent upon whether children were engaged in dialogic reading. While the Agent non-DR condition led to lower level of engagement at a marginally significant level ($\beta = -0.48$, $p < 0.1$), having dialogic reading with the agent enhanced the engagement level significantly ($\beta = 0.64$, $p < 0.05$). Children who had higher expressive vocabulary scores and those who used conversational technologies regularly were observed to have higher levels of engagement compared to their counterparts.

We then examined the scores resulting from the itemized coding of engagement (Models 3-10 in Table 5). In terms of vocalizations, dialogic reading unsurprisingly led to a significantly higher level of narrative-related vocalization ($\beta = 0.99$, $p < 0.001$, Model 3), and reading with an agent (either dialogically or non-dialogically) was associated with a decreased level of narrative-relevant vocalization ($\beta = -0.63$, $p < 0.001$, Model 3). Reading with an agent appeared to also result in less irrelevant vocalization ($\beta = -0.65$, $p < 0.001$, Model 5). When including the interaction term (Model 6), the model further suggested that dialogic reading helped reduce the instances of irrelevant vocalization ($\beta = -0.56$, $p < 0.05$). Regular users of conversational technologies were observed to have fewer instances of irrelevant vocalization.

Children's positive expressions were not affected by the dialogic reading or the agent factor (Model 7 and Model 8), nor was children's visual attention (Model 9 and Model 10).

Engagement as a Mediator between Condition and Story Comprehension (RQ3)

Finally, we conducted structural equation modeling (SEM) to formally test whether engagement explains the effect of reading condition on story comprehension. Given that the narrative-relevant and irrelevant vocalizations are the only two coded variables significantly correlated with comprehension (see Table 2), we specifically focused on these two variables in the SEM analysis. This choice was also supported by the rationale of the purpose of dialogic reading, which is to increase the amount of vocalization (Hargrave & Sénéchal, 2000). We used reading condition as a multi-categorical predictor with the Agent DR condition as the reference group.

With this rationale in mind, we fitted a model with narrative-relevant vocalizations and irrelevant vocalizations as mediators between the different group assignments and our outcome, comprehension. Our model specification included all three groups having direct paths to the outcome, as well as indirect paths through vocalizations to the outcome (See Figure 2). Three covariates in the regression analysis above, participant age, expressive vocabulary score, and prior experience with conversational technologies, were also included. Specifically, there were paths from the three covariates to the dependent variable (i.e., story comprehension) and to the mediators (i.e., narrative relevant vocalizations and irrelevant vocalizations). This model has great fit ($\chi^2(7) = 1.1, p = .30$, CFI = 1.00, TLI = 0.99, RMSEA = .02[.00, .25], SRMR = 0.01), according to Keith (2014) .

Children's group assignment had differing relationships to each of the mediators in comparison to the reference group (i.e., Agent DR). Participants in the Human DR group had, on average, higher rates of narrative-relevant vocalizations (β = 0.23, $p < 0.05$) than the Agent DR group, while children in the Agent non-DR group had, on average, lower rates of narrative-relevant comments (β = -0.49, $p < .001$). Children in the Human non-DR group had, on average, higher rates of irrelevant comments (β = 0.37, $p < 0.001$). As for the mediators, the narrative comments mediator was positively associated with the outcome and achieved marginal significance (β = .12, $p < 0.1$), and the irrelevant comments mediator was negatively associated with the outcome (β = -0.15, $p < 0.05$). The Agent non-DR group was the only group that was directly related to the story comprehension outcome (β = -0.20, $p < 0.05$). Participant age (β = 0.27, $p < 0.01$) and expressive vocabulary score (β = 0.58, $p < 0.001$) were also significantly associated with story comprehension. Participant prior agent usage was associated children's irrelevant vocalizations at a marginally significant level (β = -0.15, $p < 0.1$).

The analysis of the total effects corroborated the regression results indicating that Agent non-DR condition negatively impacted story comprehension (a total effect of β = -0.23, $p < 0.001$). The lower comprehension score resulting from assignment to the Agent non-DR group compared to the Agent DR group could be partially explained by the decreased level of narrative-relevant vocalization. Specifically, the indirect path from the Agent non-DR condition through narrative-relevant vocalization to story comprehension was -0.06 ($p < 0.1$). Additionally, there was a significant direct path from Agent non-DR to story comprehension (β = -0.20, $p < 0.05$). This direct effect suggests that there were other mediating factors that were not captured by the narrative-relevant vocalizations. On the other hand, the lower comprehension score resulting from the Human non-DR condition was completely mediated by an indirect effect through irrelevant vocalization (β = -0.06, $p < 0.05$). There is no significant direct effect or indirect effect from Human DR to story comprehension as compared to the reference group (i.e., Agent DR).

Discussion

The purpose of this study was to examine the effects of dialogic reading with a disembodied conversational agent or an adult on children's reading engagement and story comprehension. Dialogic reading, during which children are read a storybook and engaged in relevant conversation, has long been viewed as an ideal context to foster children's early language and literacy development. Our study demonstrated that a properly designed conversational agent can assume the role of a dialogue partner during children's storybook reading with benefits comparable to that of an adult dialogue partner. Given that smart speakers are affordable and already owned by many families, these findings are promising for the deployment of this technology in supporting children's language development, especially for children from families who have limited time, language skills, or resources to themselves engage in dialogic reading

Our first research question examined the effects of dialogic reading and conversational agents on children's story comprehension. Consistent with prior research, we found that children who listened to a story together with dialogue outperformed those who just listened to the story without dialogue (Flack et al., 2018; Mol et al., 2008; Noble et al., 2019; Towson et al., 2017). This validated the design of dialogic strategies (questions and feedback) used in our study and the use of a conversational agent as dialogic reading partners. Further, our results suggest that the conversational agent replicated the benefits of dialogue resulting from an adult partner, given that the effects of dialogic reading did not vary by dialogue with an adult or the agent. This is in line with the emerging body of research demonstrating the potential benefits of artificially intelligent learning companions. However, in contrast to prior research on these benefits that typically involved robots (e.g., Breazeal et al., 2016; Westlund et al., 2017), the conversational agent used in our study was disembodied and thus not capable of utilizing non-verbal expressions to facilitate the dialogue. However, we did not detect significant interaction between the children's baseline language proficiency and the effects of dialogic reading with an agent on their story comprehension. While the non-significant interaction effect may suggest the robustness of our results across subgroups with varying language proficiency, it may, on the other hand, result from the fact that even the lower proficiency children in our sample were within the norm for their chronological age. Specifically, the median age-adjusted EOWPVT-4 score was 115, which is equivalent to an 83% percentage rank among the national, normative population and the score of the first quartile was 103, which is still above 50% percentage rank. As such, the homogeneity of this sample's language proficiency may have obscured our ability to uncover the heterogenous effects of dialogic reading with conversational agents.

We also uncovered the effects of dialogic reading and conversational agents on children's engagement. An interesting pattern emerged in terms of global engagement. Non-dialogic reading with an agent is detrimental to children's overall engagement. However, dialogue with an agent increases children's engagement to the levels found when children read with a human. This finding provides empirical support for the notion that opportunities for contingent dialogue with agents may simulate the social presence of a human partner and bring about similar benefits for engaged learning (Brunick et al., 2016).

When examining vocalizations, as expected, dialogic reading resulted in significantly higher levels of narrative-relevant vocalization. This suggests that children were receptive to such reading strategies, as demonstrated repeatedly from studies in the face-to-face setting or computer-based environments (e.g., Calvert et al., 2019; Peebles et al., 2018). Interestingly, it appeared that dialogic reading also reduced the instances of irrelevant vocalizations that may be an indicator of distraction (Reich et al., 2019). This may be because dialogic reading "directs" children's vocalizations along the narrative, thus helping children focus on the reading. In regard to the conversational agents, children did not generate vocalizations as frequently, neither narrative-relevant nor irrelevant, as those reading with an adult. This finding of fewer child vocalizations with an agent was consistent with Aeschlimann et al. (2020), who showed that preschool-aged children were less likely to provide vocal information to a smart speaker than to an adult researcher. There are two possible explanations for this: either children do not know how to appropriately talk to non-human agents (Beneteau et al., 2019; Cheng et al., 2018) or they are not interested in doing so (Cameron et al., 2015). The social presence of a human partner may encourage children to provide on-topic responses but may also invite children to voluntarily extend the conversation beyond the reading context. Though bringing up irrelevant comments may be developmentally appropriate for young children (Godwin et al., 2016) and generate excitement (Xu & Warschauer, 2020), our study showed that doing so shifted children's attention away from the story and dampened children's learning. However, it should be noted the irrelevant vocalizations occurred only rarely (i.e., among 2% of the segments). As such, the true effects of dialogic reading with agents on irrelevant vocalizations may have become harder to detect due to the lessened variability in these variables. We should interpret the results on this variable with caution.

In our analysis, dialogue did not lead to a significantly higher level of visual attention during reading, while other studies suggested that children more frequently fixated on the educational content displayed on the screen when an adult co-viewer commented on the content

(Neuman et al., 2019). However, in Neuman et al., the comments were not designed to elicit children's verbal responses, but rather to label and explain the vocabulary. As such, we speculated that the dialogue moments in our study, which elicited verbal responses, may have triggered children to look at their reading partner (either the agent or adult) as they replied to the questions and listened to feedback, thus deviating the children's eye fixation from the book. To test this speculation, we recalculated children's visual attention by including their time spent looking at their respective conversational partner. However, the new visual attention variable remained consistent with the original one: children in the dialogic conditions still had a relatively lower level of visual attention than non-dialogic conditions. Specifically, Agent DR condition had 0.76 of the time looking at the book or the agent (SD = 0.13), Agent non-DR condition were attentive for 0.80 of the time (SD = 0.15), Human DR condition for 0.70 (SD = 0.18), and Human non-DR for 0.76 (SD = 0.17). As such, it was not evident that the reduced visual attention to the book in the dialogic conditions was attributed to children looking at their conversational partner. Nevertheless, we did notice from our video recordings that children shifted their eyes away from the book and looked straight ahead when they were thinking hard to formulate their responses. Supporting this observation, Table 2 shows that instances of visual attention were negatively correlated with both narrative-relevant vocalization ($r = -0.31$, $p < 0.001$) and irrelevant vocalization ($r = -0.27$, $p < 0.01$). While many studies have shown a significant positive correlation between children's fixation on the book and their learning (Justice et al., 2008), our findings suggest the importance of a holistic view in understanding children's visual attention and engagement during conversation-rich reading activities.

Our mediation analysis corroborated the effects of condition on children's vocalization and points to interesting mechanisms through which dialogic reading with conversational agents may support language development. The advantage of dialogic reading with conversational agents is explained through a two-pronged mechanism: the increased narrative-relevant vocalizations (compared to "non-DR" groups) and the decreased irrelevant vocalizations (compared to the "Human" groups). The first part of this finding replicates Calvert et al. (2019), which indicated that asking children questions during televised stories promoted learning because of children's increased relevant talk. The second part of the finding suggests that agents can enhance learning through limiting off-task behaviors.

Taken together, our study provides evidence that disembodied conversational agents can effectively engage children in dialogic reading activities. At an applied level, these findings suggest we take advantage of the prevalence of smart speakers in children's homes and integrate these devices as part of children's informal learning experiences. While we do not intend to use artificial intelligence to replace children's storytime with their parents or teachers, properly designed agents can sometimes stand in the role as an engaging dialogic partner for children when adults are otherwise unavailable. Moreover, this kind of conversational agent can be used as a model for parents or teachers' looking to improve their dialogic reading strategies.

Limitations and Future Directions

There are several directions future studies may build on and extend from the current study. First, the current study was carried out in a controlled manner where children had dialogic reading with scripted questions and feedback. This design limited the ecological validity of the findings, although it increased the internal validity of researching the effects of human versus agent reading partners by holding the conversation consistent. Future studies could be carried out in a more naturalistic setting, in which a familiar adult reads with the child as they normally would. We would expect variation in how much and how well dialogic questions and feedback were utilized by the adult. As such, we may compare virtual agents against skilled and unskilled human partners who are not constrained to a script. Second, as discussed before, the children participated in our study are from higher socioeconomic backgrounds. Future research may want to further investigate whether dialogic reading with an agent can help at risk children from lower socioeconomic backgrounds. These at-risk children may lag in language and literacy development, potentially making dialogic scaffolding particularly valuable for them. Third, our study focused on immediate outcomes after a one-time, short intervention, while future research may want to implement the agent dialogic reading partner for a longer period of time at schools, public libraries, or homes. Given that other longer term dialogic reading interventions (typically lasting 4-8 weeks) have proven successful in promoting children's general language ability, such as receptive and expressive language, vocabulary, and narrative skills, it is plausible that agent-based interventions could also bring similar benefits to children.

Conclusion

This study examined whether and how a smart speaker based conversational agent can facilitate language development by engaging children in dialogic reading. Our findings suggest that dialogic reading with a disembodied conversational agent can replicate the benefits of an adult partner in facilitating story comprehension. Furthermore, we found that some advantages of dialogic reading with an agent arose from children's enhanced narrative-relevant vocalizations and reduced irrelevant vocalizations. Given that disembodied conversational agents are already prevalent because of their affordability, such agents represent a potentially scalable, cost-effective tool for enriching preschool-aged children's early literacy development.

References

Aeschlimann, S., Bleiker, M., Wechner, M., & Gampe, A. (2020). Communicative and social consequences of interactions with voice assistants. *Computers in Human Behavior, 112*, 106466.

Arnold, D. H., Lonigan, C. J., Whitehurst, G. J., & Epstein, J. N. (1994). Accelerating language development through picture book reading: replication and extension to a videotape training format. *Journal of educational psychology*, *86*(2), 235.

Arnold, D. S., & Whitehurst, G. J. (1994). *Accelerating language development through picture book reading: A summary of dialogic reading and its effect.* In D. K. Dickinson (Ed.), *Bridges to literacy: Children, families, and schools* (p. 103–128). Blackwell Publishing.

Bai, S., Repetti, R. L., & Sperling, J. B. (2016). Children's expressions of positive emotion are sustained by smiling, touching, and playing with parents and siblings: A naturalistic observational study of family life. *Developmental psychology*, *52*(1), 88.

Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *Ai & Society*, *21*(1-2), 217-230.

Beneteau, E., Guan, Y., Richards, O. K., Zhang, M. R., Kientz, J. A., Yip, J., & Hiniker, A. (2020). Assumptions Checked: How Families Learn About and Use the Echo Dot. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(1), 1-23.

Beneteau, E., Richards, O. K., Zhang, M., Kientz, J. A., Yip, J., & Hiniker, A. (2019, May). Communication breakdowns between families and Alexa. In *Proceedings of the 2019*

CHI Conference on Human Factors in Computing Systems (pp. 1-13).

Berk, L. E., & Winsler, A. (1995). Scaffolding Children's Learning: Vygotsky and Early Childhood Education. NAEYC Research into Practice Series. Volume 7. National Association for the Education of Young Children, 1509 16th Street, NW, Washington, DC 20036-1426 (NAEYC catalog# 146).

Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology*, 101(2), 294–304.

Bodrova, E., & Leong, D. J. (2005). High quality preschool programs: What would Vygotsky say?. *Early Education and Development*, *16*(4), 435-444.

Breazeal, C., Harris, P. L., DeSteno, D., Kory Westlund, J. M., Dickens, L., & Jeong, S. (2016). Young children treat robots as informants. *Topics in cognitive science*, *8*(2), 481-491.

Brunick, K. L., Putnam, M. M., McGarry, L. E., Richards, M. N., & Calvert, S. L. (2016). Children's future parasocial relationships with media characters: the age of intelligent characters. *Journal of Children and Media*, *10*(2), 181-190.

Calvert, S. L., Putnam, M. M., Aguiar, N. R., Ryan, R. M., Wright, C. A., Liu, Y. H. A., &

Barba, E. (2019). Young Children's Mathematical Learning From Intelligent Characters. *Child Development*.

Cameron, D., Fernando, S., Collins, E., Millings, A., Moore, R., Sharkey, A., ... & Prescott, T. (2015, January). Presence of life-like robot expressions influences children's enjoyment of human-robot interactions in the field. In *Proceedings of the AISB Convention 2015*. The Society for the Study of Artificial Intelligence and Simulation of Behaviour.

Cheng, Y., Yen, K., Chen, Y., Chen, S., & Hiniker, A. (2018, June). Why doesn't it work?

voice-driven interfaces and young children's communication repair strategies. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (pp. 337-348).

Farver, J. A. M., Xu, Y., Lonigan, C. J., & Eppe, S. (2013). The home literacy environment and Latino head start children's emergent literacy skills. *Developmental Psychology*, *49*(4), 775.

Flack, Z. M., Field, A. P., & Horst, J. S. (2018). The effects of shared storybook reading on word learning: A meta-analysis. *Developmental psychology*, *54*(7), 1334.

Garg, R., & Sengupta, S. (2020, April). Conversational Technologies for In-home Learning: Using Co-Design to Understand Children's and Parents' Perspectives. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13)

Godwin, K. E., Almeda, M. V., Seltman, H., Kai, S., Skerbetz, M. D., Baker, R. S., & Fisher, A. V. (2016). Off-task behavior in elementary school children. *Learning and Instruction*, *44,* 128-143.

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child development*, *90*(3), 985-992.

Guthrie, J. T., & Klauda, S. L. (2014). Effects of classroom practices on reading comprehension, engagement, and motivations for adolescents. *Reading research quarterly*, *49*(4), 387-416.

Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly*, *15*(1), 75-90.

Hong, Z. W., Huang, Y. M., Hsu, M., & Shen, W. W. (2016). Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL classrooms. *Journal of Educational Technology & Society*, *19*(1), 337-349.

John-Steiner, V., & Mahn, H. (1996). Sociocultural approaches to learning and development: A Vygotskian framework. *Educational psychologist*, *31*(3-4), 191-206.

Justice, L. M., Pullen, P. C., & Pence, K. (2008). Influence of verbal and nonverbal references to print on preschoolers' visual attention to print during storybook reading. *Developmental Psychology*, *44*(3), 855.

Kaderavek, J. N., Guo, Y., & Justice, L. M. (2014). Validity of the children's orientation to book reading rating scale. *Journal of Research in Reading*, *37*(2), 159-178.

Keith, T. Z. (2014). *Multiple Regression and Beyond: An Introduction to Multiple Regression and Structural Equation Modeling*. Routledge.

Kendeou, P., Van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, *101*(4), 765.

Kory, J., & Breazeal, C. (2014, August). Storytelling with robots: Learning companions for preschool children's language development. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (pp. 643-648). IEEE.

Lee, K. M., Jung, Y., Kim, J., & Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International Journal of Human-Computer Studies*, *64*(10), 962-973.

Lever, R., & Sénéchal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology*, *108*(1), 1-24.

Lovato, S. B., & Piper, A. M. (2019). Young Children and Voice Search: What We Know From Human-Computer Interaction Research. *Frontiers in Psychology, 10*, 8.

Martin, N. A., & Brownell, R. (2011). *Expressive one-word picture vocabulary test-4 (EOWPVT-4)*. Academic Therapy Publications.

McNeill, J., & Flower, S. (1999). Let's Talk: Encouraging Mother-Child Conversations During Story Reading. *Journal of Early Intervention, 22*(1), 51-69.

Michaelis, J. E., & Mutlu, B. (2018). Reading socially: Transforming the in-home reading experience with a learning-companion robot. *Science Robotics*, *3*(21).

Mol, S. E., Bus, A. G., De Jong, M. T., & Smeets, D. J. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early education and development*, *19*(1), 7-26.

Neuman, S. B., Samudra, P., Wong, K. M., & Kaefer, T. (2019). Scaffolding attention and partial word learning through interactive coviewing of educational media: An eye-tracking study with low-income preschoolers. *Journal of Educational Psychology*.

Noble, C., Sala, G., Peter, M., Lingwood, J., Rowland, C., Gobet, F., & Pine, J. (2019). The impact of shared book reading on children's language skills: A meta-analysis. *Educational Research Review*, *28*, 100290.

Peebles, A., Bonus, J. A., & Mares, M. L. (2018). Questions+ answers+ agency: Interactive touchscreens and Children's learning from a socio-emotional TV story. *Computers in Human Behavior*, *85*, 339-348.

Petersen, D. B., Gillam, S. L., & Gillam, R. B. (2008). Emerging procedures in narrative assessment: The index of narrative complexity. *Topics in Language Disorders*, *28*(2), 115-130.

Phillips, B. M., & Lonigan, C. J. (2009). Variations in the home literacy environment of preschool children: A cluster analytic approach. *Scientific Studies of Reading*, *13*(2), 146-174.

Reich, S. M., Yau, J. C., Xu, Y., Muskat, T., Uvalle, J., & Cannata, D. (2019). Digital or print? A comparison of preschoolers' comprehension, vocabulary, and engagement from a print book and an e-Book. *AERA Open*, *5*(3), 2332858419878389.

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, *85*(3), 956-970.

Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018, June). " Hey Alexa, What's Up?" A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 857-868).

Sénéchal, M., Thomas, E., & Monker, J. (1995). Individual differences in 4-Year-Old Children's Acquisition of Vocabulary During Storybook Reading. *Journal of Educational Psychology*, *87*(2), 218-229.

Sengupta, S., & Garg, R. (2019, March). Impact of Voice-based Interaction on Learning Practices and Behavior of Children. In *IUI Workshops*.

Şimşek, Z. C., & Işıkoğlu Erdoğan, N. (2015). Effects of the dialogic and traditional reading techniques on children's language development. *Procedia - Social and Behavioral Sciences*, *197*, 754-758.

Tate, T. P., Collins, P., Xu, Y., Yau, J. C., Krishnan, J., Prado, Y., ... & Warschauer, M. (2019). Visual-Syntactic Text Format: Improving Adolescent Literacy. *Scientific Studies of Reading*, *23*(4), 287-304.

Tewari, A., & Canny, J. (2014, April). What did spot hide? a question-answering game for preschool children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1807-1816).

Towson, J. A., Fettig, A., Fleury, V. P., & Abarca, D. L. (2017). Dialogic reading in early childhood settings: A summary of the evidence base. *Topics in Early Childhood Special Education*, *37*(3), 132-146.

Unrau, N. J., & Quirk, M. (2014). Reading motivation and reading engagement: Clarifying commingled conceptions. *Reading Psychology*, *35*(3), 260-284.

Vygotsky, L. S. (2012). Thought and language. MIT press.

Wasik, B., Bond, M.A., & Hindman, A. (2006). The Effects of a Language and Literacy Intervention on Head Start Children and Teachers. *Journal of Educational Psychology, 98*(1), 63-74.

Westlund, J. M. K., Dickens, L., Jeong, S., Harris, P. L., DeSteno, D., & Breazeal, C. L. (2017). Children use non-verbal cues to learn new words from robots as well as people. *International Journal of Child-Computer Interaction*, *13*, 1-9.

Whitehurst, G. J. (1992). Dialogic reading: An effective way to read to preschoolers. *Reading Rockets. http://www. readingrockets. org/article/dialogic-reading-effective-way-read-preschoolers*.

Willoughby, D., Evans, M. A., & Nowak, S. (2015). Do ABC eBooks boost engagement and learning in preschoolers? An experimental study comparing eBooks with paper ABC and storybook controls. *Computers & Education*, *82*, 107-117.

Xu, Y., & Warschauer, M. (2020, June). Exploring young children's engagement in joint reading with a conversational agent. In *Proceedings of the Interaction Design and Children*

Conference (pp. 216-228).

Xu, Y., Yau, J. C., & Reich, S. M. (2020). Press, swipe and read: Do interactive features facilitate engagement and learning with e-Books?. *Journal of Computer Assisted Learning*, 1-14.

Yuan, Y., Thompson, S., Watson, K., Chase, A., Senthilkumar, A., Brush, A. B., & Yarosh, S. (2019). Speech interface reformulations and voice assistant personification preferences of children and parents. *International Journal of Child-Computer Interaction*, *21*, 77-88.

Zevenbergen, A. A., & Whitehurst, G. J. (2003). *Dialogic reading: A shared picture book reading intervention for preschoolers.* In A. van Kleeck, S. A. Stahl, & E. B. Bauer (Eds.), *Center for Improvement of Early Reading Achievement, CIERA. On reading books to children: Parents and teachers* (p. 177–200). Lawrence Erlbaum Associates Publishers.

Zhou, N., & Yadav, A. (2017). Effects of multimedia story reading and questioning on preschoolers' vocabulary learning, story comprehension and reading engagement. *Educational Technology Research and Development*, *65*(6), 1523-1545.

Table 2.1

Background Information by Condition

	Full Sample	Agent DR	Agent non-DR	Human DR	Human non-DR	ANOVA/ CHI Square
Age	58.10	59.50	57.59	58.29	56.91	F (3, 116) = 2.09,
	(9.53)	(8.8)	(10.41)	(9.12)	(9.50)	*p* = 0.12
EOWPVT	69.18	70.58	70.70	66.71	68.77	F (3, 116) = 0.34,
	(17.22)	(17.43)	(19.71)	(17.09)	(14.82)	*p* = 0.80
Predominant Home Language						X^2 (3) = 1.40, *p* = 0.70
English	78.63%	75.76%	85.19%	80.65%	73.08	
Other	21.37%	24.24%	14.81%	19.35%	26.92%	
Female	49.57%	57.58%	48.15%	48.39%	42.31%	X^2 (3) = 1.19, *p* = 0.75
Race						X^2 (18) = 18.28,
White	36.75%	33.33%	44.44%	32.26%	38.46%	*p* = 0.44
Asian	30.77%	27.27%	33.33%	41.94%	19.23%	
Hispanic	6.84%	12.12%	3.70%	0.00%	11.54%	
Black	0.85%	0.00%	0.00%	3.23%	0.00%	
Two or more	21.37%	24.24%	11.11%	22.58%	26.92%	
Other	1.71%	0.00%	3.70%	0.00%	3.85%	
Decline	0.85%	3.03%	3.70%	0.00%	00.00%	
Regular Conversational Agent Usage						X^2 (3) = 0.54, *p* = 0.91
Yes	43.59%	45.45%	40.74%	48.39%	38.46%	
No	55.56%	54.55%	59.26%	51.61%	57.69%	
Decline	0.85%	0.00%	0.00%	0.00%	3.85%	
N	117	33	27	31	26	

Note. Standard deviation in parentheses.

Table 2.2

Descriptive Statistics of Outcome Measures for Full Sample

	Comp	Global	RV	IV	PE	Mean	SD	Range
Comprehension	1					10.56	5.29	(0, 22)
Global Eng	0.15	1				3.03	0.19	(2.32, 3.71)
Relevant Voc	0.23*	0.44***	1			0.11	0.09	(0, 0.34)
Irrelevant Voc	-0.21*	-0.20*	0.10	1		0.02	0.04	(0, 0.22)
Positive Exp	-0.03	0.53***	0.41***	0.05	1	0.11	0.17	(0, 1)
Attention	0.01	0.36***	-0.31***	-0.27**	-0.01	0.75	0.16	(0.27, 0.98)

Note. Coefficients are Pearson correlations.
*** $p < 0.001$. ** $p < 0.01$. * $p < 0.05$.

Table 2.3

Outcome Measures by Condition

	Agent DR	Agent non-DR	Human DR	Human non-DR	ANOVA
Story comprehension					
Mean (SD)	11.74 (4.95)[a]	8.52 (5.08)[b]	11.86 (5.69)[a]	9.63 (4.53)[b]	$F(3, 116) = 2.89$,
Median [Min, Max]	12 [2, 20]	8 [1, 18]	11.33 [0, 22]	9 [2, 20]	$p < 0.05$*
Global engagement rating					
Mean (SD)	3.07 (0.15)[a]	2.97 (0.16)[b]	3.03 (0.20)[a b]	3.05 (0.21)[a]	$F(3, 116) = 1.83$,
Median [Min, Max]	3.05 [2.75, 3.50]	2.99 [2.54, 3.27]	3.07 [2.32, 3.35]	3.04 [2.54, 3.71]	$p = 0.15$
Relevant vocalization					
Mean (SD)	0.13 (0.04)[a]	0.03 (0.06)[b]	0.18 (0.07)[c]	0.09 (0.11)[a]	$F(3, 116) = 26.15$,
Median [Min, Max]	0.12 [0, 0.24]	0 [0, 0.23]	0.16 [0.02, 0.32]	0.04 [0, 0.34]	$p < 0.001$***
Irrelevant vocalization					
Mean (SD)	0.01 (0.01)[a]	0.01 (0.02)[a]	0.02 (0.04)[b]	0.04 (0.06)[c]	$F(3, 116) = 3.93$,
Median [Min, Max]	0 [0, 0.06]	0 [0, 0.08]	0.01 [0, 0.15]	0.01 [0, 0.22]	$p = 0.05$*
Positive expression					
Mean (SD)	0.14 (0.21)[a]	0.10 (0.16)[a]	0.10 (0.12)[a]	0.09 (0.15)[a]	$F(3, 116) = 0.36$,
Median [Min, Max]	0.05 [0, 1]	0.01 [0, 0.55]	0.04 [0, 0.39]	0.04 [0, 0.69]	$p = 0.78$
Visual attention					
Mean (SD)	0.74 (0.13)[a]	0.79 (0.15)	0.70 (0.17)[a]	0.76 (0.17)[a]	$F(3, 116) = 1.61$,
Median [Min, Max]	0.75 [0.43, 0.96]	0.83 [0.44, 0.98]	0.74 [0.27, 0.93]	0.82 [0.48, 0.98]	$p = 0.19$

Note. Post-hoc pair-wise comparisons with Bonferroni correction were conducted. The same letter superscriptions denoting means that were not significantly different from each other at $p < 0.05$ level.
*** $p < 0.001$. ** $p < 0.01$. * $p < 0.05$.

Table 2.4

Regression Analysis of The Condition Effects on Story Comprehension

	Story Comprehension		
	Model 1	Model 2	Model 3
DR	**0.60*** **	**0.58** **	**0.41***
	(0.13)	(0.19)	(0.19)
Agent	-0.17	-0.20	-0.27
	(0.15)	(0.19)	(0.19)
DR*Agent		0.04	0.12
		(0.24)	(0.26)
DR*Expressive Vocab			0.32
			(0.22)
Agent* Expressive Vocab			0.15
			(0.21)
DR*Agent* Expressive Vocab			-0.25
			(0.27)
Age	**0.20***	**0.20***	**0.20***
	(0.08)	(0.08)	(0.08)
Expressive Vocab	**0.59*** **	**0.60*** **	**0.42***
	(0.07)	(0.07)	(0.18)
Prior Agent Usage	0.03	0.03	-0.07
	(0.12)	(0.12)	(0.13)
Intercept	**-0.25†**	**-0.24†**	-0.13
	(0.14)	(0.16)	(0.16)
R^2	0.63	0.63	0.63

Note. Standardized coefficients reported. Standard errors in parentheses. Statistically significant coefficients bolded.
*** $p < .001$. ** $p < .01$. * $p < .05$. † $p < 0.1$.

Table 2.5

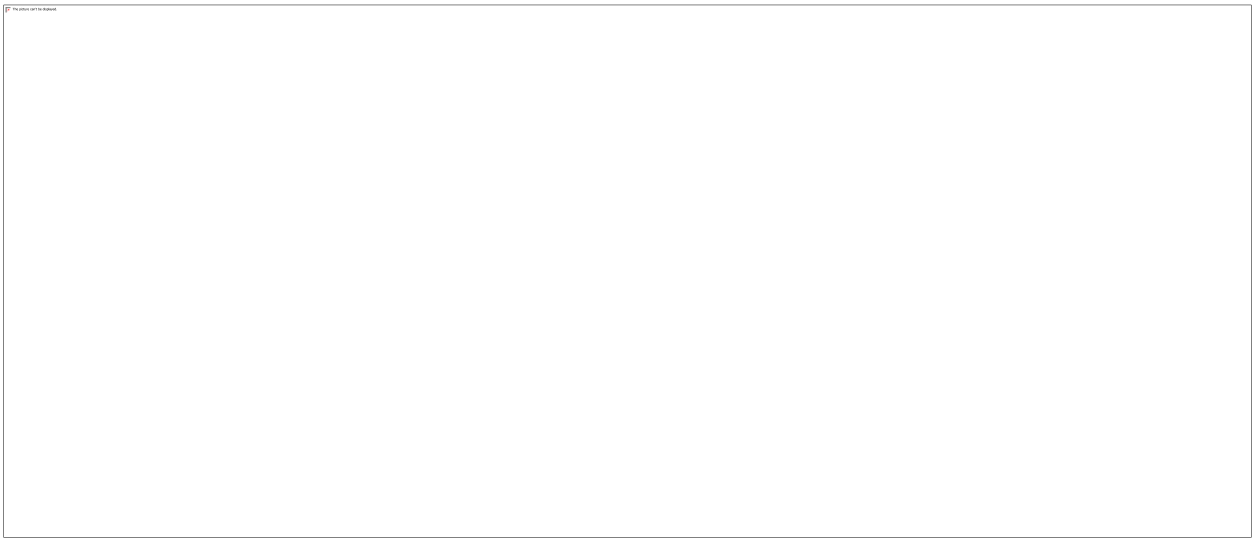
Effects of Condition on Engagement

	Global Engagement		Relevant Voc		Irrelevant Voc		Positive Exp		Visual Attention	
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
DR	0.27	-0.08	**0.99*** **	**0.88*** **	-0.28	**-0.56***	0.22	0.14	-0.31	-0.41
	(0.19)	(0.78)	(0.16)	(0.23)	(0.18)	(0.26)	(0.20)	(0.29)	(0.19)	(0.28)
Agent	-0.11	**-0.48†**	**-0.63*** **	**-0.75** **	**-0.65*** **	**-0.95*** **	0.15	0.07	0.22	0.11
	(0.19)	(0.28)	(0.16)	(0.24)	(0.18)	(0.27)	(0.20)	(0.30)	(0.19)	(0.27)
DR*Agent		**0.64***		0.21		0.52		0.14		0.19
		(0.30)		(0.32)		(0.36)		(0.39)		(0.38)
Age	-0.08	-0.07	-0.04	-0.03	-0.08	-0.06	-0.13	-0.13	-0.04	-0.04
	(0.10)	(0.10)	(0.09)	(0.09)	(0.10)	(0.09)	(0.11)	(0.11)	(0.10)	(0.10)
Exp Vocab	**0.17†**	**0.17†**	0.10	0.10	-0.24	-0.24	0.13	0.13	0.04	0.04
	(0.07)	(010)	(0.08)	(0.08)	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.10)
Prior Usage	**0.38†**	**0.37†**	0.19	0.19	**-0.28†**	**-0.30†**	0.00	0.00	-0.15	-0.16
	(0.20)	(0.19)	(0.17)	(0.18)	(0.15)	(0.16)	(0.20)	(0.21)	(0.20)	(0.20)
Intercept	-0.20	0.00	-0.24	-0.17	**0.69** **	**0.85*** **	-0.12	-0.07	-0.02	0.04
	(0.23)	(0.25)	(0.19)	(0.21)	(0.21)	(0.24)	(0.23)	(0.26)	(0.22)	(0.25)
R^2	0.08	0.11	0.36	0.36	0.21	0.23	0.05	0.05	0.17	0.17

Note. Standardized coefficient presented. Standard errors in parentheses.
*** $p < 0.001$. ** $p < 0.01$. * $p < 0.05$. † $p < 0.1$.

Figure 2.1

SEM Analysis of Reading Condition, Vocalizations, and Story Comprehension



Note. Solid lines are statistically significant paths, dashed lines are marginally significant paths, and dotted lines are non-significant paths. Covariates include Age, Expressive vocabulary, Prior Agent Usage.
† $p < 0.10$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

CHAPTER 3 COMMUNICATION PATTERNS[1]

Study Abstract

This study examined how an automated social agent can read stories to children via a smart speaker while asking questions and providing contingent feedback. Using a randomized experiment among 90 children aged three to six years, this study compared these children's story comprehension and verbal engagement in storybook reading with a conversational agent versus an adult. The conversational agent's guided conversation was found to be as supportive in improving children's story comprehension as that provided by an adult language partner. At the same time, this study uncovered a number of differences in children's verbal engagement when interacting with a conversational agent versus with an adult. Specifically, children who read with the conversational agent responded to questions with better intelligibility, whereas those who read with an adult responded to questions with higher productivity, lexical diversity, and topical relevance. And the two groups responded to questions with a similar level of accuracy. In addition, questions requiring high cognitive demand amplified the differences in of verbal engagement between the conversational agent and adult partner. The study offers important implications for developing and researching conversational agent systems to support children's language development.

Introduction

Children's development of language skills in preschool years has a profound impact on their later literacy proficiency and overall academic success. Early language skills center on the ability to understand and convey meaning in oral language form (Kim, 2017; Kim, Park, & Wagner, 2014). Extensive research shows that children's development of language skills begins in homes

[1] A version of this chapter was published in *Computers and Education.*

long before children start formal instruction (Fan, Antle, Hoskyn, Neustaedter, & Cramer, 2017; Gest, Freeman, Domitrovich, & Welsh, 2004; Roth, Speece, & Cooper, 2002; Whorrall & Cabell, 2016). **Storybook reading** by family members, typically parents, provides a comfortable environment for stimulating children's language skills. During storybook reading, parents sit together with and read to their children, ideally engaging the children in **guided conversation** where parents serve as children's language partner by posing questions and providing responsive feedback (Golinkoff, Hoff, Rowe, Tamis-LeMonda, & Hirsh-Pasek, 2019; Lever & Sénéchal, 2011). This kind of guided conversation substantially amplifies the learning benefits associated with storybook reading (for a review, see Mol, Bus, de Jong, & Smeets, 2008). However, parents may not always have the language skills, time, or inclination to engage in such conversation-rich storybook reading with their children (Cooter, 2006; Manz, Hughes, Barnabas, Bracaliello, & Ginsburg-Block, 2010; Zevenbergen & Whitehurst, 2003).

In recent years, researchers believe **intelligent systems** with a conversational interface[2] can potentially provide children with additional language learning opportunities, as they have become increasingly powerful and are capable of simulating some interpersonal communications. A growing body of research has developed conversational interfaces that can engage children in a variety of conversations as part of the experiences (see Belpaeme, Kennedy, Ramachandran, Scassellati, & Tanaka, 2018; Kennedy, Baxter, Senft, & Belpaeme, 2016, for review). Some intelligent systems developed in these studies can perform storybook reading tasks adaptable to a child's language level (e.g., Kory & Breazeal, 2014; Kory, Jeong, & Breazeal, 2013); others can employ game-like interactions for vocabulary and language learning

[2] This paper does not consider text-based conversational agents (i.e. "chatbots") (e.g., Hu, Xu, Liu, You, Guo, & Sinha et al., 2018; Xu, Liu, Guo, Sinha, & Akkiraju, 2017), given its focus on young children who cannot read or type.

(e.g., Freed, 2012; Movellan, Eckhardt, Virnes, & Rodriguez, 2009). Studies have demonstrated the feasibility and educational potential of intelligent systems as language partners (Gordon, Spaulding, Westlund, Lee, Plummer, Martinez et al., 2016; Kanero, Geçkin, Oranç, Mamus, Küntay, & Göksun 2018; Kory & Breazeal, 2014).

Most of the intelligent systems in existing studies have an embodied representation as a virtual avatar (Mack, Cummings, Rembert, & Gilbert, 2019; Pauchet, Şerban, Ruinet, Richard, Chanoni, & Barange, 2017) or as a physical robotic body (Belpaeme et al., 2018; Freed, 2012; Movellan et al., 2009; Kory & Breazeal, 2014; Shamekhi, Liao, Wang, Bellamy, & Erickson, 2018). These experimental systems are often designed for narrowly specific scenarios (e.g., a robot to teach food related French vocabularies; Freed, 2012), and thus are rarely adopted by the general public (De Graaf, Ben Allouch, & Van Dijk, 2017; Jacques, Følstad, Gerber, Grudin, Luger, Monroy- Hernández, & Wang, 2019). On the contrary, **conversational agents (CAs)** in a smart speaker form[3], such as Google Home and Amazon Echo, are already used by many families as consumer-oriented voice assistants (Brush, Hazas, & Albrecht, 2018). According to a report, over 150 million households in the U.S. owned smart speakers in early 2020 (Kinsella, 2020). Studies have found that children enjoy their spontaneous interactions with CAs in their homes; children initiated questions (e.g., “Hey Google, does unicorn exist?”; Lovato, Piper, & Wartella, 2019) or commanded CAs to perform small tasks (e.g., “Hey Alexa, play a Christmas song”; Sciuto, Saini, Forlizzi, & Hong, 2018). Despite the popularity of these affordable and versatile smart speakers, little research has been carried out to build CA systems based on smart speakers to support children’s language development. Therefore, the ultimate objective of this

[3] In this paper, “disembodied conversational agents”, “CAs”, and “smart speakers” are used interchangeably.

research is to examine the potential fully automated CAs in the form of a widely-adopted smart speaker that engages children in guided conversation in storybook reading (Blewitt, Rump, Shealy, & Cook, 2009; Chien, 2013; Zhou & Yadav, 2017).

Literature Review

Storybook Reading With Guided Conversations for Children's Language Learning

Storybook reading is an effective way of fostering children's language development in their early years (Bus, 2001; Chang & Huang, 2016; Yen, Y. Chen, Cheng, S. Chen, Y.-Y. Chen, Ni, & Hiniker, 2018). For young children who are not able to decode text independently, storybook reading typically involves them listening to their parents reading out loud a picture book while looking at images. This activity cultivates young children's ability to comprehend oral narratives, thus laying the foundation for understanding the more complex text in higher grade levels. Storybook reading by parents, such as bedtime stories, is a highly routinized activity engaged in by families across cultures (Shanahan & Lonigan, 2010).

In its basic form, storybook reading involves children merely listening to their parents reading the text verbatim (Lenhart, Lenhard, Vaahtoranta, & Suggate, 2018). But this form can be enriched with additional interactive strategies. One effective interactive strategy is to engage children in **guided conversation** (Zevenbergen & Whitehurst, 2003), in which parents ask children prepared questions and provide responsive feedback with a goal of stimulating children's active participation in the reading process. Through back-and-forth conversation, children reflect on and vocally express their understanding of the story. A meta-analysis that reviewed 16 studies has suggested an added value on children's language development resulting from incorporating guided conversation in storybook reading activities (Mol et al., 2008).

Specifically, researchers believe that guided conversation benefits both children's **story comprehension** as well as **verbal engagement** during the storybook reading activity (Vukelich, 1976). In these studies, story comprehension is typically assessed by a battery of questions developed specifically for a story. For example, Lever and Sénéchal found that children who had guided conversation with an experimenter performed significantly better in retelling story elements than those who were not asked any questions during the storybook reading (Lever & Sénéchal, 2011). When analyzing children's verbal engagement in guided conversation, researchers commonly focus on the quality of children's responses to questions asked by parents along one or more of the five aspects, namely language productivity, lexical diversity, topical relevance, accuracy, and intelligibility (Westerveld & Roberts, 2017). A study, for example, suggested that children were more engaged in a guided conversation, as indicated by greater quantity and topical relevance in their responses, if they had higher language proficiency (Westerveld & Roberts, 2017). These prior studies have established useful metrics for evaluating the effectiveness of guided conversation, which guides the development of measures used in this study.

Some studies have investigated the types of questions parents asked in a storybook reading exercise (Birbili & Karagiorgou, 2009). In general, studies suggest that parents should ask questions at different **cognitive demand levels** (Blewitt et al., 2009). Low-cognitive-demand questions typically revolve around a specific story fact, and high-cognitive-demand questions require children to make predictions and inferences based on information that is only implicit in the text. The different cognitive processes required to answer low- and high-demand questions lead to specific patterns in children's verbal engagement (Raphael, 1986). For example, a study found that the children's responses to low-cognitive-demand questions are more concise and simpler than those to a high-cognitive-demand question (Raphael, Highfield & Au, 2006). This differential pattern has suggested that researchers should take questions' cognitive demand level into consideration when designing and evaluating CAs that engage children in story-related dialogues.

In summary, traditional research suggests that an effective reading partner can increase children's language development through engaging children in guided conversation. Yet, this kind of guided conversation is not as common as may have been expected: Parents do not always pause the story, ask questions, and comment on their children's response. This could be due to parents either assuming their child can learn well enough by simply listening to parent reading, or lacking the skills or time to incorporate such interactive opportunities (Golinkoff et al., 2019). The recent development of intelligent systems with voice interfaces that can carry out natural conversation may provide an alternative approach to enrich children's in-home reading experiences.

Embodied Intelligent Systems as Children's Language Learning Partners

Using embodied intelligent systems, both robots and virtual avatars, to enhance children's language learning through a voice interface has been a popular research topic in recent years (Papadopoulos, Lazzarino, Miah, Weaver, Thomas, & Koulouglioti, 2020). A number of initiatives have developed robotic intelligent systems to carry out structured language learning activities. For example, Kory and Breazeal (2014) developed a robotic learning companion for preschool children's oral language development. The robot was designed to tell children stories with different vocabulary complexities and teach children these words. The study found that children learned the vocabulary words that the robot had introduced in their conversation. Michaelis and Mutlu (2017) implemented a robot that was designed to make pre-programmed comments at particular points in a story as a child read aloud. Another group of researchers developed a robotic intelligent agent to support children's French language learning (Freed, 2012). The robot played a food-selection game with children and then talked about that food item with the children in French. The study found that this game-like conversation helped children learn these French words (Freed, 2012). Some other projects have developed intelligent agents embodied in avatars. Allen and colleagues utilized an avatar agent to speak with students in authentic situations, with a goal of improving students' comprehension, pronunciation, and vocabulary in a foreign language (Allen, Divekar, Drozdal, Balagyozyan, Zheng, Song et al., 2019). Authors incorporated an agent in a children's science animation series to teach children scientific vocabularies, and this agent was embodied in the series' main character (Xu & Warschauer, 2020d).

In addition to this prior work contributing to system development, other studies have evaluated the effectiveness of embodied intelligent systems in children's learning context by comparing systems' performance with human learning partners. In terms of learning outcomes, for example, Westlund and colleagues found that children learn unfamiliar words equally well whether with a robot or with a human interlocutor (Westlund et al., 2017). Hong and colleagues also suggested that incorporating a robot teaching assistant in a classroom led to students' similar level of reading and writing improvement as compared to having a human assistant (Hong, Huang, Hsu, & Shen, 2016). In terms of children's verbal engagement with intelligent systems, for example, Hyde and colleagues found that children produced a comparable amount of utterances whether their on-screen conversation was with another human or with an avatar whose speech was operated by an experimenter (Hyde, Kiesler, Hodgins, & Carter, 2014). Tewari and Canny found in their study that children produced even more utterances that were relevant when playing a game with an animal character agent as compared to children playing a game with a familiar human (Tewari & Canny, 2014). They speculated that children's high-level language productions with this particular agent may stem from the more immersive experience of conversing directly with the game's character.

However, these aforementioned embodied systems relied heavily on non-verbal communication (e.g., eye gaze, body orientation) or anthropomorphism features (Tan, Wang, & Sabanovic, 2018) to engage children in learning activities, and these features are not supported by CAs. Despite many studies suggesting that embodied systems' non-verbal cues help establish social relationships with learners and thus positively affect learning (e.g., Gordon et al., 2016; Kennedy et al., 2016), such non-verbal behaviors may also place more cognitive load on the children, which may inhibit children's capacity to process information related to the learning and concentrate on the conversation (Kennedy, Baxter, & Belpaeme, 2015). These two results lead to conflicting hypotheses regarding how the effectiveness of disembodied CAs without non-verbal capacity may compare to that of embodied systems.

Disembodied Conversational Agents (CAs) and Child-CA Interaction

The research on children's interaction with smart speakers has been growing due to these devices' increasing prevalence in many households over the past few years. These studies utilized various methodologies, including parent or child interviews, observations, diary instruments, or in-home audio recordings, and the majority of them focused on unstructured conversations initiated by children with general voice assistant tools (e.g., Amazon Alexa, Google Assistant, Apple Siri). In general, studies found that children commonly either command the voice assistant to perform specific tasks or ask questions to receive answers (Lovato & Piper, 2019). For example, through analyzing audio recordings of children talking with the smart speakers deployed in their home, Beneteau and colleagues categorized children's interactions into three themes, namely entertainment, assistance, and information seeking (Beneteau, Richards, Zhang, Kientz, Yip, & Hiniker, 2020). This categorization scheme was echoed in Lovato's survey study (Lovato & Piper, 2015) and in Garg's study combining interviews with user log data (Garg & Sengupta, 2020a). Another study conducted by Lovato and colleagues specifically focused on children's information seeking behaviors with smart speakers and found that children turned to the CAs for information on a variety of topics, including language, culture, science, and math (Lovato et al., 2019). Although the CA studies reviewed above did not involve educational systems tailored for age-appropriate learning, they still indicated some learning opportunities for children as children initiated conversations with CAs (Garg & Sengupta, 2020b).

In addition to exploring how children readily use CAs, some studies also investigate how children perceive CAs. At least two studies have found that children generally perceive CAs as having cognitive ability; children in both studies indicated that the CAs they interacted with were "smart" and "knowledgeable" (Druga, Williams, Breazeal, & Resnick, 2017; Xu & Warschauer, 2020c). Children in these two studies also perceived CAs as "friendly," truthful," and sociable companions (Druga et al., 2017; Xu & Warschauer, 2020c).

Nevertheless, children were found to sometimes encounter challenges when they interact with CAs. Some children were not aware that smart speakers can not capture or interpret non-verbal expression; thus they attempted to use both verbal and non-verbal communication when responding to the CA. As CAs could not register the non-verbal responses, the conversation flow may have suffered. However, the CAs' reliance on speech may actually be positive, since this reliance—once understood by children—encourages children to practice verbal communication that is vital for their language development (Xu & Warschauer, 2020b). Indeed, two studies found that preschool-aged children made efforts to have their speech understood by CAs; they adjusted sentence structures, modified word choice, or spoke more articulately (Beneteau et al., 2019; Cheng, Yen, Y. Chen, S. Chen, & Hiniker, 2018).

Research on children's interactions with CAs shows that CAs are being seamlessly integrated into children's lives and into the family unit. This favorably positions CAs to be adapted to engage children in focused learning experiences. Indeed, an emerging yet limited body of research develops smart speaker CAs to support specific language learning goals (Smutny & Schreiberova, 2020). Most of these studies leveraged CAs to teach adults foreign languages (Fryer, Ainley, Thompson, Gibson, & Sherlock, 2017; X. L. Pham, T. Pham, Q. M Nguyen, T. H. Nguyen, & Cao, 2018), yet more research is needed to understand how young children respond to learning activities scaffolded by CA partners.

Research Questions

This project focus on the design of a CA that can engage with children in a guided conversation in storybook reading and the evaluation of the effectiveness of this system. The evaluation is guided by two sets of questions that focus on children's comprehension after reading and verbal engagement during reading:

RQ1: Does guided conversation with the CA improve children's story comprehension? If so, how does this improvement compare to that resulting from conversing with a human partner?

RQ2a: Do children's verbal engagement behaviors with a human partner resemble or differ from their behaviors with a CA partner?

RQ2b: Does the similarity or difference in verbal engagement with a CA versus human partner apply to both low- and high-cognitive-demand questions?

Development of the CA Reading Partner

The automated CA system was developed to simulate the dialogue flow of a human conversational partner. The system is built upon Dialogflow open source client library (Cloud, 2020). The CA's natural language understanding module was based on a generic pretrained model built in the Dialogflow engine, then retained with training utterances specific to the CA's conversation context (Lee, 2018; Sabharwal & Agrawal, 2020). These training utterances were collected from a pilot study of what children might say as a response to a particular question prompt. The CA was then able to learn from a small set of training utterances and naturally expand them to many more similar phrases so that the intent of children's verbal responses can be accurately captured and classified.

The CA engages children in a fantasy story *Three Bears in a Boat* authored by David Soman. This story was chosen because of the appropriate level of narrative complexity for the age group and potential story interest. To eliminate the confounding effects of the CA with the effects of voice quality (Cambre & Kulkarni, 2019; O'Neal et al., 2019), the CA used a female recorded voice instead of machine synthetic voice.

Nine open-ended questions were asked throughout the storytelling. Six of these were low-cognitive-demand questions, while the other three were high cognitive-demand questions. For example, the following is a paragraph from the story: "One day, when their mother was out, the three bears did something they really shouldn't have, and with a crash, their mother's beautiful blue seashell lay scattered in pieces across the floor." A low-cognitive-demand question asked, "What did the bears break?" And the answer to that question was "seashell", which was found directly in the text. A high-cognitive-demand question asked children to make an inference based on the given information in the story or to summarize the information (e.g., “How did the bears search for the seashell?”)

The CA performs end-to-end language processing that transcribes children’s voice input into text utterance, classifies the utterance’s intent, and selects a response to that intent. As indicated in Fig. 1, for each of the nine questions, four intent categories were defined to classify a child’s response utterances. These four categories were (1) a set of intents for correct answers, (2) a set of intents for incorrect answers, (3) an intent for when children explicitly express their inability to answer a question, and (4) an intent category for classifying all other intents (e.g., a child does not respond to the question at all or provides an off-topic response). For each intent within each question, there could be many variations of utterances that contained similar semantic meaning. After classifying a child’s responses as belonging to one of the intents, the agent then provided differential feedback that specifically addressed that response.

The CA’s language model was optimized during a three-round field testing involving 20 children. These children’s various responses to the CA’s nine pre-defined questions were collected. For example, the correct answers to the question “What do you think is going to happen with the weather?” describe inclement weather. Possible answers to the question may be “Stormy”, “Bad”, “Windy”, “Rainy”, etc. and thus these intents were created in the initial CA. However, during the pilot run, children were also found to commonly refer to the inclement weather as being scary (e.g., “It’s kind of scary.”; “The bears are afraid of this weather.”; “The bears are too scared and they closed their eyes.”). Thus, “Scary” was added as another intent to capture this group of utterances. This iterative process lasted three rounds, and the agent achieved an inter-rater reliability with a human coder of 0.88, assessed by Cohen's Kappa. A Cohen's Kappa above 0.80 has been considered as excellent agreement (McHugh, 2012).

Method

This section describes the experimental design, measures, and participants of the study.

Experimental Design

This study used a three-condition between-subject experimental design, where participants were randomly assigned to one of the three conditions:

- **"Human-Story"** where children were read a story by a human partner without any guided conversation;
- **"Human-Conversation"** where children were read the same story, plus engaged in guided conversation with a human partner; or
- **"CA-Conversation"** where children were read the same story, and engaged in the same guided conversation with the CA.

In all conditions, children met individually with a trained human experimenter in a designated quiet area at their school. Prior to the experiment session, the participant received an expressive vocabulary assessment (Expressive One Word Picture Vocabulary Test [Martin & Brownell, 2011], see Section 3.2.2 for more detail) as their baseline language proficiency.

At the beginning of the experiment session, the CA or a human experimenter had casual conversation with children about the child's age and favorite colors, following the same protocol. This activity aimed to build rapport between the child and their reading partner (Human or CA). During the storybook reading activity, children in "Human-Story" group were only read the story by a human experimenter without being asked questions, whereas children in "Human-Conversation" and "CA-Conversation" groups were asked a same list of questions and received scripted feedback based on their answers. The smart speaker used in the "CA-Conversation" condition was a Google Home Mini device. There was a human experimenter present in the room in the "CA-Conversation" condition to ensure the child's safety but not to interact with the child.

In all of the three conditions, a physical copy of the storybook was placed in front of the child as that the child could look at the pages as they followed along the narration. Figure 2 shows the experiment session setup of the "Human-Conversation" and the "CA-conversation" condition.

The reading activity took about 20 minutes. Immediately after the reading activity, children's comprehension was assessed using an assessment battery developed by the research team (see Section 3.2.3 for more details). The whole experimental session was video recorded with consent from parents or legal guardians in order to conduct video coding to analyze children's verbal engagement patterns (see Section 3.2.4 for more details).

Experiment Measurements

Background Information

A parent survey was utilized to collect background information on children's date of birth (month and year) and home language (i.e., English only, English as second language, bilingual) . These two factors have been traditionally shown to associate with children's learning and engagement in storybook reading (Cain, Oakhill, & Bryant, 2000; Farnia & Geva, 2013). This survey also asked for information about children's prior experience with CAs, since this factor has been found to influence children's interactions with the CA system (Bartneck, Suzuki, Kanda, & Nomura, 2007). A child was classified as a heavy CA user if parents indicated that the child used CAs more than a few days a week.

Baseline Language Proficiency

Children's baseline oral language skills were measured by the Expressive One Word Picture Vocabulary Test Fourth Edition (EOWPVT-4), which is an experimenter-administered, norm-referenced picture-naming assessment. Each child was asked to name objects, actions, and concepts that were depicted graphically, and the test lasted 15-20 minutes depending on the child's English proficiency. The internal reliability (Cronbach's coefficient alpha) of EOWPVT-4 for 3- to 6-year-olds is 0.95 (Martin & Brownell, 2011). Children's oral language skills are positively associated with children's performances in storybook reading activities (Kendeou, Van den Broek, White, & Lynch, 2009).

Story Comprehension

Children's comprehension level of the story after the storybook reading was measured as an indicator of a proximal learning outcome, similar to the research approach in Zhou and Yadav (2017). A questionnaire was developed, with a total of 10 items to measure how much a child understands the story[4]. Together, these items aim to assess children's ability to 1) memorize main story events and make inferences, 2) sort narrative sequence, and 3) retell part of the story. There were eight items on memorization and inferences. Children were first asked to freely recall the answers. If they could not recall the answer correctly, the researcher provided three multiple-choice options for children to select from. Two points were given to each item that was answered correctly through free recall and one point was given if answered correctly with multiple-choice options. There was one narrative sequence sorting item, where children were asked to place images from the book in the order they occurred in the story. Children earned two points for completely correct order and one point for partially correct order. There was one item to prompt

[4] These 10 questions are different from the nine questions asked during the guided conversation activity.

children to retell a part of the story, where children could earn one point for mentioning each key element in their answer up to four points.

An overall story comprehension score was calculated by summing the number of points across all the items and used this score as a dependent variable for the analysis. The range is from 0 to 22 points (16 points maximum for the 8 memorization and inference-making items, 2 points maximum for the single sequence sorting item, and 4 points maximum for the story-retelling item). Cronbach's coefficient alpha is 0.87 for this story comprehension assessment.

Verbal Engagement

Children's verbal engagement is a measure of how children responded to the CA's questions during storybook readings, which was coded from the video-taped interaction sessions. Only the Human-Conversation and CA-Conversation sessions have this measurement, because the Human-Story condition does not have guided conversation. Five sub-dimensions of verbal engagement were coded, based on the literature on parent-child storybook reading (Westerveld & Roberts, 2017; Vukelich, 1976), namely productivity, lexical diversity, topical relevance, accuracy, and intelligibility. The unit of coding was a child's response to a single prompt, and each child had nine responses.

The reliability of the coding was established using two coders. These two coders, both native English speakers, were undergraduate research assistants. Neither of them were authors of this paper. Coder A coded all of the videos, while Coder B coded a subset of the videos (30%). Coders met once every week to compare codes and discuss any discrepancies in coding. The operationalization and inter-rater reliability (i.e., Inter-class correlation) for each sub-dimension are detailed below.

Productivity. Children's language productivity was captured by the length of utterances in words. The total number of words was counted in each response, including repetitive words. The length of utterances is counted as 0 if the response does not contain verbal expressions. Meaningless speech input (e.g., filter words like Uhhh, Umm, Ahha) was also excluded from the word count. Inter-class correlation = 1.

Lexical Diversity. Children's lexical diversity was captured by the number of unique words in children's responses. The repetitive words in the utterance were removed, and only the unique words were counted. Lexical diversity was coded as 0 if no verbal expression was present, and meaningless speech input was excluded from the word count. Inter-class correlation = 1.

Topical Relevance. The relevance of children's response to a prompt will be coded using three categories, which indicate how well children's responses maintain the semantic flow of conversation. Childish language, imperfect grammar, or answer correctness was not penalized within the relevance code. A response that was directly addressed to the question received a score of 2, a response that was not directly addressed to the question but aligned with the overall theme of the story received a score of 1, and a response that was not related to the question or overall theme received a score of 0. Responses that did not contain verbal expressions was considered as irrelevant and received a score of 0. Inter-class correlation = .94.

Accuracy. The correctness of children's response to a prompt was coded as a dichotomous variable, indicating whether a response is correct or incorrect. Specifically, correct answers received a score of 1 and incorrect answers received a score of 0. Responses that did not contain verbal expressions were considered as incorrect and received a score of 0. Inter-class correlation = 1.

Intelligibility. The intelligibility of children's utterances for each prompt was rated by a 0 to 2 scale, following the method proposed by Flipsen (FlipsenJr, 2002). A score of 0 indicated that a child's utterance was largely unintelligible, and the coders could understand less than 50% of the utterance; a score of 1 indicated that a child's utterance was mostly intelligible except for one or two words; a score of 2 indicated a child's utterance was articulate, and the coder could understand every single word. Responses that did not contain verbal expressions were excluded from this coding. Inter-class correlation = .87.

In the analyses of this paper, these five sub-dimensions were analyzed separately, with each of them being a dependent variable.

Results

This section first presents the full sample descriptive statistics of the outcomes measures. Findings were then reported regarding the CA's effects on story comprehension (RQ1),verbal engagement behaviors with CA versus with human partner (RQ2a), and the interaction effects of questions' cognitive demand on verbal behaviors (RQ2b).

Descriptive Statistics of Outcome Measures for Full Sample

The descriptive statistics for the full sample are presented in Table 2. Children's average score in story comprehension was 11.2, indicating that these children in the sample correctly answered half of the comprehension items correctly. In terms of children's verbal engagement as they engaged in guided conversation, the average length of utterance (i.e., productivity) was 4.4 words and the average number of unique words (i.e., lexical diversity) was 3.7 words. The average score of topical relevance was 1.5 out of 2, indicating that most children were able to generate answers that addressed the questions. Children in this study on average responded to half of the in-story questions accurately, evidenced by an accuracy rate of 0.5. Also, these children generally articulated their answers with good intelligibility, resulting in an intelligibility score of 1.9 out of 2.

The Pearson correlation coefficients between study variables and their significance levels are displayed in Table 2. Children's story comprehension was significantly positively correlated with all verbal engagement measures. Among the verbal engagement variables, productivity, diversity, relevance, and accuracy were significantly correlated with each other, while intelligibility was only significantly correlated with relevance and accuracy but not productivity or diversity.

The Effect of CAs on Story Comprehension

The first research question examined the extent to which having guided conversation with a learning partner during storybook reading may enhance children's story comprehension and whether the benefits of guided conversation differed depending on the nature of the learning partner (i.e., a CA or a human partner).

Descriptively, children who had guided conversation with either a CA or human language partner correctly answered story comprehension questions more frequently than did children in the group without guided conversation (see Table 3). When comparing the performance of the two groups with guided conversation, children in two groups answered approximately the same number of items correctly. The difference in score was only 0.13, which was much smaller than 1 (i.e., 1 item).

The regression analyses first compared whether the two groups of children who had guided conversation performed better in story comprehension than their counterparts who did not engage in guided conversation (i.e., Human-Story group). This was considered as a baseline analysis to validate the benefits of guided conversation with low- and high-cognitive-demand questions, regardless of the nature of language partners who carried out the conversation. The "Human-Story" group was used as the reference group in our regression models (Model 1 in Table 4). The results indicated that both the "CA-Conversation" (β = 0.44, p = 0.03) and the "Human-Conversation" groups (β = 0.64, p = 0.01) 537 scored significantly higher than the "Human-Story" group. The higher comprehension score achieved by the two groups with guided conversation confirmed the advantage of incorporating guided conversation in storybook reading.

A post-hoc analysis was then conducted to compare the comprehension scores between children who had guided conversation with the CA and those who had guided conversation with a human partner, by using "Human-Conversation" as the reference group in the original regression model. The result indicated that the comprehension scores of children in the "CA-Conversation" group were not statistically different from those of children in the "Human-Conversation" group (β = -0.20, p = 0.29). These results suggested that the guided conversation carried out by CA could yield similarly effective learning as a human partner.

The Effect of CAs on Verbal Engagement in Guided Conversation

The second set of research questions (RQ2a and RQ2b) focused on children's verbal engagement behaviors in guided conversation. RQ2a examined whether children conversing with CA partners exhibited similar or different verbal engagement patterns as they would when talking with a human partner, and RQ2b examined whether any difference in verbal engagement varied based on the question's cognitive demand level. Table 3 presents descriptive statistics of children's verbal engagement between CA-Conversation and Human-Conversation conditions, as well as disaggregates the between-condition verbal engagement measures by low- and high-cognitive-demand questions. Descriptively, children produced longer, more lexically diverse, and more relevant responses when conversing with a human partner, yet children conversing with a CA language partner responded more intelligibly. The accuracy rate between the two conditions resembled each other. When the questions' cognitive demand level is taken into consideration, the difference in productivity and lexical diversity between the CA and Human groups became larger for high-cognitive-demand questions.

Multilevel linear analyses were employed to formally test whether children conversing with a CA exhibited similar verbal engagement patterns as they would when talking with a human partner, and whether any difference in verbal engagement varied based on the question's cognitive demand. For each of the five engagement metrics, the analyses first focused on the effects of language partner and questions' cognitive demand (Models 1, 3, 5, 7, and 9 in Table 5). The analyses then focused on examining the interaction effects between the nature of language partner and questions' cognitive demand, by including an additional cross-level interaction term between the nature of language partner and questions' cognitive demand (i.e., CA-Conv × High cog; Models 2, 4, 6, 8, and 10 in Table 5). All models were controlled for children's age, expressive vocabulary score, home language, and prior CA experiences, given the documented relations between these variables and children's verbal responses.

Productivity

The multilevel model analysis suggested a significant effect of the nature of learning partners on language productivity. Specifically, reading with a human partner resulted in children responding in greater length ($\beta = -0.28$, $p = 0.04$) than they did to the CA partner (Model 1 in Table 5), suggesting some benefits of a human language partner in promoting language productivity over CAs. Questions that require high cognitive demand elicited responses that were significantly longer ($\beta = 0.78$, $p = 0.00$) than low-demand questions (Model 1 in Table 5).

Furthermore, the cross-level interaction between the nature of language partner and the questions' cognitive demand was significant (Model 2 in Table 5). The difference in language productivity between "CA-Conversation" and "Human-Conversation" conditions was significantly amplified when children were answering high-cognitive-demand questions (β = -0.44, p = 0.002; Figure 3-A). Specifically, children's response length did not differ significantly for questions that required low cognitive demand (β = -0.14, p = 0.36). However, human partners' advantages in eliciting more language production became prominent when the conversation was cognitively challenging.

Lexical Diversity

Regarding the effect of a CA versus human partner on lexical diversity (Model 3 in Table 5), children's responses contained more unique words when conversing with a human partner than with the CA (β = -0.35, p = 0.01), indicating some advantages of a human partner in encouraging more lexically diverse utterances. In terms of the effect of questions' cognitive demand levels (Model 3 in Table 5), children were found to respond to high-cognitive-demand questions using more unique words (β = 0.83, p = 0.00).

Furthermore, the model with the interaction effect (Model 4 in Table 5) indicated that the difference in lexical diversity between "CA-Conversation" and "Human-Conversation" was significantly larger among questions that require high cognitive demand (β = -0.35, p = 0.01; Figure 3-B). Conversing with a human partner did not elicit more lexically diverse responses from children than did the CA if the questions required low cognitive demand (β = -0.24, p = 0.11). Yet among the questions that were cognitively challenging, human partners were more likely to invite responses with higher lexical diversity than were CAs.

Topical Relevance

In terms of the effect of a CA versus human partner on topical relevance (Model 5 in Table 5), children's responses were more topically relevant when they conversed with a human partner than with a CA partner (β = -0.30, p = 0.04). However, when focusing on the effect of questions' cognitive demand (Model 5 in Table 5), a child's ability to generate topically relevant answers did not significantly differ by low- and high-cognitive-demand questions (β = -0.04, p = 0.60).

The cross-level interaction between questions' cognitive demand and learning condition was not significant, as indicated in Model 6 in Table 5 (β = -0.01, p = 0.95). Specifically, this non-significant interaction effect suggested that human learning partners, in general, were more likely to elicit relevant responses from children, and this pattern was consistent for both low- and high-cognitive-demand questions (Figure 3-C).

Accuracy

In terms of the effect of a CA versus human partner on response accuracy (Model 7 in Table 5), there was no significant differences in response accuracy between children conversing with the CA and those conversing with a human partner (β = -0.08, p = 0.38). In terms of the effect of questions' cognitive demand level (Model 7 in Table 5), unsurprisingly, children were more likely to answer lower cognitive demand questions accurately compared to higher cognitive demand questions (β = -0.28, p = 0.00). In terms of the interaction effect between questions' cognitive demand and the nature of language partner (Model 8 in Table 5), our finding indicated that, when answering low-cognitive-demand and high-cognitive-demand questions, children had a comparable level of accuracy in the "CA-Conversation" and "Human-Conversation" conditions (β = -0.04, p = 0.79; Figure 3 -D).

Intelligibility

In terms of the effects of a CA versus human partner on intelligibility of children's responses (Model 9 in Table 5), children's responses appeared to be more intelligible when conversing with a CA than with a human partner ($\beta = 0.23$, $p = 0.04$). This suggested CAs' advantages in encouraging children to articulate their utterances. Questions' cognitive demand level did not significantly influence the intelligibility of children's responses ($\beta = -0.14$, $p = 0.10$; Model 9 in Table 5). When including the interaction effect (Model 10 in Table 5), children showed a similar level of intelligibility when conversing with the CA or a human partner, regardless of whether they answered low- or high-cognitive-demand questions ($\beta = 0.07$, $p = 0.67$; Figure 3-E).

Robustness Check

We ran a robustness check excluding the 5 children who had indicated they had already read the story. All results remained consistent with the findings reported above.

Discussion

This section discusses the interpretation of the findings with regard to why children can learn from a CA partner and demonstrated certain verbal engagement behaviors in the guided conversations. As one of the first studies to design and evaluate a CA learning partner, the findings of this study provide novel design implications for further improving CAs in an affordable smart speaker format and deploying such systems to enrich young children's everyday literacy learning.

The Learning Benefits of Conversing With Disembodied CAs

This study demonstrated that the children who had guided conversation, whether with the CA or a human, comprehended the story better than the group who did not engage in guided conversation. This result was not surprising given that the vast education literature documenting the added value of guided conversation over non-interactive storybook reading where parents merely read the text (for reviews, see Arnold & Whitehurst, 1994; Mol et al., 2008). However, the study's findings also extend this traditional line of research by demonstrating that a CA can potentially facilitate children's language learning as effectively as a human partner in this study's context. The positive effects of guided conversation in this study also validated the CA dialogue design strategy that incorporates high- and low-cognitive-demand questions (Raphael, 1986), proving the usefulness of developing intelligent systems grounded in education theories (Callaghan & Reich, 2018). One important factor enabling the CA in this study to replicate these benefits is its capacity to respond to children adaptively based on the children's answers. This kind of adaptive response helped identify and clarify children's misconceptions and reinforce an accurate understanding of the story (Aksan, Kochanska, & Ortmann, 2006; Funamoto & Rinaldi, 2015).

The fact that the CA, with only a voice interface, can benefit children's story comprehension as much as face-to-face human partners can reinforces the importance of verbal dialogue in promoting children's language skills. Yet, this result should not be interpreted as undermining the role non-verbal cues ordinarily play in boosting learning effectiveness (Dunn, Rodriguez, Miller, Gerhardt, Vannatta, Saylor et al., 2011; Negi, 2009; Kahlbaugh & Haviland, 1994). Instead, this result arises out of the storybook reading scenario in general. During storybook reading, young children typically look at a picture book while listening to the story. Therefore, children's visual channel primarily concentrates on the illustrations (Paivio, 1991), which substantially facilitates their understanding of the narration (Takacs & Bus, 2018), thus leaving limited room for processing other non-verbal information provided by a human partner (Hanson, 1989). The minimal non-verbal information children do receive from human partners during storybook reading may not be sufficient to translate into the short-term learning benefits this study assessed, such as immediate recall of story elements. Yet it is plausible that non-verbal cues may influence how children verbally engage with their reading partner, which is discussed below.

Verbal Engagement Behaviors With Disembodied CAs

The findings of this study revealed nuances in how children verbally respond differently to a natural human and to a CA. Specifically, children were found to generate longer and more lexically diverse responses when conversing with a human partner than with a CA. The human partner's ability to leverage social cues (e.g., looking at children as children formulate responses [Guo & Feng, 2013]) could contribute to this difference. Moreover, children were found to provide more relevant responses to a human partner. One speculation is that the social presence of a human partner may have encouraged children to make an effort to maintain the conversation flow (Groom, 2008; Kim, Berkovits, Bernier, Leyzberg, Shic, Paul, & Scassellati 2013; Zhou, Mark, Li, & Yang, 2019). Yet interestingly, despite the differences in response relevance, children answered questions from the CA and the human partner with a similar level of accuracy, corroborating the finding on post-storybook reading comprehension that the CA benefits children's learning as well as the human does. Taken together, this suggests that the lower relevance of children's responses to CAs' questions was not due to cognitive factors but may be related to social or behavioral factors. Lastly, CAs were found to enhance children's intelligibility. This may be due to children's perceptions of the CAs' listening ability: Studies have suggested that people are likely to talk more clearly and slowly if they perceive their partners as needing additional support in interpreting their utterances (Rooy, 2009). This pattern was also identified in children's communication with CAs: Young children adjusted their speech style if they perceived CAs as encountering difficulties in understanding them (Beneteau et al., 2019; Cheng et al., 2018).

There was evidence that the cognitive demand required to participate in the conversation, on top of the nature of the language partner, jointly shapes children's verbal engagement. Specifically, high-cognitive-demand questions amplified the effects of human partners on eliciting longer and more lexically diverse responses from children. This may also be attributed to the social presence of a human partner discussed before. An adult figure is often perceived as more authoritative by children, which may encourage children to devote greater cognitive effort in attempting challenging questions (Davis, 2003).

Despite the nuances discussed above, the descriptive statistics suggest that children's responses to the CA were not fundamentally different from their responses to a human partner. Children in both conditions replied to the prompting with multi-word responses, kept their responses quite relevant to the question, and uttered their responses intelligibly. This implies that children, regardless of whether they are conversing with a CA or a human partner, follow a shared convention during the conversation. This finding resonates with the prominent "Computers as Social Actors (CASA)" paradigm (Nass, Steuer, & Tauber, 1994; Nass, Moon, & Morkes, 1997), which suggests that human users, especially children, tend to treat intelligent systems as human beings. Numerous studies on children's interactions with embodied intelligent systems (e.g., robots and avatars) have supported this paradigm (Admoni & Scassellati, 2017; Fink, Lemaignan, Dillenbourg, Rétornaz, Vaussard, Berthoud et al., 2014; Heerink, Diaz, Albo-Canals, Angulo, Barco, Casacuberta, & Garriga, 2012; Melson, Kahn, Friedman, Roberts, Garrett, & Gill, 2009; Spolaôr & Benitti, 2017; Tewari & Canny, 2014). The current study thus extends the application of the CASA paradigm to include disembodied CAs that do not have anthropomorphic figures and are restricted to verbal communication. This extension is also supported by other theories that suggest an intelligent system's verbal ability is a central factor that shapes how users judge the system's sociability and intelligibility and thus how they interact with that system (Araujo, 2018).

Designing Better CAs for Early Childhood Language Development

The current study sheds light on three implications for future design of CA language partners for young children.

First, it is important to design CAs with a clear theoretical rationale for meeting children's unique learning needs. In this study, the CA was tailored to a storybook reading context, incorporating evidence-based strategies that take into consideration the cognitive demand required by the conversation. The CA's ability to improve children's learning confirms the importance of a theoretically-driven design approach. Unfortunately, according to a recent review of over 500 voice-based apps on the market (Xu & Warschauer, 2020a), many of the apps purporting to benefit language learning were not grounded in research, thus limiting these apps' abilities to fulfill their intended educational goal.

Second, it is important to fully leverage a CA's conversation capacity to compensate for its inability to utilize non-verbal expressions. In the current study, the CA did not fully simulate a human partner in eliciting children's elaborate, complex, and relevant responses. As discussed before, it is possible that the CA's disadvantages may arise from its disembodiment which prevents it from leveraging non-verbal cues. Developers can compensate for this lack by improving on such CAs' conversational expressiveness (Lin, Ginns, Wang, & Zhang, 2020). For example, CAs may be designed to clearly explain to children how to best answer a question (e.g., "Listen to the question carefully and try to say as much as you can!") or provide follow-up prompts to encourage longer or more appropriate responses (e.g., “Great job! Can you say some more?” or “That is a good idea! But how about what we’ve just talked about?”). CAs may also leverage natural acoustic features (i.e., tone, prosody, speech speed), such as asking a question with a tone of genuine curiosity, which may entice children to more thoroughly express their thoughts.

Third, it is also important for developers to recognize the unique properties of CAs that make them especially useful learning tools regardless of whether they precisely mimic a human partner. Some researchers have proposed that CAs may be particularly valuable for providing children with opportunities to practice their language skills since CAs require children to communicate verbally (Vaquero, Saz, Lleida, Marcos, & Canalís, 2006). This supports the claim of Clark and colleagues (Clark, Munteanu, Wade, Cowan, Pantidi, Cooney et al., 2019), who suggest that developers need not and should not attempt to develop CAs that exactly emulate human-to-human interactions. Instead, CAs should be envisioned as a new form of language partner, one that could complement and enrich children’s everyday conversational experiences.

Future Directions

These initial studies shed light on future directions. First, this dissertation only focused on children’s immediate outcomes—comprehension of the story they have just listened to as well as perceptions of the conversational agent’s animacy. While my findings provide important evidence regarding children’s positive learning effects as well as perceptions of conversational agents, it is unclear whether interacting with a conversational agent reading partner may lead to long-run benefits to children’s language development and the establishment of enduring social bonding between children and the agent. Future studies may want to include a delayed post-test or carry out a more intense intervention to examine long-term effects of interactions with conversational agents.

Second, this dissertation was conducted in an experimental manner where the human partner's conversational behaviors were scripted. This design increased the internal validity of the findings, reducing the confounding effects resulting from the variations in ways of human partners actually carrying out the guided conversation. Nevertheless, in a naturalistic setting, parents or other adults may unitize guided conversation in varying degrees and with varying approaches. Future studies may compare children's learning and engagement with conversational agents with those with their care givers or teachers who routinely read with them. In addition, in this study, children had brief casual conversation with the conversational agent, which might be viewed as a training opportunity that familiarized children with the scheme of interacting with the conversational agent. Future studies may want to further explore how to best design such warm-up interactions to better support children who were either reluctant to participate in the conversation or encountered obstacles (e.g., relied on non-verbal expressions) during their conversation with the CA.

Conclusion

This study demonstrated the potential of smart speaker CAs in carrying out guided conversation in storybook reading activities to nurture children's language development. Given that smart speakers are already accessible in many homes, the endeavor to augment smart speakers' usefulness as learning tools may have profound impact on children and on the market. Encouragingly, this study demonstrated that the CA developed in this study based on education literature was equally supportive as a human partner in enhancing children's story comprehension. However, nuanced patterns in children's verbal engagement were also identified:

CAs and human partners have their own advantages in some respects. Among the first experimental studies comparing children's learning and engagement with CAs versus adults, this study provides initial evidence on the potential of smart speakers as effective reading partners. Nevertheless, the precious parent-child interactions can not be replaced by artificially intelligent systems; yet CAs may supplement parents' current practices and thus enrich children's early literacy experiences. Understanding how children learn from and engage with CAs is an important step in gaining a more complete picture of the role that intelligent systems play in children's educational landscape in today's world.

References

Admoni, H., & Scassellati, B. (2017). Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction, 6*(1), 25–63. https://doi.org/10.5898/jhri.6.1.admoni

Aksan, N., Kochanska, G., & Ortmann, M. R. (2006). Mutually responsive orientation between parents and their young children: Toward methodological advances in the science of relationships. *Developmental Psychology*, *42*(5), 833–848. https://doi.org/10.1037/0012-1649.42.5.833

Allen, D., Divekar, R. R., Drozdal, J., Balagyozyan, L., Zheng, S., Song, Z., Zou, H., Tyler, J., Mou, X., Zhao, R., Zhou, H., Yue, J., Kephart, J. O., & Su, H. (2019). The rensselaer Mandarin project — A cognitive and immersive language learning environment. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 9845–9846. https://doi.org/10.1609/aaai.v33i01.33019845

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, *85*, 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Arnold, D. H., & Whitehurst, G. J. (1994). *Accelerating language development through picture book reading: A summary of dialogic reading and its effect*.

Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & Society*, *21*(1–2), 217–230. https://doi.org/10.1007/s00146-006-0052-7

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21), eaat5954. https://doi.org/10.1126/scirobotics.aat5954

Beneteau, E., Boone, A., Wu, Y., Kientz, J. A., Yip, J., & Hiniker, A. (2020). Parenting with Alexa: Exploring the introduction of smart speakers on family dynamics. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376344

Beneteau, E., Richards, O. K., Zhang, M., Kientz, J. A., Yip, J., & Hiniker, A. (2019). Communication Breakdowns Between Families and Alexa. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–13. https://doi.org/10.1145/3290605.3300473

Birbili, M., & Karagiorgou, I. (2009). Helping children and their parents ask better questions: An intervention study. *Journal of Research in Childhood Education*, *24*(1), 18–31. https://doi.org/10.1080/02568540903439359

Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology*, *101*(2), 294–304. https://doi.org/10.1037/a0013844

Brush, A. J., Hazas, M., & Albrecht, J. (2018). Smart homes: Undeniable reality or always just around the corner? *IEEE Pervasive Computing*, *17*(1), 82–86. https://doi.org/10.1109/mprv.2018.011591065

Bus, A. G. (2001). Joint caregiver-child storybook reading: A route to literacy development. *Handbook of Early Literacy Research*, 171–191.

Cain, K., Oakhill, J., & Bryant, P. (2000). Investigating the causes of reading comprehension failure: The comprehension-age match design. *Reading and Writing*, *12*(1–2), 31–40.

Callaghan, M. N., & Reich, S. M. (2018). Are educational preschool apps designed to teach?: An analysis of the app market. *Learning, Media and Technology*, *43*(3), 280–293. https://doi.org/10.1080/17439884.2018.1498355

Cambre, J., & Kulkarni, C. (2019). One voice fits all?: Social implications and research challenges of designing voices for smart devices. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–19. https://doi.org/10.1145/3359325

Change, C.-J., & Huang, C.-C. (2016). Mother–child talk during joint book reading in two social classes in Taiwan: Interaction strategies and information types. *Applied Psycholinguistics*, *37*(2), 387–410. https://doi.org/10.1017/s0142716415000041

Cheng, Y., Yen, K., Chen, Y., Chen, S., & Hiniker, A. (2018). Why doesn't it work?: Voice-driven interfaces and young children's communication repair strategies. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 337–348. https://doi.org/10.1145/3202185.3202749

Chien, C.-W. (2013). Using Raphael's QARs as differentiated instruction with picture books. *English Teaching Forum*, *51*, 20–27. ERIC.

Clark, L., Munteanu, C., Wade, V., Cowan, B. R., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., & Murad, C. (2019). What makes a good conversation?: Challenges in designing truly conversational agents. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–12. https://doi.org/10.1145/3290605.3300705

Cloud, G. (2020). *Dialogueflow documentation*. Technical Report.

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. *Academic Press*.

Cooter, K. S. (2006). When mama can't read: Counteracting intergenerational illiteracy. *The Reading Teacher*, *59*(7), 698–702. https://doi.org/10.1598/rt.59.7.9

Davis, H. A. (2003). Conceptualizing the role and influence of student-teacher relationships on children's social and cognitive development. *Educational Psychologist*, *38*(4), 207–234. https://doi.org/10.1207/s15326985ep3804_2

de Graaf, M., Ben Allouch, S., & van Dijk, J. (2017). Why do they refuse to use my robot?: Reasons for non-use derived from a long-term home study. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 224–233. https://doi.org/10.1145/2909824.3020236

Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). "Hey Google is it ok if I eat you?": Initial explorations in child-agent interaction. *Proceedings of the 2017 Conference on Interaction Design and Children*, 595–600. https://doi.org/10.1145/3078072.3084330

Dunn, M. J., Rodriguez, E. M., Miller, K. S., Gerhardt, C. A., Vannatta, K., Saylor, M., Scheule, C. M., & Compas, B. E. (2011). Direct observation of mother-child communication in pediatric cancer: Assessment of verbal and non-verbal behavior and emotion. *Journal of Pediatric Psychology*, *36*(5), 565–575. https://doi.org/10.1093/jpepsy/jsq062

Fan, M., Antle, A. N., Hoskyn, M., Neustaedter, C., & Cramer, E. S. (2017). Why tangibility matters: A design case study of at-risk children learning to read and spell. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1805–1816. https://doi.org/10.1145/3025453.3026048

Farnia, F., & Geva, E. (2013). Growth and predictors of change in English language learners' reading comprehension. *Journal of Research in Reading*, *36*, 389–421. https://doi.org/10.1111/jrir.12003

Fink, J., Lemaignan, S., Dillenbourg, P., Rétornaz, P., Vaussard, F., Berthoud, A., Mondada, F., Wille, F., & Franinović, K. (2014). Which robot behavior can motivate children to tidy up their toys?: Design and evaluation of "ranger." *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction - HRI '14*, 439–446. https://doi.org/10.1145/2559636.2559659

FlipsenJr, P. (2002). Longitudinal changes in articulation rate and phonetic phrase length in children with speech delay. *Journal of Speech, Language, and Hearing Research*, *45*(1), 100–110. https://doi.org/10.1044/1092-4388(2002/008)

Freed, N. A. (2012). *"This is the fluffy robot that only speaks French": Language use between preschoolers, their families, and a social robot while sharing virtual toys* [Ph.D. thesis].

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior*, *75*, 461–468. https://doi.org/10.1016/j.chb.2017.05.045

Funamoto, A., & Rinaldi, C. M. (2015). Measuring parent-child mutuality: A review of current observational coding systems. *Infant Mental Health Journal*, *36*(1), 3–11. https://doi.org/10.1002/imhj.21481

Garg, R., & Sengupta, S. (2020a). He is just like me: A study of the long-term use of smart speakers by parents and children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(1), 1–24. https://doi.org/10.1145/3381002

Garg, R., & Sengupta, S. (2020b). Conversational technologies for in-home learning: Using co-design to understand children's and parents' perspectives. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376631

Gest, S. D., Freeman, N. R., Domitrovich, C. E., & Welsh, J. A. (2004). Shared book reading and children's language comprehension skills: The moderating role of parental discipline practices. *Early Childhood Research Quarterly*, *19*(2), 319–336. https://doi.org/10.1016/j.ecresq.2004.04.007

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development*, *90*(3), 985–992. https://doi.org/10.1111/cdev.13128

Gordon, G., Spaulding, S., Westlund, J., Lee, J., Plummer, L., Martinez, M., Das, M., & Breazeal, C. (2016). Affective personalization of a social robot tutor for children's second language skills. *Thirtieth AAAI Conference on Artificial Intelligence*.

Groom, V. (2008). What's the best role for a robot. *Proceedings of the Fifth International Conference on Informatics in Control, Automation and Robotics Service*, 323–328. Automation and Robotics (ICINCO). https://doi.org/10.5220/0001507103230328

Guo, J., & Feng, G. (2013). How eye gaze feedback changes parent-child joint attention in shared storybook reading? *Eye Gaze in Intelligent User Interfaces*, 9–21. https://doi.org/10.1007/978-1-4471-4784-8_2

Hanson, L. (1989). Multichannel learning research applied to principles of television production: A review and synthesis of the literature. *Educational Technology*, *29*, 15–19.

Heerink, M., Diaz, M., Albo-Canals, J., Angulo, C., Barco, A., Casacuberta, J., & Garriga, C. (2012). A field study with primary school children on perception of social presence and interactive behavior with a pet robot. *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, 1045–1050. https://doi.org/10.1109/roman.2012.6343887

Hong, Z., Huang, Y., Hsu, M., & Shen, W. (2016). Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL classrooms. *Journal of Educational Technology & Society*, *19*(1), 337–349.

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., & Akkiraju, R. (2018). Touch your heart: A tone-aware chatbot for customer care on social media. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–12. https://doi.org/10.1145/3173574.3173989

Hyde, J., Kiesler, S., Hodgins, J. K., & Carter, E. J. (2014). Conversing with children: Cartoon and video people elicit similar conversational behaviors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1787–1796. https://doi.org/10.1145/2556288.2557280

Jacques, R., Følstad, A., Gerber, E., Grudin, J., Luger, E., Monroy-Hernández, A., & Wang, D. (2019). Conversational agents: Acting on the wave of research and development. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–8. https://doi.org/10.1145/3290607.3299034

Kahlbaugh, P. E., & Haviland, J. M. (1994). Nonverbal communication between parents and adolescents: A study of approach and avoidance behaviors. *Journal of Nonverbal Behavior*, *18*(1), 91–113. https://doi.org/10.1007/bf02169080

Kanero, J., Geçkin, V., Oranç, C., Mamus, E., Küntay, A. C., & Göksun, T. (2018). Social robots for early language learning: Current evidence and future directions. *Child Development Perspectives*, *12*(3), 146–151. https://doi.org/10.1111/cdep.12277

Kendeou, P., Van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, *101*(4), 765–778. https://doi.org/10.1037/a0015956

Kennedy, J., Baxter, P., & Belpaeme, T. (2015). The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*, 67–74. https://doi.org/10.1145/2696454.2696457

Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (2016). Social robot tutoring for child second language learning. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 231–238.

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., & Scassellati, B. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders*, *43*(5), 1038–1049. https://doi.org/10.1007/s10803-012-1645-2

Kim, Y.-S. G. (2017). Why the simple view of reading is not simplistic: Unpacking component skills of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading*, *21*(4), 310–333. https://doi.org/10.1080/10888438.2017.1291643

Kim, Y.-S., Park, C. H., & Wagner, R. K. (2014). Is oral/text reading fluency a “bridge” to reading comprehension? *Reading and Writing*, *27*(1), 79–99. https://doi.org/10.1007/s11145-013-9434-7

Kinsella, B. (2020, April 28). *Nearly 90 million U.S. adults have smart speakers, adoption now exceeds one-third of consumers*. Voicebot.Ai.

Kory, J., & Breazeal, C. (2014). Storytelling with robots: Learning companions for preschool children’s language development. *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 643–648. https://doi.org/10.1109/roman.2014.6926325

Kory, J. M., Jeong, S., & Breazeal, C. L. (2013). Robotic learning companions for early language development. *Proceedings of the 15th ACM on International Conference on Multimodal Interaction - ICMI ’13*, 71–72. https://doi.org/10.1145/2522848.2531750

Lee, H. (2018). *Voice user interface projects: Build voice-enabled applications using Dialogflow for Google Home and Alexa Skills Kit for Amazon Echo*. Packt Publishing.

Lenhart, J., Lenhard, W., Vaahtoranta, E., & Suggate, S. (2018). Incidental vocabulary acquisition from listening to stories: A comparison between read-aloud and free storytelling approaches. *Educational Psychology*, *38*(5), 596–616. https://doi.org/10.1080/01443410.2017.1363377

Lever, R., & Sénéchal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners’ oral narrative construction. *Journal of Experimental Child Psychology*, *108*(1), 1–24. https://doi.org/10.1016/j.jecp.2010.07.002

Lin, L., Ginns, P., Wang, T., & Zhang, P. (2020). Using a pedagogical agent to deliver conversational style instruction: What benefits can you obtain?. *Computers & Education*, 143, 103658. https://doi.org/10.1016/j.compedu.2019.103658

Lovato, S., & Piper, A. M. (2015). "Siri, is this you?": Understanding young children's interactions with voice input systems. *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 335–338. https://doi.org/10.1145/2771839.2771910

Lovato, S. B., & Piper, A. M. (2019). Young children and voice search: What we know from human-computer interaction research. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00008

Lovato, S. B., Piper, A. M., & Wartella, E. A. (2019). Hey Google, do unicorns exist?: Conversational agents as a path to answers to children's questions. *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 301–313. https://doi.org/10.1145/3311927.3323150

Mack, N. A., Cummings, R., Rembert, D. G. M., & Gilbert, J. E. (2019). Co-Designing an intelligent conversational history tutor with children. *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. https://doi.org/10.1145/3311927.3325336

Manz, P. H., Hughes, C., Barnabas, E., Bracaliello, C., & Ginsburg-Block, M. (2010). A descriptive review and meta-analysis of family-based emergent literacy interventions: To what extent is the research applicable to low-income, ethnic-minority or linguistically-diverse young children? *Early Childhood Research Quarterly*, *25*(4), 409–431. https://doi.org/10.1016/j.ecresq.2010.03.002

Martin, N. A., & Brownell, R. (2011). *Expressive one-word picture vocabulary test-4 (EOWPVT-4)*. Academic Therapy Publications.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia medica: Biochemia medica, 22(3), 276-282.

Melson, G. F., Kahn, P. H., Beck, A., Friedman, B., Roberts, T., Garrett, E., & Gill, B. T. (2009). Children's behavior toward and understanding of robotic and living dogs. *Journal of Applied Developmental Psychology*, *30*(2), 92–102. https://doi.org/10.1016/j.appdev.2008.10.011

Michaelis, J. E., & Mutlu, B. (2017). Someone to read with: Design of and experiences with an in-home learning companion robot for reading. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 301–312. https://doi.org/10.1145/3025453.3025499

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development*, *19*(1), 7–26. https://doi.org/10.1080/10409280701838603

Movellan, J., Eckhardt, M., Virnes, M., & Rodriguez, A. (2009). Sociable robot improves toddler vocabulary skills. *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction - HRI '09*. https://doi.org/10.1145/1514095.1514189

Nass, C. I., Moon, Y., & Morkes, J. (1997). Computers are social actors: A review of current. *Human Values and the Design of Computer Technology*, *72*, 137.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*, 72–78. https://doi.org/10.1145/191666.191703

Negi, J. S. (2009). The role of teachers' non-verbal communication in ELT classroom. *Journal of NELTA*, 101–110. https://doi.org/10.3126/nelta.v14i1.3096

O'Neal, A. L. (2019). *Is Google Duplex too human?: Exploring user perceptions of opaque conversational agents* [Ph.D. thesis].

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 45*(3), 255–287. https://doi.org/10.1037/h0084295

Papadopoulos, I., Lazzarino, R., Miah, S., Weaver, T., Thomas, B., & Koulouglioti, C. (2020). A systematic review of socially assistive robots in pre-tertiary education. *Computers & Education*, 103924. https://doi.org/10.1016/j.compedu.2020.103924

Pauchet, A., Şerban, O., Ruinet, M., Richard, A., Chanoni, É., & Barange, M. (2017). Interactive narration with a child: Avatar versus human in video-conference. *Intelligent Virtual Agents,* 343–346. https://doi.org/10.1007/978-3-319-67401-8_44

Pham, X. L., Pham, T., Nguyen, Q. M., Nguyen, T. H., & Cao, T. T. H. (2018). Chatbot as an intelligent personal assistant for mobile language learning. *Proceedings of the 2018 2nd International Conference on Education and E-Learning - ICEEL 2018*, 16–21. https://doi.org/10.1145/3291078.3291115

Raphael, T. E. (1986). Teaching question answer relationships, revisited. *Reading Teacher, 39*(6), 516–522.

Raphael, T. E., Highfield, K., & Au, K. H. (2006). *Question-Answer relationships*. Scholastic.

Rooy, S. C.-V. (2009). Intelligibility and perceptions of English proficiency. *World Englishes, 28*(1), 15–34. https://doi.org/10.1111/j.1467-971x.2008.01567.x

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *The Journal of Educational Research*, *95*(5), 259–272. https://doi.org/10.1080/00220670209596600

Sabharwal, N., & Agrawal, A. (2020). *Cognitive Virtual Assistants Using Google Dialogflow* (pp. 13–54). Apress. https://doi.org/10.1007/978-1-4842-5741-8

Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). “Hey Alexa, what’s up?”: A mixed-methods studies of in-home conversational agent usage. *Proceedings of the 2018 on Designing Interactive Systems Conference*, 857–868. https://doi.org/10.1145/3196709.3196772

Shamekhi, A., Liao, Q. V., Wang, D., Bellamy, R. K. E., & Erickson, T. (2018). Face value? Exploring the effects of embodiment for a group facilitation agent. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18*, 1–13. https://doi.org/10.1145/3173574.3173965

Shanahan, T., & Lonigan, C. J. (2010). The national early literacy panel: A summary of the process and the report. *Educational Researcher, 39*(4), 279–285. https://doi.org/10.3102/0013189x10369172

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. Computers & Education, 103862. https://doi.org/10.1016/j.compedu.2020.103862

Spolaôr, N., & Benitti, F. B. V. (2017). Robotics applications grounded in learning theories on tertiary education: A systematic review. *Computers & Education*, *112*, 97–107. https://doi.org/10.1016/j.compedu.2017.05.001

Takacs, Z. K., & Bus, A. G. (2018). How pictures in picture storybooks support young children's story comprehension: An eye-tracking experiment. *Journal of Experimental Child Psychology*, *174*, 1–12. https://doi.org/10.1016/j.jecp.2018.04.013

Tan, H., Wang, D., & Sabanovic, S. (2018). Projecting life onto robots: The effects of cultural factors and design type on multi-level evaluations of robot anthropomorphism. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 129–136. https://doi.org/10.1109/roman.2018.8525584

Tewari, A., & Canny, J. (2014). What did spot hide?: A question-answering game for preschool children. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1807–1816. https://doi.org/10.1145/2556288.2557205

Vaquero, C., Saz, O., Lleida, E., Marcos, J., & Canalís, C. (2006). Vocaliza: An application for computer-aided speech therapy in Spanish language. *IV Jornadas En Tecnologia Del Habla*, 321–326.

Vukelich, C. (1976). The development of listening comprehension through storytime. *Language Arts*, *53*(8), 889–891.

Westerveld, M. F., & Roberts, J. M. A. (2017). The oral narrative comprehension and production abilities of verbal preschoolers on the autism spectrum. *Language, Speech, and Hearing Services in Schools*, *48*(4), 260–272. https://doi.org/10.1044/2017_lshss-17-0003

Whorrall, J., & Cabell, S. Q. (2016). Supporting children's oral language development in the preschool classroom. *Early Childhood Education Journal*, *44*(4), 335–341. https://doi.org/10.1007/s10643-015-0719-0

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. https://doi.org/10.1145/3025453.3025496

Xu, Y., & Warschauer, M. (2020a). A content analysis of voice-based apps on the market for early literacy development. In Proceedings of the 19th ACM International Conference on Interaction Design and Children (IDC '20). Association for Computing Machinery, New York, NY, USA.

Xu, Y., & Warschauer, M. (2020b). Exploring young children's engagement in joint reading with a conversational agent. In *Proceedings of the 19th ACM International Conference on Interaction Design and Children* (IDC '20). Association for Computing Machinery, New York, NY, USA. https://doi.org/ 10.1145/3392063.3394417

Xu, Y., & Warschauer, M. (2020c). What are you talking to?: Understanding children's perceptions of conversational agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. April 25-30, 2020, Honolulu, HI. ACM. https://doi.org/10.1145/3313831.3376416

Xu, Y., & Warschauer, M. (2020d). "Elinor is talking to me on the screen!" Integrating conversational agents into children's television programming. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. April 25-30, 2020, Honolulu, HI. ACM. https://doi.org/10.1145/3334480.3383000

Yen, K., Chen, Y., Cheng, Y., Chen, S., Chen, Y.-Y., Ni, Y., & Hiniker, A. (2018). Joint media engagement between parents and preschoolers in the U.S., China, and Taiwan. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), 1–19. https://doi.org/10.1145/3274461

Zevenbergen, A., & Whitehurst, G. (2003). Dialogic reading: A shared picture book reading intervention for preschoolers. *On Reading Books to Children: Parents and Teachers*, 177–200.

Zhou, M. X., Mark, G., Li, J., & Yang, H. (2019). Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems*, *9*(2–3), 1–36. https://doi.org/10.1145/3232077

Zhou, N., & Yadav, A. (2017). Effects of multimedia story reading and questioning on preschoolers' vocabulary learning, story comprehension and reading engagement. *Educational Technology Research and Development*, *65*(6), 1523–1545. https://doi.org/10.1007/s11423-017-9533-2

Table 3.1

Participant Background Information by Experimental Condition

	Full Sample	Human-Story	Human-Conversation	CA-Conversation
Age in months	58.1 (9.33)	56.9 (9.5)	58.4 (9.1)	59.5 (8.8)
Home language				
English only	67.4%	60.0%	71.0%	69.7%
Bilingual	10.1%	16.0%	9.7%	6.1%
ESL	22.5%	24.0%	19.3%	24.2%
Heavy CA use	36.2%	33.3%	34.2%	37.9%
EOWPVT	68.2 (17.5)	68.8 (17.1)	66.7 (17.1)	70.6 (17.4)
N	90	26	31	33

Note. Standard deviations in parentheses.

Table 3.2

Intercorrelations Among Study Variables

	Comp.	Pro.	Div.	Rel.	Acc.	Int.	Mean	SD	Range
Comprehension	1	0.28**	0.35**	0.59***	0.66***	0.32**	11.17	5.23	(0, 22.00)
Productivity		1	0.96***	0.52***	0.48***	0.03	4.38	2.55	(0, 13.78)
Diversity			1	0.55***	0.49***	0.02	3.73	1.89	(0, 11.33)
Relevance				1	0.87***	0.27*	1.47	0.53	(0, 2)
Accuracy					1	0.34*	0.54	2.19	(0, 1)
Intelligibility						1	1.90	1.65	(0, 2)

Note. Pearson correlation coefficients and significance levels reported.
p <.05 denoted as *, p <.01 denoted as **, p <.001 denoted as ***

Table 3.3

Descriptive Statistics of Story Comprehension and Verbal Engagement Variables by Experimental Condition

		Human-Story		Human-Conversation		CA-Conversation	
Comprehension		9.62	(4.62)	11.86	(5.78)	11.73	(5.03)
Productivity	Low cog.			3.24	(2.71)	3.10	(1.99)
	High cog.			7.84	(4.67)	5.81	(4.05)
	Combined			4.77	(2.53)	4.00	(2.56)
Diversity	Low cog.			2.89	(2.57)	2.62	(1.44)
	High cog.			6.35	(3.32)	4.89	(2.57)
	Combined			4.11	(1.99)	3.38	(1.75)
Relevance	Low cog.			1.56	(0.44)	1.40	(0.59)
	High cog.			1.53	(0.63)	1.36	(0.68)
	Combined			1.55	(0.46)	1.39	(0.59)
Accuracy	Low cog.			0.58	(0.25)	0.59	(0.28)
	High cog.			0.46	(0.28)	0.44	(0.29)
	Combined			0.52	(0.27)	0.51	(0.30)
Intelligibility	Low cog.			1.89	(0.17)	1.94	(0.12)
	High cog.			1.82	(0.30)	1.89	(0.27)
	Combined			1.87	(0.17)	1.93	(0.15)

Note. Standard deviations in parentheses. Verbal engagement measures not applicable in Human-Story condition.

Table 3.4

Linear Regression Model on Story Comprehension Measures

	Comprehension
Human-Conv	**0.64****
	(0.22)
CA-Conv	**0.44****
	(0.22)
EOWPVT	**0.62*****
	(0.11)
Age	**0.27***
	(0.12)
English only	-0.07
	(0.30)
ESL	0.23
	(0.31)
Heavy CA use	0.02
	(0.19)
R^2	0.62

Note. “Human-Story” is the reference group. Coefficients are standardized. Standard error in parentheses.
p <.05 denoted as *, *p* <.01 denoted as **, *p* <.001 denoted as ***. Significant coefficients bolded.

Table 3.5

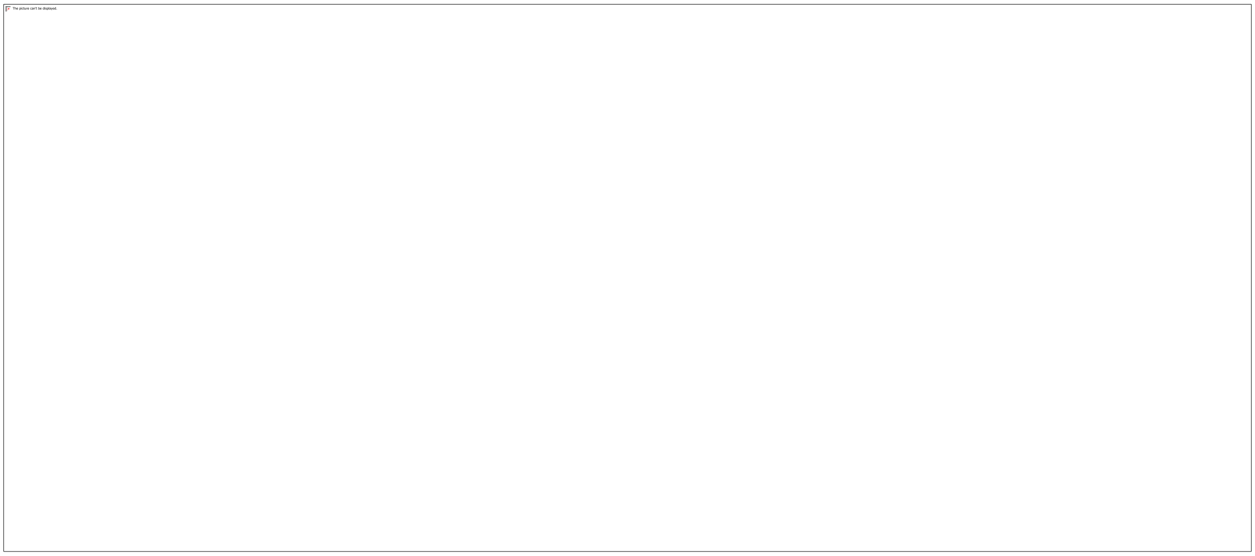
Multilevel Linear Models on Verbal Engagement Measures

	Productivity		Diversity		Relevance		Accuracy		Intelligibility	
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
CA-Conv	**-0.28***	-0.14	**-0.35***	-0.24	**-0.30***	**-0.30***	-0.08	-0.07	**0.23***	0.20
	(0.14)	(0.15)	(0.14)	(0.15)	(0.15)	(0.15)	(0.11)	(0.12)	(0.11)	(0.13)
High cog	**0.78*****	**0.58*****	**0.83*****	**0.67*****	-0.04	-0.04	**-0.28*****	**-0.30*****	-0.14	-0.11
	(0.07)	(0.1)	(0.07)	(0.1)	(0.07)	(0.1)	(0.08)	(0.11)	(0.09)	(0.12)
CA-Conv × High cog		**-0.44****		**-0.35***		-0.01		-0.04		0.07
		(0.14)		(0.14)		(0.15)		(0.16)		(0.17)
EOWPVT	**0.21***	**0.21***	**0.26***	**0.26***	**0.24***	**0.24***	**0.30*****	**0.30*****	0.13	0.13
	(0.11)	(0.11)	(0.10)	(0.10)	(0.11)	(0.11)	(0.08)	(0.08)	(0.08)	(0.08)
Age	0.05	0.05	0.04	0.04	0.19	0.19	0.10	0.10	0.02	0.02
	(0.10)	(0.10)	(0.10)	(0.10)	(0.11)	(0.11)	(0.08)	(0.08)	(0.08)	(0.08)
English only	-0.09	-0.09	-0.16	-0.16	0.13	0.13	0.03	0.03	0.16	0.16
	(0.31)	(0.31)	(0.30)	(0.30)	(0.32)	(0.32)	(0.23)	(0.23)	(0.26)	(0.26)
ESL	0.20	0.20	0.08	0.08	0.35	0.35	0.23	0.23	0.20	0.16
	(0.32)	(0.32)	(0.32)	(0.32)	(0.33)	(0.33)	(0.24)	(0.24)	(0.27)	(0.26)
Heavy CA use	-0.09	-0.09	-0.05	-0.05	0.02	0.02	-0.10	-0.10	-0.01	-0.01
	(0.17)	(0.17)	(0.16)	(0.16)	(0.17)	(0.17)	(0.13)	(0.13)	(0.13)	(0.13)

Note. Each row is an independent variable, and each column is a dependent variable. For each dependent variable, two models were performed: the latter one has an interaction effect (CA-Conv × High cog). EOWPVT, Age, English only, ESL, and Heavy CA use are 5 control variables. Coefficients are standardized coefficient. Standard error in parentheses.
p <.05 denoted as * is considered statistically significant, p <.01 denoted as **, p <.001 denoted as ***. Significant coefficients bolded.

Figure 3.1

Dialogue Flow Design of the CA's Guided Conversation Module



Note. CA pauses at particular points of the story, asks a question, classifies the child's response, and selects a feedback response for the child. We zoom in to the four intent categories only for Question 1. In total there are nine questions.

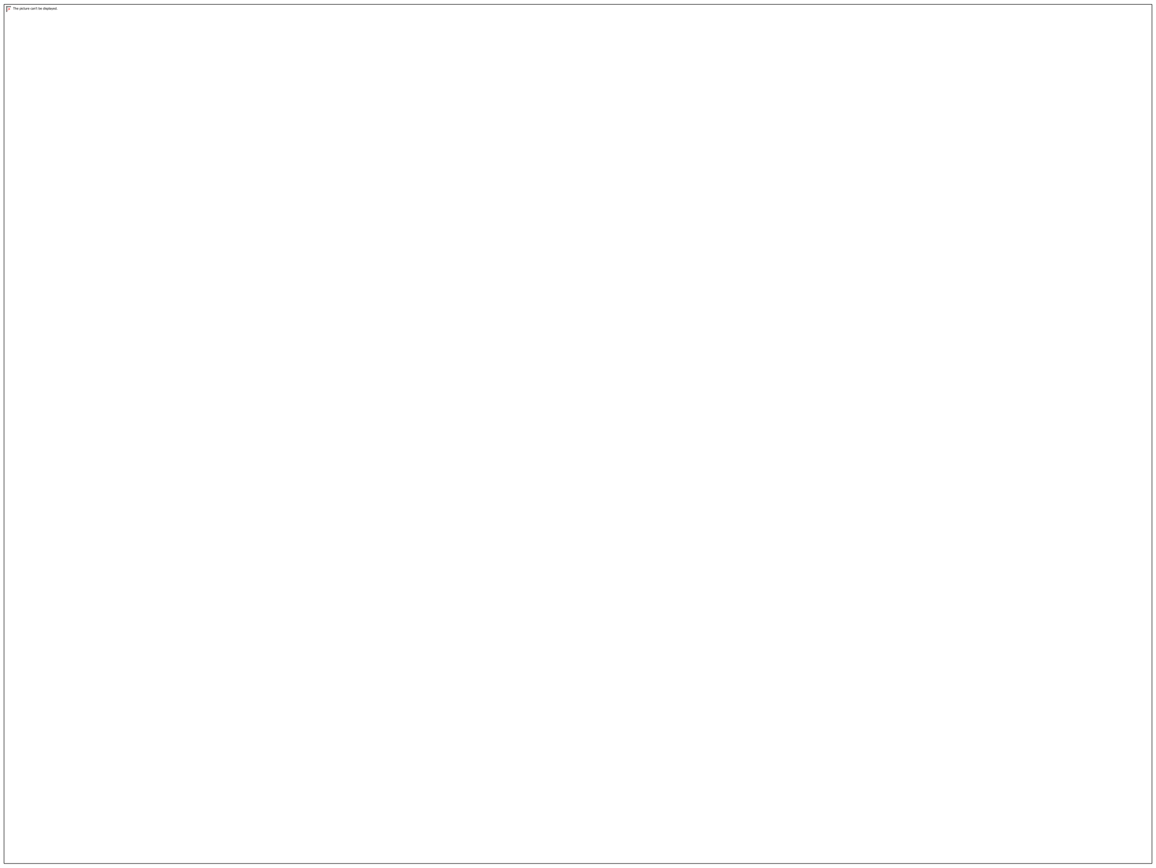
Figure 3.2

Experiment Session Setup

Note. A child participant in the Human-Conversation group (left); and another child participant in the CA-Conversation group, the smart speaker system is highlighted (right)

Figure 3.3

Verbal Engagement by the Nature of Language Partner and Questions' Cognitive Demand Levels



CHAPTER 4 VERBAL AND NON-VERBAL INTERACTIONS[5]

Study Abstract

Joint book reading is a highly routinized activity that is nearly universal among families. Conversational agents (CAs) can potentially act as joint-reading partners by engaging children in story-related, scaffolded conversations. In this project, we examine children's interactions with a CA reading partner. We identify patterns in children's language production, flow maintenance, and affect when responding to the CA. We then lay out a set of affordances and challenges for developing CAs as conversation partners. We propose that, rather than attempting to develop CAs as an exact replicate of human conversational partners, we should treat child-agent interaction as a new genre of conversation and calibrate CAs based on children's actual communicative practices and needs.

Introduction

Joint book reading is a highly routinized activity engaged in by families across cultures. Joint reading provides a focused and interactive literacy environment, which is believed to boost children's language development and long-run academic success. One key ingredient to such benefits is the meaningful conversation between the child and parent during joint reading. In recent years, the rapid development of artificial intelligence (AI) has made conversational agents (CAs) more capable of simulating natural interpersonal interactions [49]. Studies suggest that children respond to CAs socially and treat CAs as companions or guides [46, 60, 62]. Children's social reactions to CAs raise the question: Can CAs serve as suitable language partners for children in joint reading activities, complementing the role of parents or other mentors?

---

[5] A version of this chapter was published in *the Proceedings of the 19th ACM International Conference on Interaction Design and Children*

In this project, we developed a smart-speaker CA narrating a picture book while engaging children in story-related, scaffolded conversations in order to facilitate comprehension and engagement. We then conducted an observational study of 33 children's individual interactions with the. In our analysis, we approached conversation as an interaction between two parties (i.e., the child and the CA) and focused on the children's responses to the CA in three dimensions that are traditionally identified as revealing engagement levels in conversations [7, 26, 31, 53, 63]. These three dimensions are language production that captures the quantity of children's vocalization, flow maintenance that details the semantic and temporal appropriateness of children's responses, and affect that indicates children's emotional engagement during the conversation. We seek to answer the following question: How do children respond to a CA reading partner during conversation, in terms of children's language production, flow maintenance, and affect? We also note significant developmental differences within children aged 3 to 6 years, and thus further ask: Do the younger children within this age group (3- to 4-year-olds) respond to the CA reading partner differently than do older children (5- to 6-year-olds)?

CA Dialogue Flow

Our CA reading partner itself contains no visual element but is designed to be used alongside a printed picture book. This combination increases children's print exposure and potentially enhances their engagement and learning.

Figure 1 displays the general workflow of the child-CA communication. Children were first invited to respond to an open-ended question (Initial Prompt hereafter) and received feedback for providing an answer that the CA could interpret. If the response could not be understood by the CA, the CA would ask children a scaffolded follow-up question (Follow-up

Prompt hereafter) and then would give feedback based on the child's response. If the CA could not understand a child's response to the Follow-up Prompt, the CA would give the child vague, generic feedback that explained the question but did not directly address the child's answer.

Method

Participants

Thirty-three children aged 3 to 6 years were recruited from childcare centers in a university community. The mean age of the participants was 4.5 years, and 19 of them (58%) were girls. Twenty-three children (70%) spoke only English at home. According to parent reports, 30% of the children had never interacted with a CA, 27% had done so monthly, 12% had done so weekly, and 30% had daily interaction with a CA. We divided these children into two groups based on their age. The younger group (3- and 4-year-olds) consisted of 16 children, and the older group (5- and 6-year-olds) consisted of 17 children.

Coding Framework

The development of a coding framework was guided by prior research that collectively emphasizes the verbal and non-verbal aspects of conversations (Brennan & others, 2005; Heldner & Edlund, 2010; Kang et al., 2009; Ruusuvuori, 2013; Wanska et al., 1989). The resultant coding framework consists of three dimensions, namely **language production**, **flow maintenance**, and **affect**. These three dimensions are believed to work in conjunction to signify the extent to which a speaker is engaged in meaningful and productive communication (Sánchez et al., 2006). Prior research has included some or all of these dimensions to analyze children's communication with voice interfaces such as robots and other CAs (Beneteau et al., 2019; Robins et al., 2004; Sidner et al., 2004). Below, we will detail how each of the dimensions was informed by prior work, and how the coding was operationalized.

The first coding dimension was **language production**, which captures a child's production of verbal responses to the CA's prompts. As suggested by Brennan & others (2005), active verbal responses are generally prerequisite for fluid conversation. Thus, we coded whether a child verbally responded to the prompt. In addition, studies suggested that the word length of a response is one of the most important indicators of conversation engagement (Kang et al., 2009). We therefore also coded the total number of words in each of the children's responses.

The second dimension was **flow maintenance**, which focuses on the semantic flow and temporal flow of the conversation. According to Wanska et al. (1989), to maintain the semantic flow, a speaker needs to respond to his partner in a topically relevant way. This indicates that a speaker is monitoring the content of his partner's statement and making an effort to link his own response to his partner's (Wanska et al., 1989). According to Heldner & Edlund (2010), to maintain temporal flow, a speaker's timing of responses should follow a turn-taking pattern without any overlapping speech or any silence between turns (i.e., no-overlap-no-gap). We therefore coded the topic relevance and timing of children's responses. For example, in response to the question "*What shape is the island the bears need to look for,*" a relevant answer would be a shape (e.g., triangle) or some recognizable object (e.g., hat, crown). Responses that were not considered relevant included those that did not reference some shape or did not stay within the broader theme of the story. The timing of response included two codes: whether a child responded too quickly (before the CA came to a full stop) and whether a child responded with a substantial delay (after approximately 2 seconds when CA believed the child was giving up their turn).

The third dimension was **affect**, which focuses on children's varied emotional responses throughout the conversation. According to Ruusuvuori (2013), a speaker's emotional engagement in a conversation (both when speaking and being spoken to) can be revealed through several affective markers, including laugh tokens, lexical choices, tones of voice, and facial expressions. We examined these markers during children's responses to the CA and when they were listening to the CA's feedback and then categorized children's affective state as belonging to one of four categories: positive, negative, confused, and neutral. These four categories are believed to be salient affective states as children engage in learning processes (Halliday et al., 2018). Positive emotion was identified through the presence of any of the following: positive facial expressions, positive body cues, presence of laugh, rising tone, or positive connotations. Negative emotion was identified through the presence of any of the following: negative facial expressions, negative body cues, falling tone, or negative connotations (Aviezer et al., 2012). Confusion was identified through any facial expressions (e.g., eyebrow raise-arched, side mouth stretch) or verbal expressions (e.g., "Umm?" "Why?") that indicated confusion (Rozin & Cohen, 2003). Neutral emotion was coded when no significant signs of emotion were present (Leppänen & Hietanen, 2004).

Coding Procedure

Our primary data sources were the video-taped interaction sessions and their transcriptions. The unit of analysis is a child's response to a single prompt. If a child successfully answered an Initial Prompt, they would not receive a Follow-up Prompt for that same question. In total, we analyzed 330 responses to Initial Prompts and 205 responses to Follow-up Prompts, thus resulting in a total of 535 coding fragments. For each coding fragment, we coded the three dimensions of communication and included detailed notes for each dimension. This process generated both quantitative and qualitative coding data, enabling statistical analyses accompanied by contextual evidence.

We established the reliability of the coding using two coders who were informed with the overall objective of the study to examine children's engagement with a CA reading partner. Coder A coded and took notes on all of the videos, while Coder B coded a subset of the videos (30%) to establish the inter-rater reliability. Coders met once every week to compare codes and discuss any discrepancies in coding. The inter-rater reliability (Cohen's Kappa for categorical codes and Interclass correlation for numeric codes) between Coder A and Coder B for each item is between 0.88 and 1. To establish reliability of the qualitative coding, Coder B reviewed the notes initially taken by Coder A and discussed any disagreements or necessary clarifications. This process was repeated until both coders agreed that the notes accurately reflected the actual interactions.

Results

In this section, we first detail the CA's performance in order to demonstrate the CA's accuracy as a language partner. We then answer our first research question by presenting statistics from the quantitative coding along with descriptive notes contextualizing the statistics. In addition, we answer our second research question using an ANOVA analysis for numeric coding data (i.e., response length in words) and Chi-square analyses for the rest of the coding items with categorical data to determine whether a significant difference exists between the younger children and the older children along the three coding dimensions.

CA's Performance

The performance of the CA was determined by how successfully the CA could categorize children's responses into predefined intent categories. There were three possible outcomes: accurate categorization, inaccurate categorization, or categorization failed. "Accurate categorization" indicates that the CA was able to categorize a child's response to a pre-defined intent, and this categorization was accurate. “Inaccurate categorization” indicates that the CA was also able to categorize a response, but this categorization was inaccurate. “Categorization failed” indicates that the CA was not able to categorize a child’s response as any of the pre-defined intent categories.

As displayed in Table 1, the majority of the responses to Initial and Follow-up Prompts were accurately categorized by the CA, with 76.7% and 83.7% accuracy, respectively. Inaccurate categorization occurred very rarely for Initial or Follow-up Prompts, with 0.3% and 0.7% inaccuracy. All instances of inaccurate categorization were due to the CA’s inaccurate speech-to-text translation. Another 22.7% of responses to Initial Prompts and 15.6% of responses to Follow-up Prompts were identified as “categorization failed.” There were three reasons for categorization failure. First, children’s verbal responses were absent or incomplete. For example, children nodded their head to indicate “yes,” shook their head for “no,” or shrugged their shoulders for “I don’t know.” Children sometimes provided a verbal response that could only be understood when combined with non-verbal expressions. For example, saying “This one,” and pointing to the picture at the same time. Second, a child’s response was not anticipated. For example, a child answered “dinosaur” to the question “What shape is the island the bears need to look for?”, with dinosaur outside of the overall theme of the story and only brought up by this single child. Not surprisingly, Follow-up Prompts resulted in a higher rate of intent detection than Initial Prompts, largely due to the more restricted questions that eliminated the likelihood of a child providing unanticipated answers. Third, the voice response was translated incorrectly to text. One example for this case was the CA mis-registering a child’s correct answer of “shell” to “sound,” thus leading to an out-of-context response.

The CA appeared to perform better with the older group of children. Young children's utterances being categorized with a lower success rate may primarily be due to young children's less articulate pronunciation and higher likelihood of providing unanticipated answers.

Language Production

Presence of Verbal Expressions

Children actively responded to the CA with verbal expressions: they verbally responded to over 85% of the CA's prompts (see Table 2). The response rate for Follow-up Prompts (89.3%) was higher than that for Initial Prompts (86.2%), probably due to the scaffolded nature of the Follow-up Prompts. We also found that older children were more likely to verbally respond to the prompts.

When children did not respond verbally to a prompt, they almost always instead relied on non-verbal expressions. Nonverbal responses were quite common when children did not know the answer (e.g., shrugging, shaking head). Children also sometimes gestured to convey information (i.e., pointing to an image in the book). Since the CA was not able to understand such responses, they triggered the CA's programmed scaffolding mechanism. Only a few of children's failures to respond verbally were due to the child's disengagement or intentional avoidance. For example, one child became distracted and looked at the ceiling, missing the question altogether. Another child appeared to realize that he would receive a multiple-choice question if he did not answer the Initial Prompts. After attempting two questions, he stopped responding to any of the Initial Prompts and instead waited for the CA to give him scaffolded questions, all of which he answered correctly.

Response Length

The average length of responses to Initial Prompts was 4 words, and the average for Follow-up Prompts was 2 words (see Table 2). Initial Prompts generally solicited longer responses that tended to be complete sentences or phrases. For example, when asked the question “What do the bears ride on to travel across the sea?” most children’s responses included a verb or a preposition; rather than simply replying “sailboat,” children replied “ride on a sailboat” or “on a sailboat”. Follow-up Prompts tended to result in children giving shorter responses, and many children tended to give single-word responses. For example, when the question on the bears’ transportation was rephrased as “Do the bears ride on a sailboat or do they swim across the sea?” children tended to respond by simply saying “sailboat” or “swim,” rather than “on a sailboat” or “swim across the sea.” Among both Initial and Follow-up Prompts, older children and younger children generated responses of comparable length.

Flow Maintenance

Topic Relevance

In our observation, children were able to directly answer the majority of questions (see Table 3). Children were much more likely to generate relevant responses to Follow-up Prompts (89.6%) than to Initial Prompts (76.7%). The increase of topic relevance among Follow-up Prompts suggests that our scaffolding mechanisms worked well to support children’s communication. For example, when asked “Where did the bears find the blue seashell?” one child provided an answer that was topically irrelevant to the story (an answer about a dinosaur). The CA then asked, “Did they find it on an island or did they find it under the sea?” The child responded appropriately, and the conversation flow was maintained.

When looking at the topic relevance by age group, we found that, unsurprisingly, older children were better able to directly answer the Initial Prompts, which were open-ended questions. However, with the scaffolding prompts, the age difference in topic-relevance became non-significant. The topic relevance of younger children's responses increased by 23 percent (from 65% to 88%). For older children, scaffolded prompts only slightly increased the already high proportion of relevant responses by 5 percent (from 87% to 92%).

Timing of Response

Gaps and Pauses. Children needed time to organize their thoughts when answering a question, and this resulted in children sometimes not initiating a response within a short period of time or beginning a response but pausing to think before completing it. Because CAs must rely solely on the duration of gaps and pauses to determine when a child's turn is either abandoned or completed, CAs would simply miss the utterances spoken after the gaps or pauses. Overall, gaps and pauses were observed among responses to 21.6% of Initial Prompts, and this number was 11.5% for Follow-up Prompts (see Table 4). The occurrence of gaps and pauses was higher among the Initial Prompts than among the Follow-up Prompts with multiple choices, probably because Initial Prompts were generally more challenging for children. As expected, younger children had significantly more gaps and pauses in responding to the open-ended Initial Prompts than older children. Younger children were observed to have gaps and pauses among 29.7% of Initial Prompts while older children had gaps and pauses among 14.8% of Initial Prompts. This difference was probably due to younger children's less advanced reading comprehension. With the Follow-up Prompt, the frequency of gaps and pauses became more similar between younger and older children (13.3% for the younger group and 9.5% for the older group).

Rushed Responses. Children sometimes responded to a question too quickly, before the CA fully completed its turn. This kind of rushed response was observed among 8% of Initial Prompts but among 24% of Follow-up Prompts (see Table 4). The commonality of rushed responses to Follow-up Prompts may be due to the questions' lower difficulty or wording that contained the original open-ended question and a set of possible answers in question form. For example, a Follow-up Prompt asked "What did the bears break? Did they break a blue seashell or did they break a honey jar?" and one child responded "blue seashell" immediately after the first part of the question, given that the questioning tone invited the child to respond. The CA did not register the child's answer and continued to complete the scaffolded question. The child then replied "yes" immediately after the CA mentioned the blue seashell, but the CA also did not hear this response. Younger children appeared to have more trouble determining when the CA had completed its question and could register the child's response for the Follow-up Prompts that contained questioning tone in the middle of them. Specifically, younger children's rushed responses occurred among 31.6% of Follow-up Prompts while older children's only occurred among 15.5% of Follow-up Prompts.

Affect

Affect While Responding

As shown in Table 5, most of the time, children expressed no emotion at all when responding to the CA. Children showed neutral affect among 74.7% of Initial Prompts and 85.6% of Follow-up Prompts. These neutral affective states were categorized by a lack of facial expressions and body gestures, a flat tone of voice, and matter-of-fact word choices. This lack of affect may be due to the design of our CA's prompts: these prompts primarily asked about specific content in the story, thus leaving little room for children's emotional expression.

Nevertheless, positive emotional responses were not uncommon, which was observed among 25.3% of Initial Prompts and 14.4% of Follow-up Prompts. Children sometimes exhibited pride in having given what they were confident was a correct response. For example, when a child was recalling a set of places where the bears had searched on an island, she nodded her head and clapped her hands with every additional place she recalled. Another child smiled and nodded to herself once she finished answering, as if she was satisfied with her own answer. Positive emotions were also observed during some "distracted moments" when children's comments expressed excitement over something tangentially related to a detail in the book but not directly related to the prompt. For example, when asked, "Why do you think the bears stopped at this island?" a child commented excitedly while pointing at a Ferris wheel on that page, "This island has a Ferris wheel! I saw a Ferris wheel! I have been on a Ferris wheel." Although these conversational moments may, on the one hand, indicate a child's disengagement with the story, the child sharing a related personal experience with the CA may, on the other hand, suggest that the CA's prompt did provide children an opportunity to express their enthusiasm. We observed less neutrality and more positivity among Initial Prompts, which may be due to that Initial Prompts were less restricted, thus allowing children to include their feelings and attitudes.

Young children appeared to be more likely to show positive affect as they responded to the Initial Prompts than older children. When answering Initial Prompts, younger children had positive affect among 31.1% of their responses while older children showed positive emotion among 21.8% of their responses. This may be due to that younger children would have a greater sense of accomplishment for answering a question that seemed to be challenging for them. It may also be related to younger children's tendency to insert information that interests them in the conversation. As expected, when it comes to the Follow-up Prompts that were generally easier and more restricted, the age difference between positive (or neutral) affect diminished.

We did not observe any negative affect or confusion during children's responses to the CA's prompts, either to Initial or Follow-up Prompts. However, one caveat to interpret the absence of negative affect and confusion was that we only examined the affect while children were actually responding to the CA. It is possible that some children who were not responding were confused by a particular question or did not feel like answering the question.

Affective Reaction to Feedback

We analyzed children's reactions to CA feedback resulting from two types of CA intent categorization (i.e., accurate and failed categorization) described in the "CA's Performance" section above. The CA's accurate categorization of a child's response resulted in feedback that was appropriate and specific, while the CA's failure to categorize a child's response resulted in feedback that was vague and generic. Instances of inaccurate categorization were not examined here since they occurred very rarely in our study.

We first looked at children's reactions to specific feedback to their correct and incorrect responses (see Table 6 for correct responses and Table 7 for incorrect responses). When children received specific feedback that indicated their answer was correct, they typically expressed positive emotion (e.g., laughing, cheering, clapping, dancing, saying "Yay!"). This positive emotion was observed among 75.2% of Initial Prompts and 73.4% of Follow-up Prompts. However, children's affect was less impacted when they received specific feedback for an incorrect response, as neutral affect was observed among 82.4% of Initial Prompts and 83.4% of Follow-up Prompts. Negative emotion only occurred among 15 percent of feedback that indicated a child answered a prompt incorrectly (15.9% of Initial Prompts and 15.3% of Follow-up Prompts). In a few cases, children also showed confusion after receiving feedback for their incorrect answers (1.7% of Initial Prompts and 1.3% for Follow-up Prompts).

Interestingly, compared to older children, younger children's emotion was more likely to be enhanced when they received positive feedback from the CA for their correct answer. However, younger and older children reacted in a similar way when they received feedback indicating their answer was incorrect.

We then looked at children's reactions to the CA's vague and generic feedback (i.e., feedback that did not address children's response at all and that instead simply told them the answer to a question) resulting from failed categorizations (see Table 8). We rarely observed any emotional changes when children received this kind of feedback (i.e., neutral affect, 94% for Initial Prompts and 90% for Follow-up Prompts), regardless of whether children actually answered the question correctly or incorrectly. Occasionally, we observed confusion among children (6.5% for Initial Prompts and 9.9% for Follow-up Prompts), especially those who appeared confident about their response. For example, one child correctly answered that the bears were "hugging each other," but the CA mistranslated "hugging" as "hacking." The CA then replied, "The bears were hugging each other to make themselves feel better," and the child commented, "Why? Why didn't it say I'm right?" This pattern was similar across age groups.

Discussion

In this paper, we describe both the design of a CA that can engage children in joint reading and a user study of how children interacted with this CA. Based on our findings, we now turn to a discussion of how automated conversational interfaces could play the role of language partners for young children and how to best design such interfaces.

Leverage CA's Natural Language Ability

Our study suggests that, if designed properly, a CA can perform satisfactorily as a joint reading partner for children. In our observation, children actively participated in conversation with the CA and frequently generated on-topic responses. Children were generally able to respond to the CA within the proper time frame. Children also showed positive affect while speaking to the CA or listening to the CA's feedback. We attribute children's engagement to how we designed the CA to invite children's verbal engagement and respond to children.

Inviting Children's Responses

Our CA used a combination of open-ended questions (i.e., Initial Prompts) and multiple-choice questions (i.e., Follow-up Prompts), which worked together to support children's interactions with the CA. The initial open-ended questions were designed to encourage children's free expression of their thoughts related to the story (Applegate et al., 2002; Gazella & Stockman, 2003). Indeed, we found that children articulated their understandings more fully and in a more grammatically complex way when responding to open-ended questions. Interestingly, we also observed that children's responses to the open-ended questions commonly involved some personal connection the child had to the topic. Although these responses sometimes did not directly answer the question, they were almost always accompanied by children's increased affective engagement, and we believe this engagement makes the joint-reading experience more relatable for the children (Ketch, 2005). However, we also note that this excitement may not be directly linked to the topics being discussed (Etta & Kirkorian, 2019), but may, in general, reflect children's enjoyment of having open-ended conversation with a digital learning partner.

Despite the benefits of using open-ended questions, this approach is not without costs. While open-ended questions can stimulate thinking, responding to them may also require more cognitive resources, sometimes exceeding a child's capacity. Additionally, the freedom in formulating a response may lead children astray from the topic at hand. As such, broadly open-ended questions may lead children to either not answer the questions or answer them with lower topical relevance. Moreover, unlike humans, CAs' response quality may degrade substantially when discussing unrestricted topics; a CA's performance is largely reliant on the designers' ability to predict children's responses, and open-ended questions can result in greater variation and unpredictability. Indeed, in our observations, children's responses to unrestricted Initial Prompts contained more unpredicted content, resulting in the CA's decreased accuracy rate when categorizing responses for Initial Prompts.

In order to balance the drawbacks of open-ended questions, we introduced more restricted multiple-choice questions as Follow-up Prompts. These types of questions can help ease the potential cognitive obstacles and redirect children's attention to the story content. They also have benefits for the CA's performance, since they keep children's likely responses within what the CA is capable of categorizing, thereby preventing possible conversation breakdown. This strategy of including restricted questions as a way to recover from impediments in the preceding conversational turns resonates with the notion of adaptability commonly suggested in conventional educational pedagogy (Fogleman et al., 2011; Fuchs et al., 1992). However, adaptability traditionally involves the more experienced language partner tailoring the conversation to meet the child's ability; in CA-child communication, including restricted questions also adjusts the child's response to accommodate the CA's ability in understanding.

Responding to Children

The CA was intended to provide specific feedback to children's responses. Specific feedback acknowledges what a child has said and then moves the conversation forward based on the child's input. The CA's capability of providing specific feedback depended on how accurately the CA could interpret children's responses and map those to intent categories. As discussed before, we attempted to achieve this goal through creating fine-grained categorization of children's possible language input so that the CA could provide precise feedback based on children's utterance. In our observations, we found that specific feedback kept children emotionally engaged, while vague feedback resulting from failed categorization generally did not facilitate engagement.

While the value of specific feedback has been emphasized in adult-child communication, we think that specific feedback is especially important in human-machine interaction. As voice interfaces cannot provide other social cues (e.g., eye-gaze, facial expression) through which children can infer that their responses have been correctly registered, feedback plays a role to reassure children that the agent understands their input. This viewpoint has been confirmed in multiple HCI papers that examined how CAs should respond to users to demonstrate CAs' good listenership and understanding (Branham & Mukkath Roy, 2019; Clark et al., 2019).

Interaction Challenges Inherent to Voice Interfaces

While CAs' natural language abilities make it possible to simulate a human reading partner, there exists some interaction challenges inherent to voice interfaces. However, it is still possible to improve children's conversation experiences through optimizing the conversational design.

First, CAs in the form of a smart speaker do not have the capability to capture and interpret non-verbal expressions. Children were not fully aware of this inability, and the children thus tended to use both verbal and non-verbal communication when responding to the CA. On the one hand, children's engagement in multi-channel expressions is similar with what has been identified in child-parent conversations during story reading, suggesting that the CA elicited children's natural responses. On the other hand, children's use of non-verbal expressions may lead to the CA's failure to register their responses, and the conversation flow can suffer. One possible way to eliminate this issue would be having the CA emphasize its ability to listen and prime the children to respond verbally. For example, the CA may explicitly include the words "tell" or "say" within questions, such as "Please tell me whether the bears broke a blue seashell or a honey jar." The CAs' reliance on speech may actually be positive, since this reliance—once understood by children—encourages children to practice verbal communication vital for their language development.

Second, research on human-to-human communication suggests that the most common turn-taking pattern is one-party-at-a-time and that speaker changes typically occur without any silence in between and without any overlapping speech (i.e., no-gap-no-overlap). Although infrequent, human-to-human interactions can sometimes involve some overlap and gaps, but CAs are not currently capable of allowing this flexibility. In communicating with CAs, the turn-taking schema must be followed rigidly. As such, awareness of conversational timing is especially important for interacting with CAs. Unsurprisingly, we observed that children sometimes did not follow the "no-gap-no-overlap" rule. This violation may be, in part, due to children's unfamiliarity of CA's restrictions, and may also in part arise from the design of question prompts. For prompts that elicit responses from the children prior to the prompt's completion, we suggest avoiding any questioning tone if there is no intention to immediately invite a child's response. For example, our CA originally asked, "Did they break a blue seashell (?) or did they break a honey jar?" This could be rephrased as "The bears broke something: a blue seashell or a honey jar. Which one did they break?" As for gaps, the question prompts that are particularly difficult led to gaps before or pauses during a child's response. We suggest that developers ensure the difficulty level of all prompts is such that children can maintain relatively constant responses. This suggestion is consistent with the literature on parent-child interaction that recommends parents ask questions within a child's "zone of proximal development" (Kermani & Brenner, 2000; Pellegrini et al., 1985).

Age Differences in Engagement

Younger children's communication patterns differed from those of older children. The age difference in language production and flow maintenance was expected, given young children's less developed language skills and reading ability. This is in line with many studies in adult-child joint-reading that suggest younger children are less able to generate verbal responses that are topically relevant to the conversation (Miller & Chapman, 1981). In addition, younger children appeared to face more challenges when interacting with the CA, mostly due to their lack of awareness of the unique nature of artificial voice interfaces (Cheng et al., 2018; Xu & Warschauer, 2019). Despite younger children's obstacles in participating in the conversation, the younger children seemed more interested in the CA reading partner than did the older children. One possible explanation is younger children's increased tendency to perceive CAs as human-like social beings, whereas older children tend to view CAs simply as machines (Jipson & Gelman, 2007). Younger children may thus approach their interactions with the CA with greater enthusiasm.

We also observed that the age difference among responses to Follow-up Prompts were generally smaller than those to Initial Prompts. In particular, Follow-up Prompts increased younger children's relevant responses by 23 percent as compared to the 6 percent increase among older children. This further suggests the CA's scaffolding mechanism gears towards supporting younger children who are more in need of it.

Limitation and Future Work

First, while our study has proven feasibility of using a CA to simulate parent-child interaction during joint-reading activities, a follow-up experimental study should be conducted to compare children's engagement with the CA partner compared to their engagement with a human partner, in order to examine the effectiveness of the CA partner. Second, as it is unclear how the effectiveness of unimodal CAs (speech-only) compares to robots capable of carrying out multimodel interactions, future studies may explore whether CAs and robots with the same conversation design result in children's different patterns of engagement.

Conclusion

In this project, we developed and tested a CA reading partner that can read storybooks to children and actively engage them in conversations relevant to the story. The main goal of this CA is to simulate a conversation-rich, interactive reading experience similar to that of an adult partner guiding meaningful language exchange. Through examining children's conversations with the CA, we identified patterns in how the design of the CA's questions, feedback, and scaffolding influences children's responses. Overall, children responded to the CA's conversational guidance in many ways consistent with the literature on parent-child communication. Children's natural communication with CAs is encouraging and may indicate that CAs have, from the child's perspective, effectively simulated a dialogic partner, while children's assumptions of CAs' capabilities lead to some interaction challenges. As such, it is important to leverage what CAs have in common with human partners (i.e., the natural language ability) but also recognize CAs' own unique properties as artificially intelligent interlocutors. Our study suggests that, rather than attempting to develop CAs as an exact replicate of human conversational partners, we should treat child-agent interaction as a new genre of conversation and calibrate CAs based on children's actual communicative practices and needs.

References

Applegate, M. D., Quinn, K. B., & Applegate, A. J. (2002). Levels of thinking required by comprehension questions in informal reading inventories. *The Reading Teacher*, *56*(2), 174–180.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*(6111), 1225–1229. https://doi.org/10.1126/science.1224313

Beneteau, E., Richards, O. K., Zhang, M., Kientz, J. A., Yip, J., & Hiniker, A. (2019). Communication breakdowns between families and Alexa. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 243. https://doi.org/10.1145/3290605.3300473

Branham, S. M., & Mukkath Roy, A. R. (2019). Reading between the guidelines. *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 446–458. https://doi.org/10.1145/3308561.3353797

Brennan, S. E., & others. (2005). How conversation is shaped by visual and spoken evidence. *Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions*, 95–129.

Cheng, Y., Yen, K., Chen, Y., Chen, S., & Hiniker, A. (2018). Why doesn't it work?: Voice-driven interfaces and young children's communication repair strategies. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 337–348. https://doi.org/10.1145/3202185.3202749

Clark, L., Munteanu, C., Wade, V., Cowan, B. R., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., & Murad, C. (2019). What makes a good conversation? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–12. https://doi.org/10.1145/3290605.3300705

Etta, R. A., & Kirkorian, H. L. (2019). Children's learning from interactive eBooks: Simple irrelevant features are not necessarily worse than relevant ones. *Frontiers in Psychology*, *9*, 2733. https://doi.org/10.3389/fpsyg.2018.02733

Fogleman, J., McNeill, K. L., & Krajcik, J. (2011). Examining the effect of teachers' adaptations of a middle school science inquiry-oriented curriculum unit on student learning. *Journal of Research in Science Teaching*, *48*(2), 149–169. https://doi.org/10.1002/tea.20399

Fuchs, L. S., Fuchs, D., & Bishop, N. (1992). Instructional adaptation for students at risk. *The Journal of Educational Research*, *86*(2), 70–84. https://doi.org/10.1080/00220671.1992.9941143

Gazella, J., & Stockman, I. J. (2003). Children's story retelling under different modality and task conditions. *American Journal of Speech-Language Pathology*, *12*(1), 61–72. https://doi.org/10.1044/1058-0360(2003/053)

Halliday, S. E., Calkins, S. D., & Leerkes, E. M. (2018). Measuring preschool learning engagement in the laboratory. *Journal of Experimental Child Psychology*, *167*, 93–116. https://doi.org/10.1016/j.jecp.2017.10.006

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568. https://doi.org/10.1016/j.wocn.2010.08.002

Jipson, J. L., & Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child Development*, *78*(6), 1675–1688. https://doi.org/10.1111/j.1467-8624.2007.01095.x

Kang, J. Y., Kim, Y.-S., & Pan, B. A. (2009). Five-year-olds' book talk and story retelling: Contributions of mother—child joint bookreading. *First Language*, *29*(3), 243–265. https://doi.org/10.1177/0142723708101680

Kermani, H., & Brenner, M. E. (2000). Maternal scaffolding in the child's zone of proximal development across tasks: Cross-cultural perspectives. *Journal of Research in Childhood Education*, *15*(1), 30–52. https://doi.org/10.1080/02568540009594774

Ketch, A. (2005). Conversation: The comprehension connection. *The Reading Teacher*, *59*(1), 8–13. https://doi.org/10.1598/rt.59.1.2

Leppänen, J. M., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological Research Psychologische Forschung*, *69*(1–2), 22–29. https://doi.org/10.1007/s00426-003-0157-2

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech, Language, and Hearing Research, 24*(2), 154–161. https://doi.org/10.1044/jshr.2402.154

Pellegrini, A. D., Brody, G. H., & Sigel, I. E. (1985). Parents' book-reading habits with their children. *Journal of Educational Psychology*, *77*(3), 332. https://doi.org/10.1037/0022-0663.77.3.332

Robins, B., Dickerson, P., Stribling, P., & Dautenhahn, K. (2004). Robot-mediated joint attention in children with autism: A case study in robot-human interaction. *Interaction Studies*, *5*(2), 161–198. https://doi.org/10.1075/is.5.2.02rob

Rozin, P., & Cohen, A. B. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion*, *3*(1), 68. https://doi.org/10.1037/1528-3542.3.1.68

Ruusuvuori, J. (2013). Emotion, affect and conversation. *The Handbook of Conversation Analysis*, 330. https://doi.org/10.1002/9781118325001.ch16

Sánchez, J. A., Hernández, N. P., Penagos, J. C., & Ostróvskaya, Y. (2006). Conveying mood and emotion in instant messaging by using a two-dimensional model for affective states. *Proceedings of VII Brazilian Symposium on Human Factors in Computing Systems - IHC '06*, 66–72. https://doi.org/10.1145/1298023.1298033

Sidner, C. L., Kidd, C. D., Lee, C., & Lesh, N. (2004). Where to look: A study of human-robot engagement. *Proceedings of the 9th International Conference on Intelligent User Interface - IUI '04*, 78–84. https://doi.org/10.1145/964442.964458

Wanska, S. K., Pohlman, J. C., & Bedrosian, J. L. (1989). Topic maintenance in preschoolers' conversation in three play situations. *Early Childhood Research Quarterly*, *4*(3), 393–402. https://doi.org/10.1016/0885-2006(89)90023-9

Xu, Y., & Warschauer, M. (2019). Young children's reading and learning with conversational agents. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. https://doi.org/10.1145/3290607.3299035

Table 4.1

CA Performance in Intent Categorization

	Accurate categorization		Inaccurate Categorization		Categorization failed	
	Initial Prompts	Follow-up Prompts	Initial Prompts	Follow-up Prompts	Initial Prompts	Follow-up Prompts
Full sample	76.7%	83.7%	0.3%	0.7%	22.7%	15.6%
Younger	70.3%	80.2%	1.6%	1.6%	28.1%	18.2%
Older	80.1%	86.5%	1.1%	1.0%	18.8%	12.5%
Age difference	Initial Prompts: $\chi^2(2) = 9.49$, p <0.001; Follow-up Prompts: $\chi^2(2) = 6.68$, p <0.05					

Table 4.2

Language Production

	Verbal Expressions		Response Length	
	Initial Prompts	Follow-up Prompts	Initial Prompts	Follow-up Prompts
Full	86.2%	89.3%	4.1	2.4
Younger	80.0%	82.3%	4.2	2.4
Older	91.1%	97.7%	4.1	2.4
Age difference	$\chi^2(1) = 7.53$ $p < 0.01$	$\chi^2(1) = 10.28$ $p < 0.01$	$F(1, 281) = 0.08$ $p = 0.78$	$F(1, 180) = 0.07$ $p = 0.80$

Table 4.3

Topic Relevance in Children's Responses

	Response relevant to the question	
	Initial Prompts	Follow-up Prompts
Full	76.7%	89.6%
Younger	64.9%	87.8%
Older	86.5%	91.7%
Age difference	$\chi^2(1) = 17.12$ $p < 0.001$	$\chi^2(1) = 0.38$ $p = 0.54$

Table 4.4

Timing of Children's Responses

	Gaps and Pauses		Rushed Responses	
	Initial Prompts	Follow-up Prompts	Initial Prompts	Follow-up Prompts
Full	21.6%	11.5%	8.4%	24.2%
Younger	29.7%	13.3%	5.5%	31.6%
Older	14.8%	9.5%	11.0%	15.5%
Age difference	$\chi^2(1) = 8.28$ $p < 0.01$	$\chi^2(1) = 0.31$ $p = 0.58$	$\chi^2(1) = 5.88$ $p < 0.05$	$\chi^2(1) = 6.15$ $p < 0.05$

Table 4.5

Affect in Children's Responses to the CA

	Positive Affect		Neutral Affect	
	Initial Prompts	Follow-up Prompts	Initial Prompts	Follow-up Prompts
Full	25.3%	14.4%	74.7%	85.6%
Younger	31.1%	15.3%	68.9%	84.7%
Older	21.8%	13.4%	78.2%	86.6%
Age difference	Initial Prompts: $\chi^2(3) = 8.05$, $p < 0.05$; Follow-up Prompts: $\chi^2(3) = 2.42$, $p < 0.49$			

Note. Negative affect or confusion was not observed.

Table 4.6

Children's Affective Reactions to Receiving Positive Feedback from the CA

	Positive Affect		Neutral Affect	
	Initial Prompts	Follow-up Prompts	Initial Prompts	Follow-up Prompts
Full	75.2%	73.4%	24.8%	26.6%
Younger	82.3%	81.8%	17.7%	18.2%
Older	67.9%	66.5%	32.1%	33.5%
Age difference	Initial Prompts: $\chi^2(2) = 11.65$, $p < 0.01$; Follow-up Prompts: $\chi^2(2) = 11.52$, $p < 0.01$			

Note. Negative affect or confusion was not observed.

Table 4.7

Children's Affective Reactions to Receiving Negative Feedback from the CA

	Positive Affect		Neutral Affect		Negative Affect		Confusion	
	Initial Prompts	Follow-up Prompts	Initial Prompts	Follow-up Prompts	Initial Prompts	Follow-up Prompts	Initial Prompts	Follow-up Prompts
Full	0.0%	0.0%	82.4%	83.4%	15.9%	15.3%	1.7%	1.3%
Younger	1.4%	0.0%	80.2%	79.5%	16.8%	17.9%	3.0%	2.6%
Older	0.0%	0.0%	85.6%	86.5%	14.4%	13.5%	0.0%	0.0%
Age difference	Initial Prompts: $\chi^2(3) = 7.16$, $p = 0.07$; Follow-up Prompts: $\chi^2(3) = 7.73$, $p = 0.05$							

Table 4.8

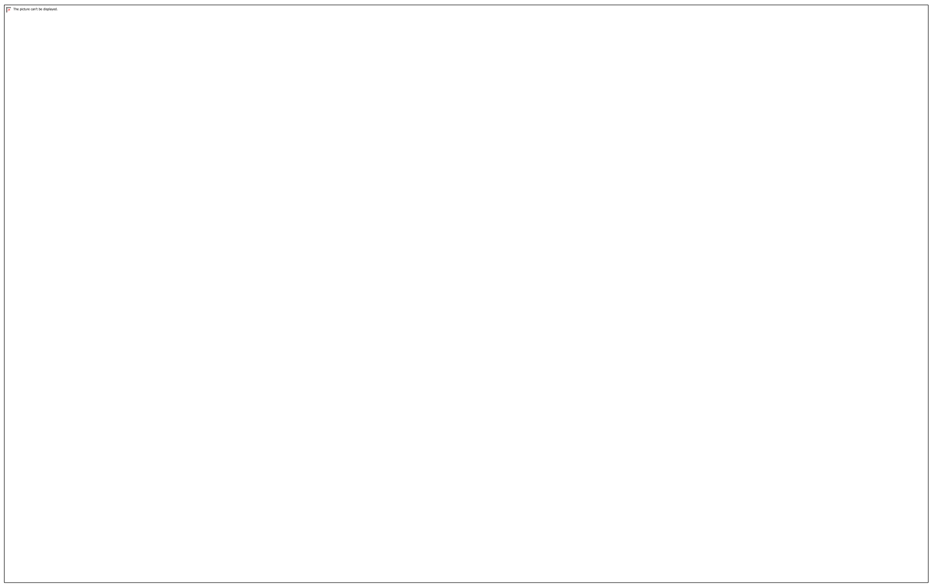
Children's Affective Reactions to Receiving Vague, Neutral Feedback from the CA

	Neutral Affect		Confusion	
	Initial Prompts	Follow-up Prompts	Initial Prompts	Follow-up Prompts
Full	93.5%	90.1%	6.5%	9.9%
Younger	94.2%	92.2%	5.8%	7.8%
Older	93.0%	88.9%	7.0%	11.1%
Age difference	Initial Prompts: $\chi^2(3) = 0.72$, $p = 0.87$; Follow-up Prompts: $\chi^2(3) = 6.67$, $p = 0.08$			

Note. Positive or negative affect was not observed.

Figure 4.1

Child-CA Dialogue Flow



CHAPTER 5 PERCEPTIONS[6]

Study Abstract

Conversational agents (CAs) available in smart phones or smart speakers play an increasingly important role in young children's technological landscapes and life worlds. While a handful of studies have documented children's natural interactions with CAs, little is known about children's perceptions of CAs. To fill this gap, we examined three- to six-year-olds' perceptions of CAs' animate/artifact domain membership and properties, as well as their justifications for these perceptions. We found that children sometimes take a more nuanced position and spontaneously attribute both artifact and animate properties to CAs or view them as neither artifacts nor animate objects. This study extends current research on children's perceptions of intelligent artifacts by adding CAs as a new genre of study and provides some underlying knowledge that may guide the development of CAs to support young children's cognitive and social development.

Introduction

As conversational agents (CAs) become increasingly prevalent in home life, whether through smart phones, tablets, or smart speakers, both scholars and the general public have noted young children's propensity to interact with them (Brunick et al., 2016; Druga et al., 2017; Lovato et al., 2019; Serholt & Barendregt, 2016).

CAs are designed to take on many of the properties previously thought to be unique to humans. Specifically, CAs support natural spoken conversation, thus displaying a high level of intelligence. Moreover, some CAs have been designed as social companions for children

[6] A version of this chapter was published in the *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*

(Kanero et al., 2018; Kim et al., 2007) and are capable of provoking social reactions, such as empathy and trust (Purington et al., 2017). However, CAs available in phones or smart speakers are neither anthropomorphic nor self-locomotive, making them physically different from a human dialogue partner. CAs' human-like capabilities without corresponding physical features create an intriguing research scenario for examining child-CA interaction. A handful of studies have found that children interacting with CAs utilize communication strategies similar to those normally used when interacting with a human interlocuter (Sciuto et al., 2018; Tewari & Canny, 2014; Xu & Warschauer, 2019). However, we know very little about children's perceptions during such interactions, particularly whether children attribute human properties to non-human CAs.

The question of how children perceive CAs is of interest to the fields of developmental psychology and human-computer interaction (HCI). First, this question is relevant to the long-standing focus within developmental psychology on the animate-inanimate (A-I) distinction in early childhood (Rakison & Poulin-Dubois, 2001). Given that CAs are highly interactive and intelligent, they may blur children's categorical distinctions between technological artifacts on the one hand and biological beings on the other (Jipson & Gelman, 2007; Severson & Carlson, 2010). CAs' blurring of these boundaries may result in children categorizing CAs as neither artifacts nor living beings (Reeves & Nass, 1996) or perceiving CAs as occupying some middling position along an animate-inanimate continuum (Kahn et al., 2004; Kim et al., 2019). Second, children's perceptions of CAs are of importance to HCI given that this field is keenly interested in developing CAs that simulate human-to-human communication (Luger & Sellen, 2016; Woodward et al., 2018). Within the field of HCI research, children's behavioral interactions with CAs are typically used to evaluate whether CAs have gotten closer to that gold standard (Hill et al., 2015; Tewari & Canny, 2014). However, these studies fail to consider children's perceptions, which is another integral facet of children's experiences with CAs as perceptions shape behavioral interactions (Kory-Westlund & Breazeal, 2019). If children ascribe life characteristics to CAs, they may then engage in more natural communication patterns with those CAs. In contrast, if children view CAs as simply machines or tools, they may approach their interactions in a less natural way. Therefore, understanding children's perceptions may help make sense of previous research on child-CA interactions and provide a more complete picture of children's relationships with CAs. Moreover, children's perceptions of CAs' properties and human/non-human status may reveal children's expectations for consumer-level CAs, which could be helpful for development of future CAs.

The present study is grounded in and extends the developmental psychology and HCI lines of research above. We build on existing research of early childhood A-I distinction and extend its application to how children understand intelligent artifacts, in particular, speaker-based CAs. We also explore children's perceptions, in particular whether children view CAs as human-like dialogic partners. Specifically, this study seeks to answer three questions. First, which domain do children perceive CAs as belonging to (e.g., artifact, living object, or something else)? Second, do children view CAs as possessing human-like cognitive, psychological, and behavioral properties? Third, how do children reason about whether CAs possess certain properties? To answer these three questions, 28 children aged 3 to 6 were invited to individually interact with a CA, after which we elicited their perceptions through a semi-structured interview and a drawing task. This study is intended to reveal children's perceptions behind their active engagement with a CA and offer theoretical and design implications. It is interesting to focus on children aged 3 to 6 primarily because children in this age group have developed a naïve framework of beliefs about living things, which emerged in the absence of formal instruction (Marshall & Brenneman, 2016), and do not yet have a sophisticated conceptualization of computational objects (Rücker & Pinkwart, 2016).

Literature Review

Understanding of Animacy in Early Childhood

Children's understanding of the A-I distinction – the distinction between living and non-living things – is probably one of the most enduring questions in developmental psychology (Rakison & Poulin-Dubois, 2001). Indeed, the ability to recognize objects as animate or inanimate is thought to be a fundamental cognitive process since it provides the foundation upon which children categorize objects in the world (Caramazza & Shelton, 1998; Rakison & Poulin-Dubois, 2001). Children's primitive understanding of the A-I distinction begins in infancy, develops rapidly during early childhood, and matures in adolescence (Opfer & Gelman, 2011). Melson et al., synthesizing research on the topic, suggested that a young child distinguishes between animate and inanimate things based on the child's perception of that thing's cognitive (thoughts), psychological (feelings or emotions), and behavioral (actions or speech) properties (2009). Inevitably, if an object displays either all or none of these properties, children find it less challenging to categorize the object as either animate or inanimate. However, objects that display only some of the properties are more likely to raise boundary questions for children. In other words, if A-I distinction is perceived as a continuum, some objects may be more clearly perceived to be on either end of the continuum, while some are perceived to fall in between (Mikropoulos et al., 2003).

Distinct properties play different roles in children's evaluation of whether an object is animate or inanimate. In studies on this topic, children are typically shown pictures of everyday objects and tasked with sorting them based on their membership in either category before then being asked to describe which of the object's properties informed that categorization. Two trends emerge from these studies. First, children tend to firmly, yet incorrectly, associate the ability to move physically with animacy itself. Second, children astutely understand that only animate things have the ability to think and feel (Rakison & Poulin-Dubois, 2001). Although these two trends have been observed in studies that did not involve artificial intelligence, it would not be surprising if children used these same principles to determine whether an intelligent artifact is animate or inanimate. As suggested in Edwards et al., when young children attempt to understand complex and novel technologies, they tend to apply a familiar schema they have developed from their daily lives (2018).

Children's Perceptions of Intelligent Artifacts

A growing body of research has focused on children's perceptions of intelligent artifacts, especially computers, CAs, and robots (Jipson & Gelman, 2007; Katayama et al., 2010; Kory-Westlund & Breazeal, 2019; Mertala, 2019; Mikropoulos et al., 2003; Scaife & van Duuren, 1995; Severson & Carlson, 2010). These three artifacts represent objects that may elicit different levels of perceptions of animacy given their differing properties. If children conceive of the A-I distinction as a continuum rather than a dichotomy, computers would lie close to the inanimate end, robots close to the animate end, and CAs somewhere in between.

On one end of the spectrum, robots appear to possess the cognitive, psychological, and behavioral capacities that elicit perceptions of animacy. They may move, learn, communicate, self-organize, and respond to emotions in humans. As such, it is not surprising that children who regularly interact with robots often view them as animate objects (Jipson & Gelman, 2007; Kory-Westlund & Breazeal, 2019; Severson & Carlson, 2010). Correspondingly, children tend to perceive robots as possessing all of the properties associated with animacy (Kahn et al., 2012; Melson et al., 2009). For example, Melson and colleagues found that the majority of children affirmed that AIBO, the robotic dog, had mental states, social awareness, and moral standing (2009). Similarly, Beran and colleagues suggested that a significant proportion of children in their study ascribed cognitive, behavioral, and psychological characteristics to robots (2011). This implies that children impose their own understanding of human nature onto such technological devices and see them as possessing similar capabilities. However, Beran's study also noted that children's assigning animacy to robots is driven more by robots' physical movements rather than by their intelligence (Beran et al., 2011). When children were asked why they considered the robot to be a living being, most children pointed out the robot's humanoid appearance and its seeming ability to move spontaneously. This is consistent with children's firm association of animacy with the ability to move (Tremoulet & Feldman, 2000). Given that smart speakers lack mobility and anthropomorphic embodiment, it is unclear whether the findings from robot studies will hold true for speech-only CAs.

On the other end of the spectrum, computers, while demonstrating some level of cognitive capability, typically lack the psychological and behavioral properties that children emphasize when evaluating an object's animacy. One study looked at the properties young children ascribed to computers (van Duuren & Scaife, 1996). Interestingly, while a considerable proportion of children believed that computers were capable of performing tasks that required intelligence, almost all children viewed computers as lacking psychological and behavioral capabilities. A second study also found that although children believed computers possessed moderate to high intelligence capability, they did not view computers as living objects (Scaife & van Duuren, 1995). Through analyzing children's drawings of how they think computers might look inside, Mertala suggested that children tended to view computers as machines, as most children depicted computers as technological objects such as monitors, wires, or keyboards (2019). Much like computers, CAs lack both the mobility and anthropomorphic embodiment that children readily associate with animacy. However, CAs' ability to engage in natural spoken conversation may lead children to ascribe CAs with cognitive and psychological properties similar to those in humans.

Children's Perceptions of Speech-Only CAs

Only two studies, to our knowledge, exist which speak to children's perceptions of smart speaker-based CAs. In the first study, Druga and colleagues examined how children perceive CAs' psychological properties (2017). The authors asked children to interact with different smart speakers during both free and structured play and found that most children viewed the CAs as friendly and genuine. This finding supports the idea that, although CAs lack mobility and anthropomorphic embodiment, children still view them as possessing psychological properties similar to those of robots. In the second study, Lee, Kim, and Lee asked participants with previous experience interacting with Amazon Alexa or Google Home devices to draw what they thought a CA looked like (2019). The participants produced drawings that fell within four general categories: human, speaker, system, and space object. This finding confirms that CAs' unique combination of features may induce some users to view CAs as living beings rather than inanimate objects. However, while the study included participants ranging in age from 4 to 51, it did not distinguish between drawings produced by young children and those produced by older children or adults. In addition, the study focused exclusively on domain membership perceptions and did not examine perceptions of CA properties.

The current study expands on each of these two projects by examining children's perceptions of CAs' domain membership and their cognitive, psychological, and behavioral properties, while also exploring children's explanations for these perceptions.

Method

Participants

Our participants consisted of 28 typically developing children between the ages of 3 and 6 recruited from preschools and afterschool programs in a university community. These participants were drawn from a larger study, with the total number of participants designated so as to have sufficient statistical power.

Parents or guardians completed a brief survey on demographic characteristics. According to parent reports, the mean age of the participants was 4.7 years, and 54% were girls. Nine children (32%) were identified as White. Nineteen children (68%) spoke only English at home, and the rest of them were bilingual or spoke English as a second language, but all children possessed sufficient oral English proficiency for daily conversation. Twenty-one percent of the participants had never interacted with a CA, 36% had done so less than once a month, 18% between weekly and monthly, and 25% had daily interaction with a CA.

Description of the Interaction Tasks

The interaction tasks provided children with direct and in-the-moment experience with a CA. We noted that this approach is among the three common methods utilized by prior research. One approach relies on children's past experiences with technology (e.g., Lee et al., 2019); however, very young children are less able to accurately recall past experiences (Rooy et al., 2007). The second approach involves showing children videos of how a technology works (e.g., Kim et al., 2019); however, young children mostly learn from authentic, direct experiences rather than events they indirectly witness (Heintz & Wartella, 2012). We believe that allowing children to interact with the CA during the study session will provide them proximal and in-person experiences, and thus we will be better able to elicit their perceptions.

Each child's interaction entailed three sessions with a Google Home Mini device and lasted approximately 40 minutes in total. The child first had a structured personal conversation with the CA, then played a structured narrative game, and finally had an unstructured dialogue. These three sessions mirrored the common interaction experiences a child typically would have with CAs in their everyday lives (Druga et al., 2018; Sciuto et al., 2018).

Personal Conversation

In the first session, the CA asked children their age, favorite color, and a simple animal question (i.e., which animal has a really long neck?). The CA was programmed to repeat children's responses. For example, when a child tells the CA that he or she is five years old, the CA responds, "Wow, you're five years old. You are such a big kid!" When a child tells the CA that his or her favorite color is blue, the CA responds, "Great choice! My favorite color is also blue." In cases where the CA failed to understand a child's voice input, whether due to fuzzy pronunciation, an irrelevant response, or some other issue, the CA adopted a fallback mechanism to move the conversation forward. Such mechanisms included more general, neutral responses that did not directly repeat the child's answers (e.g., "Great choice! That is my favorite color too.").

Narrative Game

In the second session, the CA read a ten-minute fantasy story and asked the children 9 story-related open-ended questions throughout. The CA gave responsive feedback based on a child's answer, either praising the child for a correct answer or encouraging the child to try again after an incorrect answer. In the latter case, the CA provided hints or rephrased the original question into a multiple-choice format, with the goal of simulating how an engaging adult would scaffold children's learning and conversation during shared reading. A fallback mechanism was triggered if the CA failed to capture a child's response twice in a row. The CA gave a general, neutral response without repeating the child's response and moved the story forward.

Unstructured Dialogue

In the final session, children were encouraged to freely talk to the CA and ask the CA any questions they would like. Common topics included math (e.g., "What is one hundred plus one hundred?"), culture facts, personal questions (e.g., "How old are you?"), and the child's sharing of personal information (e.g., "My favorite princess is Elsa."). These topics corroborated findings from (Lovato et al., 2019).

Procedure

Each child met individually with a trained experimenter in a designated quiet area at the child's school. At the beginning of each session, the experimenter introduced the interaction task as a game and described the Google Home Mini device as "Google." During the interaction sessions, the experimenter sat beside the child, interfering only if technical issues interrupted the child's interactions with the CA (e.g., Internet or battery issues). In the case that a child asked the experimenter questions or initiated comments, the experimenter simply answered the question or replied "okay," but avoided elaborating or extending the conversation. After the child completed all sessions, the experimenter administered a semi-structured interview and a drawing task to elicit children's perceptions, as discussed below.

Analytic Strategies

To analyze video and interview transcription data, we used a hybrid approach to thematic analysis (Swain, 2018): we incorporated both a data-driven inductive process and a deductive process where we referenced the relevant frameworks outlined in previous studies to inform our coding. For the domain membership, we referenced the framework in Khan et al. (2006) and Kim et al. (2019); for the justification of property attribution, we referenced the framework in van Duuren and Scaife (1996) and Melson et al. (2009). The inductive process produced a set of a priori codes that came from children's responses to interviews, and the deductive process allowed us to re-formulate our codes based on existing theories.

To establish inter-rater reliability, two coders were involved in the coding process: Coder One coded data from all participants, and Coder Two coded data from 30% to establish reliability. The two coders met weekly for one month to calibrate their coding. Specifically, of the child participants Coder One analyzed each week, Coder Two randomly selected 30% to perform the coding. Discrepancies in coding were used to iteratively refine the coding protocol until an inter-rater reliability of 85% was achieved.

Interview Data

For the open-ended question on what children thought they were talking to (i.e., the domain membership question), we categorized children's responses into three groups: artifacts, living objects, and a residual category for all other descriptions based on the framework in Kahn et al. (2006) and Kim et al. (2019).

For the affirmative questions on property attributions, each question was coded as an affirmation (e.g., "yes," "I think so," or nodding) or a negation (e.g., "no", "I don't think so," or shaking head). In some instances, children had difficulty deciding on a response, so we created a separate category ("I don't know") to capture this type of response.

To code children's verbal response to the open-ended follow-up questions regarding their justification for their property attribution, we developed a scheme with 9 codable categories, derived from protocols used in van Duuren and Scaife (1996) and Melson et al. (2009). Children's justifications for thinking CAs possessed a certain property were classified as 1) **domain references** if the child relied on the domain they perceived the CA to belong to (e.g., "It is just a machine."); 2) **analogical reasoning** if the child compared the CA to other familiar objects and pointed to either similarities (e.g., "It is like phones so it can talk.") or differences (e.g., "It is not like a human so it can't remember well."); 3) **biological references** if the child indicated the CA possessed or lacked body parts or internal organs (e.g., eyes, heart); 4) **physical feature references** if the child mentioned the material the CA was made of (e.g., "It is plastic.") or the appearance of the CA (e.g., "It is orange, and a person can't be orange."); 5) **mental state references** if the child pointed to the CA's mental state, such as knowing, perceiving, and emotion (e.g., "It learns a lot of things"; "It was trying to be kind and nice to me. That's her personality."); 6) **behavioral references** if the child mentioned what the CA did (e.g., "It just read stories to me.") or how the CA behaved (e.g., "It just listened to me nicely."); 7) **reciprocity** if the child believed the CA's properties were results of others' actions, in particular, the child's own actions; 8) **mechanical references** if the child believed the CA's properties were the result of human programming (e.g., "It is made to be smart."); and 9) **fantasy reasoning** if the child attributed the CA's properties to magic or a supernatural power (e.g., "Google uses magic to listen"; "It is a witch."). One response could be coded for multiple justifications, and off-topic responses and "I don't know" were coded as invalid.

Drawing Data

Data generated from the drawing task were intended to supplement the findings from interview data. Hence, we combined the two data sources when presenting the findings on children's perceptions of the CA's domain membership and properties. Given that most of children's drawings are hard to interpret without referring to their explanations, we annotated each drawing sample based on the child's verbal explanation. The drawing samples as well as the accompanying verbal accounts were coded in relation to how the CA's domain membership and cognitive, psychological, and behavioral features were exhibited in them.

Results

Domain Membership

Our first research question focuses on which domain children perceived the CA as belonging to. Children's interview answers can be grouped into three domains: 1) artifacts, 2) living objects, and 3) a residual category, which was assigned for any description that is neither artificial nor living (see Table 1).

Over half of the children (n = 16, 57%) conceived the CA as an artifact. In interviews, children provided differing levels of specificity when describing CAs: some children broadly described the CA as a "device," "machine," or "tool," while others described the CA as a specific object, such as a "phone," "speaker," "CD-player," "robot," or "app." A small proportion of children (n = 5, 18%) viewed the CA as a living object. All of these children described it as "human," and one child specifically said that the CA was a "girl." A considerable proportion (n = 7, 25%) of children indicated that the CA was neither an artifact nor a living object. A variety of responses were grouped together into this residual domain, such as "Google is some sort of girl," "something very special that can talk like us but not a person," "a sound we can't see," "magic things to talk," and "Google is Alexa, Alexa is Google. They are not other things."

We also analyzed children's drawings using the same three coding categories as interview data. Half of the children's drawings presented the CA as a technological artifact (n = 14, 50%). These drawings suggested the children did not understand CAs as simple objects but instead as complex machines that contained multiple components (Lee et al., 2019). These drawings typically depicted the actual shape of the smart speaker the children had interacted with (i.e., a circle). But within that outer shape, these children drew clusters of multiple objects, including wires, microphones, speakers, electricity, batteries, light bulbs, radios, or nails. The prevalence of these components may be because they were either visible (e.g., light and nails) or familiar to children from other devices they had experiences with (e.g., batteries and wires). Figure 1a contains several rounds of pink wires and a red line that connects the wires and makes them "work together," and Figure 1b contains a microphone, wires, and electricity within a circle. Each of these drawings presented the CA as a connected system with all of its parts functioning synergistically. One child noted as he pointed to the "wires," "microphones," "holes," "plugs," and "connectors" he drew, the CA has "a lot of things. All of these help it speak.

Almost one fifth of the children illustrated the CA as a human face or human-like figure (n = 5, 18%). However, none of these drawings depicted a complete human figure, but all contained the most vital elements of a human from a child's perspective. For example, Figure 2a only illustrates the CA as a face with two eyes, a nose, and a mouth yet without a body, and Figure 2 displays a girl who does not have arms or legs. Such incompleteness in human figures may reveal that although these children were inclined to identify CAs as living things, the children were also aware of some typical human features that the CA lacked.

The remainder of the drawings (n = 9, 32%) contained a mixture of representations of human and artifact elements or representations that could not be clearly categorized into either domain. For the drawings that contained both an artifact component and a living object component, children typically included a round outer shape similar to those drawings that represent the CA as a technological object, but included human figures or human body parts inside. For example, in Figure 3a, the rectangle and wavy lines represent a speaker and wires that "bring different parts together, so it won't fall apart," while inside that speaker is a human figure and the foods he/she requires. For the drawings that could not be categorized into either technological objects or living beings, children represented the CA as a wide range of varying things. For example, a child drew the CA as an apple with juice, flesh, and seed because the Google Home device was orange and the flashing lights looked like seeds. Another child drew the CA as lightning because there was a storm sound during the narrative game. A third child combined a variety of shapes and colors to represent the CA as sound (Figure 3b). This child indicated the CA was "rainbow sound" which is "happy and smart." Overall, this group of drawings was centered narrowly on certain micro-level features of the CA that stood out to each child (e.g., the CA's color, its lights, or a sound it made).

For the majority of children, their drawings corroborated their interview answers (n = 24, 86%). Every child who categorized the CA as human in the interview also drew the CA with clearly anthropomorphic features. Instances in which drawings differed from interview responses only occurred among children who did not identify the CA as a human.

Property Attribution

Our second research question explored children's attribution of cognitive, psychological, and behavioral properties to CAs (Table 2). The vast majority of children believed that the CA possessed cognitive ability; these children stated that the CA was smart (n = 26, 93%) or that it could remember their previous conversation well (n = 24, 86%). A slightly smaller majority of all children attributed psychological properties to the CA, indicating that the CA could like others as a friend (n = 18, 64%) and feel emotion (n = 19, 68%). Lastly, in terms of behavioral properties, the majority of children indicated that the CA possessed speech-related capabilities (listening, n = 25, 89%; talking, n = 26, 93%), but only a quarter of these children believed that the CA could see (n = 7, 25%).

We found that children's drawings also contained these cognitive, psychological, and behavioral elements. However, given the inherent difficulty in visually representing these three elements, we focus here exclusively on those drawings where children provided relevant clarifications.

In Figure 4a, for example, a child indicated the CA's cognitive properties by drawing letters within the CA (i.e., the letters "c" and "w") to signify that the CA is "smart and knows a lot of things," and in Figure 4b, another child wrote her age (4) using her favorite color (pink) explaining that the CA put this information in its memory. Figure 5a and Figure 5b shows two children's drawings indicating psychological properties. In Figure 5a, a child drew a heart and a smiley face, commenting that the CA "knows if I am happy" and is "sometimes happy but sometimes not." In Figure 5b, a child drew a rainbow, a smiley face, and rain drops, commenting that the CA "has a rainbow inside that makes it laugh and happy" and "rains inside if Google is sad." Representations of behavioral properties were rarer. In Figure 6, a child drew a large mouth and said, "This is why Google can talk so loud."

Children's Justification of Property Attributions

We then analyzed children's explanation for their attribution of cognitive, psychological, or behavioral properties to the CA. The most frequently occurring justifications across all properties referred to the CA's presumed behaviors, biological features, mental states, or the reciprocal relationships between the child and the CA (see Table 3). Unique patterns of justifications also appeared when children were deciding whether the CA had any of the three properties (see bolded, italicized numbers in Table 3):

- Cognitive properties were most frequently justified through behavioral references,
- psychological properties most frequently through references to reciprocity, and
- behavioral properties through biological references, mechanical causality, and fantasy reasoning.

In terms of cognitive property attribution, children commonly relied on their observation of the CA's behaviors, particularly its communication techniques. Two techniques we programmed were frequently mentioned by children as a sign of cognition, namely the repetition strategy that allows the CA to repeat what a child has said and the fallback strategy that ensures that the CA always responds to the child to prevent communication breakdowns. For example, a child stated that Google had a good memory because it "just repeated what I told her," and another child mentioned that "Google was smart because it always talks back to me."

When justifying attribution of psychological properties, children frequently referred to reciprocity; their own actions led to the CA's affective reactions. For example, one child commented that the reason why she thought Google liked her was because she was nice to Google, and another child said that Google may have felt sad when she was not listening to the story. These comments suggest that children believed the CA could reciprocate socially or emotionally on children's behaviors.

A more complex pattern was frequently observed in children's justification of behavioral properties. Children tended to first search for the biological features commonly associated with a particular behavior (i.e., eyes to see, ears to hear, and a mouth to talk). When children could not justify their attribution of a particular property through biological references (e.g., when the CA can listen or talk but doesn't have ears or a mouth), they tended to resolve the conflict through mechanical explanations or fantasy reasoning. For example, one child noted that "we installed a speaker so it can talk without a mouth," while another child said that the CA "talks with a magic mouth we can't see."

Discussion

CAs as Humans, Artifacts, and What Else?

Our first research question examined children's categorization of the CA's domain membership. We found that some children associated the CA with artifacts or humans, while some children provided answers that did not fit either of these two categories. Children's categorizing CAs as artifacts or humans is consistent with the traditional A-I distinction proposed in the developmental psychology research (Wright et al., 2015). The traditional distinction suggests that children develop their understanding of living or non-living things during early childhood and then use this schema to classify things they encounter in their daily lives (Wright et al., 2015). In addition to these two domains, an ambiguous status of CAs among the A-I distinction was also demonstrated by children in this study, one that does not map onto artifacts or living beings. In the interview, a number of children suggested that CAs are something unique. This was further evidenced in children's drawings, with a considerable portion representing a combination of human and artifact elements. As Kahn et al. suggest, such findings may imply that children's interactions with intelligent artifacts have led to a "new ontological category" that is cutting across prototypic categories of animate and inanimate (2006; M. Kim et al., 2019). However, as Kahn further pointed out, the English language may not be well equipped to characterize or talk about this new category (2004), and children may thus turn to use familiar, yet less accurate, terms to describe their perceived domains of CAs. Moreover, if we conceive of these categories as existing on a continuum, our evidence suggests that this new ontological category may be closer to the technological artifact side of the continuum (van

Duuren & Scaife, 1996). Every child who spoke of the CA as human also drew the CA accordingly, but approximately 20% of the other children exhibited some level of inconsistency in their depictions of the CA during the interview and drawing sessions. Taken all together, our evidence suggests that a strict distinction between animate and inanimate may fail to accurately capture children's conceptions of CAs that appear to be more nuanced and multifaceted.

Highlighted Cognition and Speech Properties

Our second research question explored what properties children perceive the CA to possess. We found that children, overall, assigned many animistic abilities to the CA. Further, children understood CAs as possessing a unique constellation of properties: almost all of the children in this study ascribed cognitive and speech-related behavioral properties to the CA, while fewer children ascribed psychological and non-speech related behavioral properties. Children overwhelmingly believed the CA to possess cognitive and speech abilities. This is not surprising as the ability to converse intelligently is a defining feature of CAs (Luger & Sellen, 2016). Nevertheless, we note that the ability to listen and talk can be considered from either a cognitive or behavioral perspective (Meichenbaum, 2017). A cognitive perspective emphasizes the mental aspects involved in listening and talking (i.e., understanding, interpreting, responding), while a behavioral perspective highlights the CA's actions (i.e., listening quietly, making sounds). It would have been difficult, if not impossible, to truly elicit the underlying perspective of the young children in this study. However, our evidence suggests a strong association between cognitive and speech-related behavioral attributions: almost all children believed the CA possessed abilities in these two categories. Also, when children justified attributing cognitive abilities to the CA, they commonly referred to the CA's speech behaviors, in particular, the communication strategies we programmed. Taken together, this implies that children in this study may understand the CA's speech behaviors as indicators of cognitive abilities.

Children's responses were more heterogeneous regarding the CA's psychological abilities; slightly over half of the children believed the CA had the ability to like others and feel emotions. In the context of the broader literature, this proportion places CAs somewhere between robots and computers in terms of perceived psychological properties. While Melson et al. (2009) and Weiss et al. (2009) suggested most children believed robots could experience happiness and sadness, Scaife and van Durren found that only 20% of five-year-olds attributed those same abilities to computers (1995). Children's differing perceptions of psychological properties of computers, CAs, and robots may be largely due to these artifacts' varying expressive abilities (Wiltshire et al., 2015). As discussed in Johnstone and Scherer, spoken languages utilize acoustic qualities (i.e., tones, pauses, pitch) that convey affective and social signals which go beyond the content of the speech (2000). In this study, we speculate that the CA's ability to engage in natural spoken conversation may have led the majority of children to ascribe psychological properties to the CA despite its lack of embodiment.

Justification of CA Properties

Our third research question looked at the explanations children provided when justifying their attribution of properties to the CA. Overall, we identified nine distinct strategies children used to decide whether the CA possessed certain abilities.

Two strategies have already been described in prior research: children may regard the CA's capabilities as either a result of programming (i.e., mechanical references) or as a result of natural intelligence (i.e., mental state references) (Levy & Mioduser, 2008). The former perspective ascribes no intentions to the artifact and considers its ability to arise from human design, while the latter ascribes intentions and awareness to the artifact itself. Relatedly, some children used justification that focused on reciprocity – the contingencies between the child's actions and those of the CA. While some justifications hinged on what appeared to be automatic responses (e.g., the CA listened only because the child spoke more loudly), many justifications hinted at a perceived social reciprocity (e.g., the CA liked a child because she was nice to the CA). These latter justifications, overlapping with mental state references, suggest that children might view CAs as psychological entities and form social relationships with CAs, a speculation which expands previous findings that show children form relationships with robots (Kahn et al., 2013; Weiss et al., 2009).

Evidence from our study suggests another perspective involving fantasy reasoning. Some children relied on magical thinking or supernatural justification to explain how the CA could have speech abilities yet lack the human body parts necessary for those abilities.

Another form of justification strategy uses empirical observation, referring to children's focus on CAs' behaviors and physical and biological features. This justification strategy echoes Rucker and Pinkwart's assertion that children's actual interactions with intelligent artifacts impact the way children construct mental models of the aliveness status of such objects (2016). We expect that the increase in children's experiences with CAs may lead to more nuanced views (Bernstein & Crowley, 2008).

Lastly, some children relied on the perceived domain of CAs and used such perception as a premise to reason their abilities (i.e., domain references and analogical reasoning). They explained that CAs possessed certain abilities because they belonged to a certain animate domain. However, such strategy did not occur frequently in our study. The infrequency of this strategy may be due to CAs' straddling the boundaries between animacy and inanimacy, thus creating difficulties for children to firmly associate them with either domain in the first place (Jipson et al., 2016).

Tentative Age Trend

Although not one of the research questions, we noticed a possible association between age and children's identification of domain membership. The oldest children in our study (i.e., 6-year-olds) all described the CA as a technological object in the interview and in the drawings, while the answers from younger children (i.e., 3- to 5-year-olds) were mixed. This may be due to the older children's more advanced understanding of programmable machines. Younger children have less awareness of this concept and tend to make sense of computational objects by personifying them (Cameron et al., 2015). Druga et al. provided evidence for this hypothesis in their study of children who observed a robot solving a maze problem (2018). One third of the children between the ages of four to seven credited the robot's successfully solving the problem to its innate cognitive capability, while none of the children aged eight to ten did. The latter group was much more likely to think that the robot was programmed to perform such strategies. However, the children's age range within each group (i.e., 4-7 years and 8-10 years) was wide, which may have obscured developments occurring within each group of children, particular the younger group. Future studies, if carefully structured, may produce more nuanced findings regarding young children's development of perceptions about CAs.

Design Implications

A number of CAs are being developed to provide young children with learning opportunities or social companionship. Findings from our study may help improve the design of such CAs in two ways. First, children's recognition of the CA's cognitive capabilities is an encouraging sign, as children have been found to selectively seek information and learn from those they believe to be intelligent (Harris & Corriveau, 2011). As demonstrated in Breazeal et al. (2016), preschoolers are more willing to trust the information from smarter robots that can provide contingent responses. These children remembered more information from and talked more with the contingent robot than with a non-contingent robot. As such, CAs should be best designed so as to elicit children's attribution of cognitive abilities. In our study, we identified some reoccurring communication techniques that children recognized as a sign of being smart, including the repetition and fallback mechanism we programmed. Children commonly commented that the CA was able to remember and understand because it "repeats" what they said; they also said that the CA always responds to them (even in the case when the CA actually failed to understand). These two techniques both amplified the CA's role as an active interlocutor that is capable of engaging in contingent interactions. Developers may consider incorporating these two communication techniques.

Second, as compared to children's overwhelming recognition of CAs' cognitive properties, children's attitudes regarding whether CAs are psychological entities were mixed. This challenges researchers to develop CAs that children are more willing to engage with socially (Borenstein & Pearson, 2013; Breazeal, 2009). Even though the disembodiment of CAs such as Alexa and Google Assistant may prevent them from leveraging the full range of psychological cues (i.e., facial expressions and body language) (Pelachaud et al., 2010), researchers can compensate for this lack by improving on such CAs' conversational expressiveness (Heerink et al., 2010). CAs may be designed to talk explicitly about their emotions or leverage natural acoustic features (i.e., tone, prosody, speech speed), which may more consistently elicit children's affective reactions and thus may be more likely to lead children to treat CAs as psychological entities. However, while it is important to increase the human-likeness of CAs, we note that children's attribution of human-like qualities to CAs may potentially open children up to undue influence (e.g., misinformation). As such, designers should keep in mind to ensure the content appropriateness of CAs' conversations.

Limitations and Future Work

We noticed three potential limitations during the course of the study. First, our study provided children with opportunities to interact with a CA within a controlled environment. However, children's perceptions may be less about CAs in general and more about the particular CA they interacted with. We addressed this issue by designing the study to cover the typical interactions a child would have with a CA. Second, children's perceptions may be associated with their differing levels of prior experience with conversational technologies. While we did not formally test for this relationship, anecdotal evidence in our study suggests such a relation. For example, when answering a question about whether the CA could listen, one child replied yes and explained that from his previous experience at home, calling Alexa's name would always wake her up. Third, while we suspect a relation between children's overall development and their perceptions of CAs, the small sample size of this study limits our ability to statistically examine this relationship. Future studies should be carried out with larger sample sizes to tackle this issue.

Conclusion

CAs, such as Apple Siri, Google Assistant, and Amazon Alexa, play an increasingly important role in young children's technological landscapes and life worlds. While a handful of studies have documented children's natural interactions with CAs, little is known about children's perceptions of CAs. To fill this gap, we examined three- to six-year-olds' perceptions of CAs' domain membership and properties, as well as their justifications for these perceptions. Overall, these three research questions yielded converging evidence that children sometimes take a more nuanced position and spontaneously attribute both artifact and animate properties to CAs. At least some children appeared unwilling to describe the CA as either a living being or an artifact. These children described the CA as either being a combination of these two categories or fitting into some other third category. Additionally, children appeared to consistently conceive of CAs as possessing a unique constellation of animate properties while lacking others. Almost all of the children in this study ascribed cognitive and speech-related behavioral properties to the CA, while fewer children ascribed psychological and non-speech related behavioral properties. This also reflects children's dilemma in determining CAs' animacy domains. Examination of children's justifications for their perceptions further revealed nuanced reasoning. Taken together, these findings extend current research on children's perceptions of intelligent artifacts by adding CAs as a new genre of study and also provide some underlying knowledge that may guide the development of CAs to support young children's cognitive and social development.

References

Beran, T. N., Ramirez-Serrano, A., Kuzyk, R., Fior, M., & Nugent, S. (2011). Understanding how children understand robots: Perceived animism in child–robot interaction. *International Journal of Human-Computer Studies*, *69*(7–8), 539–550. https://doi.org/10.1016/j.ijhcs.2011.04.003

Bernstein, D., & Crowley, K. (2008). Searching for signs of intelligent life: An investigation of young children's beliefs about robot intelligence. *Journal of the Learning Sciences*, *17*(2), 225–247. https://doi.org/10.1080/10508400801986116

Borenstein, J., & Pearson, Y. (2013). Companion robots and the emotional development of children. *Law, Innovation and Technology*, *5*(2), 172–189. https://doi.org/10.5235/17579961.5.2.172

Breazeal, C. (2009). Role of expressive behaviour for robots that learn from people. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3527–3538. https://doi.org/10.1098/rstb.2009.0157

Breazeal, C., Harris, P. L., DeSteno, D., Kory Westlund, J. M., Dickens, L., & Jeong, S. (2016). Young children treat robots as informants. *Topics in Cognitive Science*, *8*(2), 481–491. https://doi.org/10.1111/tops.12192

Brunick, K. L., Putnam, M. M., McGarry, L. E., Richards, M. N., & Calvert, S. L. (2016). Children's future parasocial relationships with media characters: The age of intelligent characters. *Journal of Children and Media*, *10*(2), 181–190. https://doi.org/10.1080/17482798.2015.1127839

Cameron, D., Fernando, S., Millings, A., Moore, R., Sharkey, A., & Prescott, T. (2015). Children's age influences their perceptions of a humanoid robot as being like a person or machine. *Biomimetic and Biohybrid Systems,* 348–353. https://doi.org/10.1007/978-3-319-22979-9_34

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34. https://doi.org/10.1162/089892998563752

Chan, K. (2006). Exploring children's perceptions of material possessions: A drawing study. *Qualitative Market Research: An International Journal*, *9*(4), 352–366. https://doi.org/10.1108/13522750610689087

Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). Hey Google is it OK if I eat you?: Initial explorations in child-agent interaction. *Proceedings of the 2017 Conference on Interaction Design and Children,* 595–600. https://doi.org/10.1145/3078072.3084330

Druga, S., Williams, R., Park, H. W., & Breazeal, C. (2018). How smart are the smart toys?: Children and parents' agent interaction and intelligence attribution. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 231–240. https://doi.org/10.1145/3202185.3202741

Edwards, S., Nolan, A., Henderson, M., Mantilla, A., Plowman, L., & Skouteris, H. (2018). Young children's everyday concepts of the internet: A platform for cyber-safety education in the early years. *British Journal of Educational Technology*, *49*(1), 45–55. https://doi.org/10.1111/bjet.12529

Gelman, S. A., & Opfer, J. E. (2011). Development of the Animate–Inanimate Distinction. *Blackwell Handbook of Childhood Cognitive Development*, *2*, 213–238. https://doi.org/10.1002/9780470996652.ch7

Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1179–1187. https://doi.org/10.1098/rstb.2010.0321

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Relating conversational expressiveness to social presence and acceptance of an assistive social robot. *Virtual Reality*, *14*(1), 77–84. https://doi.org/10.1007/s10055-009-0142-1

Heintz, K. E., & Wartella, E. A. (2012). Young children's learning from screen media. *Communication Research Trends*, *31*(3), 22.

Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, *49*, 245–250. https://doi.org/10.1016/j.chb.2015.02.026

Jipson, J. L., & Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child Development*, *78*(6), 1675–1688. https://doi.org/10.1111/j.1467-8624.2007.01095.x

Jipson, J. L., Gülgöz, S., & Gelman, S. A. (2016). Parent–child conversations regarding the ontological status of a robotic dog. *Cognitive Development*, *39*, 21–35. https://doi.org/10.1016/j.cogdev.2016.03.001

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. *Handbook of Emotions*, *2*, 220–235.

Kahn, Peter H., Friedman, B., Pérez-Granados, D. R., & Freier, N. G. (2006). Robotic pets in the lives of preschool children. *Interaction Studies*, *7*(3), 405–436. https://doi.org/10.1075/is.7.3.13kah

Kahn, Peter H., Gary, H. E., & Shen, S. (2013). Children's social relationships with current and near-future robots. *Child Development Perspectives*, *7*(1), 32–37. https://doi.org/10.1111/cdep.12011

Kahn, Peter H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., Ruckert, J. H., & Shen, S. (2012). "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental Psychology*, *48*(2), 303–314. https://doi.org/10.1037/a0027033

Kahn, P.H., Freier, N. G., Friedman, B., Severson, R. L., & Feldman, E. N. (2004). Social and moral relationships with robotic others? *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*, 545–550. https://doi.org/10.1109/roman.2004.1374819

Kanero, J., Geçkin, V., Oranç, C., Mamus, E., Küntay, A. C., & Göksun, T. (2018). Social robots for early language learning: Current evidence and future directions. *Child Development Perspectives*, *12*(3), 146–151. https://doi.org/10.1111/cdep.12277

Katayama, N., Katayama, J., Kitazaki, M., & Itakura, S. (2010). Young children's folk knowledge of robots. *Asian Culture and History*, *2*(2), 111. https://doi.org/10.5539/ach.v2n2p111

Kim, M., Yi, S., & Lee, D. (2019). Between living and nonliving: Young children's animacy judgments and reasoning about humanoid robots. *PLOS ONE*, *14*(6), e0216869. https://doi.org/10.1371/journal.pone.0216869

Kim, Y., Baylor, A. L., & Shen, E. (2007). Pedagogical agents as learning companions: The impact of agent emotion and gender. *Journal of Computer Assisted Learning*, *23*(3), 220–234. https://doi.org/10.1111/j.1365-2729.2006.00210.x

Kory-Westlund, J. M., & Breazeal, C. (2019). Assessing children's perceptions and acceptance of a social robot. *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 38–50. https://doi.org/10.1145/3311927.3323143

Lee, S., Kim, S., & Lee, S. (2019). What does your agent look like?: A drawing study to understand users' perceived persona of conversational agent. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290607.3312796

Levy, S. T., & Mioduser, D. (2008). Does it "want" or "was it programmed to..."? Kindergarten children's explanations of an autonomous robot's adaptive functioning. *International Journal of Technology and Design Education*, *18*(4), 337–359. https://doi.org/10.1007/s10798-007-9032-6

Lovato, S. B., Piper, A. M., & Wartella, E. A. (2019). Hey Google, do unicorns exist? *Proceedings of the Interaction Design and Children on ZZZ - IDC '19*, 301–313. https://doi.org/10.1145/3311927.3323150

Luger, E., & Sellen, A. (2016). "Like having a really bad PA": The gulf between user expectation and experience of conversational agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297. https://doi.org/10.1145/2858036.2858288

Marshall, P. J., & Brenneman, K. (2016). Young children's developing understanding of the biological world. *Early Education and Development*, *27*(8), 1103–1108. https://doi.org/10.1080/10409289.2016.1220772

Meichenbaum, D. (2017). Teaching thinking: A cognitive behavioral perspective. *The Evolution of Cognitive Behavior Therapy*, 85–104.

Melson, G. F., Kahn, P. H., Beck, A., Friedman, B., Roberts, T., Garrett, E., & Gill, B. T. (2009). Children's behavior toward and understanding of robotic and living dogs. *Journal of Applied Developmental Psychology*, *30*(2), 92–102. https://doi.org/10.1016/j.appdev.2008.10.011

Mertala, P. (2019). Young children's perceptions of ubiquitous computing and the Internet of Things. *British Journal of Educational Technology*, *51*(1), 84–102. https://doi.org/10.1111/bjet.12821

Mikropoulos, T. A., Misailidi, P., & Bonoti, F. (2003). *Attributing human properties to computer artifacts: Developmental changes in children's understanding of the animate*-inanimate *distinction*.

Pelachaud, C., Gelin, R., Martin, J.-C., & Le, Q. A. (2010). Expressive gesture displayed by a humanoid robot during a storytelling application. *New Frontiers in Human-Robot Interaction (AISB)*.

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). Alexa is my new BFF: Social rules, user satisfaction, and personification of the Amazon Echo. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, 2853–2859. https://doi.org/10.1145/3027063.3053246

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate–inanimate distinction. *Psychological Bulletin*, *127*(2), 209–228. https://doi.org/10.1037/0033-2909.127.2.209

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Rooy, D. L., Pipe, M.-E., & Murray, J. E. (2007). Enhancing children's event recall after long delays. *Applied Cognitive Psychology*, *21*(1), 1–17. https://doi.org/10.1002/acp.1272

Rücker, M. T., & Pinkwart, N. (2016). Review and discussion of children's conceptions of computers. *Journal of Science Education and Technology*, *25*(2), 274–283. https://doi.org/10.1007/s10956-015-9592-2

Scaife, M., & Duuren, M. van. (1995). Do computers have brains? What children believe about intelligent artifacts. *British Journal of Developmental Psychology*, *13*(4), 367–377. https://doi.org/10.1111/j.2044-835x.1995.tb00686.x

Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). Hey Alexa, what's up?: A mixed-methods studies of in-home conversational agent usage. *Proceedings of the 2018 on Designing*

Interactive Systems Conference 2018 - DIS '18, 857–868. https://doi.org/10.1145/3196709.3196772

Serholt, S., & Barendregt, W. (2016). Robots Tutoring Children. *Proceedings of the 9th Nordic Conference on Human-Computer Interaction - NordiCHI '16*, 64. https://doi.org/10.1145/2971485.2971536

Severson, R. L., & Carlson, S. M. (2010). Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category. *Neural Networks*, *23*(8–9), 1099–1103. https://doi.org/10.1016/j.neunet.2010.08.014

Swain, J. (2018). A hybrid approach to thematic analysis in qualitative research: Using a practical example. *Sage Research Methods*. https://doi.org/10.4135/9781526435477

Tewari, A., & Canny, J. (2014). What did spot hide?: A question-answering game for preschool children. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*, 1807–1816. https://doi.org/10.1145/2556288.2557205

Tremoulet, P. D., & Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, *29*(8), 943–951. https://doi.org/10.1068/p3101

van Duuren, M., & Scaife, M. (1996). "Because a robot's brain hasn't got a brain, it just controls itself" — Children's attributions of brain related behaviour to intelligent artefacts. *European Journal of Psychology of Education*, *11*(4), 365–376. https://doi.org/10.1007/bf03173278

Weiss, A., Wurhofer, D., & Tscheligi, M. (2009). "I love this dog"—children's emotional attachment to the Robotic dog AIBO. *International Journal of Social Robotics*, *1*(3), 243–248. https://doi.org/10.1007/s12369-009-0024-4

Wiltshire, T. J., Lobato, E. J. C., Garcia, D. R., Fiore, S. M., Jentsch, F. G., Huang, W. H., & Axelrod, B. (2015). Effects of robotic social cues on interpersonal attributions and assessments of robot interaction behaviors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 801–805. https://doi.org/10.1177/1541931215591245

Woodward, J., McFadden, Z., Shiver, N., Ben-hayon, A., Yip, J. C., & Anthony, L. (2018). Using co-design to examine how children conceptualize intelligent interfaces. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 575. https://doi.org/10.1145/3173574.3174149

Wright, K., Poulin-Dubois, D., & Kelley, E. (2015). The animate-inanimate distinction in preschool children. *British Journal of Developmental Psychology*, *33*(1), 73–91. https://doi.org/10.1111/bjdp.12068

Xu, Y., & Warschauer, M. (2019). Young children's reading and learning with conversational agents. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, CS10. https://doi.org/10.1145/3290607.3299035

Table 5.1

Children's Domain Membership Categorization of the CA in Interview and Drawing

Interview	Drawing	Counts
artifacts	artifacts	13
	residual	3
residual category	residual	6
	artifacts	1
living beings	living beings	5

Table 5.2
Children's Attribution of Cognitive, Psychological, and Behavioral Properties of the CA

	Yes	No	I don't know
Cognitive			
Smart	26 (92.8%)	1 (3.6%)	1 (3.6%)
Remember	24 (85.7%)	1 (3.6%)	3 (10.7%)
Psychological			
Like	18 (64.2%)	6 (21.4%)	4 (14.3%)
Emote	19 (67.8%)	6 (21.4%)	3 (10.7%)
Behavioral			
See	7 (25.0%)	21 (75.0%)	0%
Listen	25 (89.3%)	3 (10.7%)	0%
Talk	26 (92.8%)	1 (3.6%)	1 (3.6%)

Table 5.3

Children's Justifications of the CA's Properties. Bolded Numbers Indicate Salient Justification Patterns within Each Property

	Total	Cognitive		Psychological		Behavioral		
		Smart	Rmb.	Like	Emote	See	Listen	Talk
Domain references	5	0	2	0	0	2	1	0
Analogical reasoning	13	1	3	4	3	0	1	1
Biological references	27	1	1	0	0	***13***	***6***	***6***
Physical references	10	0	0	0	0	4	2	4
Mental state references	21	4	4	6	6	1	0	0
Behavioral references	28	***12***	***7***	2	0	1	4	2
Reciprocity	24	0	5	***8***	***9***	0	2	0
Mechanical references	17	3	0	0	0	0	***6***	***8***
Fantasy reasoning	11	0	0	0	0	***3***	***5***	***3***

Figure 5.1

Drawings That Illustrate Google Home Mini as Artifacts

The picture can't be displayed.

Figure 5.2

Drawings That Illustrate Google Home Mini as Living Objects

The picture can't be displayed.

Figure 5.3

Drawings That Illustrate Google Home Mini as a Combination of Artifacts and Living Objects or as Neither Artifacts nor Living Objects

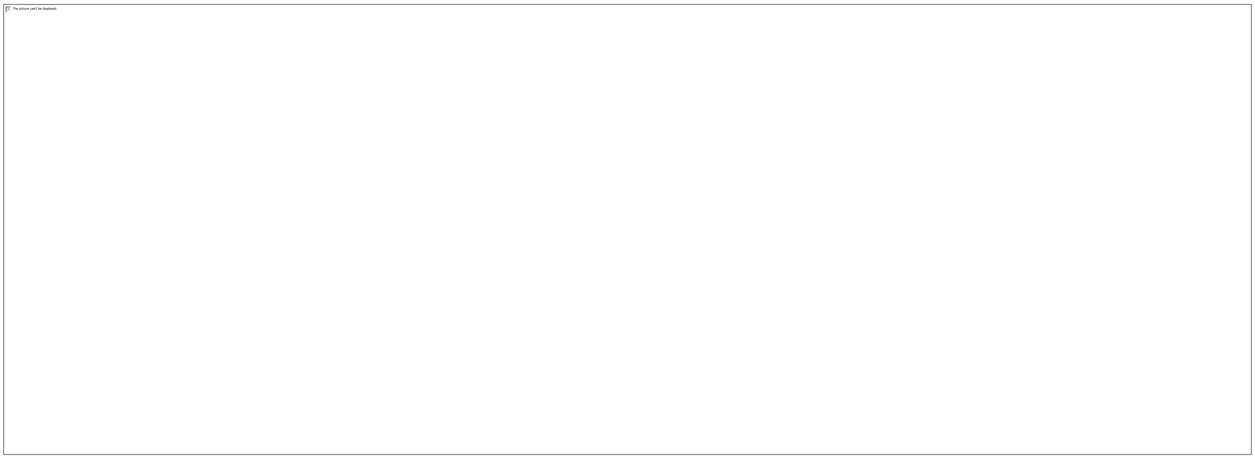


Figure 5.4

Drawing Samples That Contain Cognitive Elements

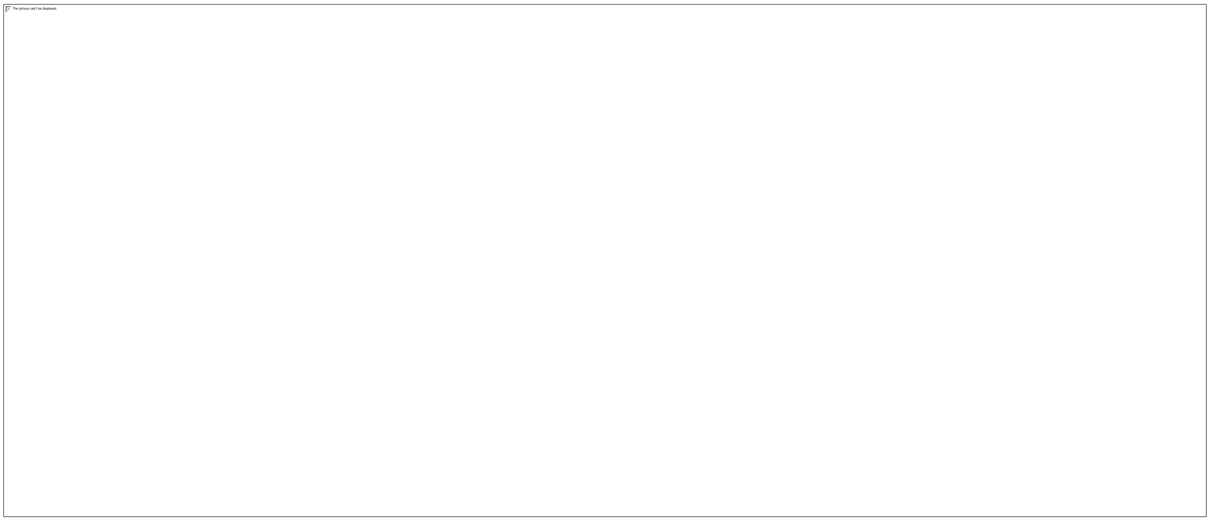


Figure 5.5

Drawing Samples That Contain Psychological Elements

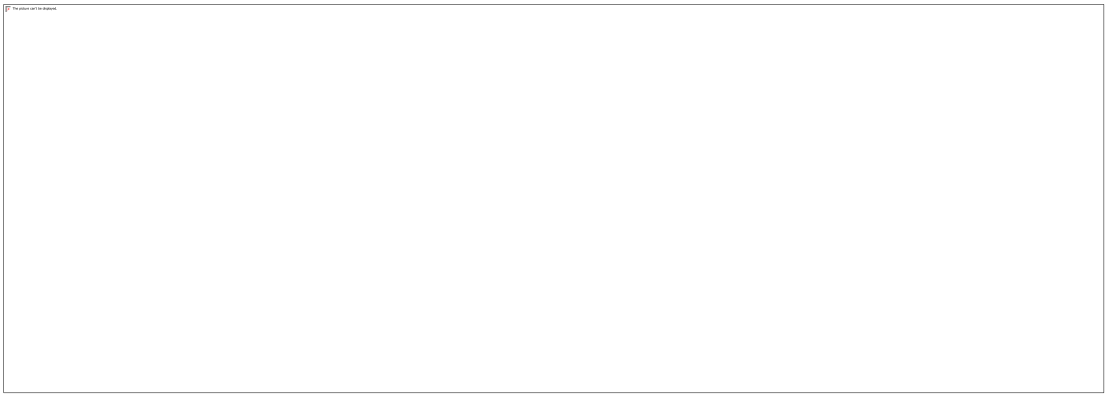


Figure 5.6

A Drawing Sample That Contains Behavioral Elements

The picture can't be displayed.

CHAPTER 6 CONCLUSION

This dissertation reveals how children learn from, respond to, interact with, and perceive conversational agent as their learning companion during shared storybook reading. The first study focuses on ***how children learn from*** the agent. I found that contingent, structured dialogue with the conversational agent led to children's enhanced story comprehension. Such benefit was largely driven by children's heightened level of vocalizations related to the story narratives and reduced off-topic vocalizations. The second study primarily focuses on how ***children respond to*** the conversational agent. I found that conversational agents promoted children's response intelligibility, while adults elicited longer, more lexically diverse, and more relevant responses. The differences in language productivity were amplified among the questions requiring high cognitive demand. The third study focuses on ***how children interact with*** the agent verbally and non-verbally. I found that children generally participated in the conversation with the agent smoothly: they generated on-topic responses and answered within the proper time frame. The result also confirmed the advantage of using a combination of open-ended questions as initial prompts to encourage children's free expression and multiple-choice questions as follow-up prompts to help ease the potential cognitive obstacles. Such scaffolding mechanisms appeared to benefit younger children more so than older ones. The fourth study focuses ***how children perceive*** the agent. I found that children in general held positive perceptions in terms of conversational agents' cognitive and psychological capabilities. Children's such perceptions establish the feasibility of developing agents to socially engage children in learning activities. Overall, the four studies provide converging evidence on the promise of leveraging AI-powered conversational technologies to support young children's language development. The findings are intended to be generalized to designing socially interactive environments for different learning domains (e.g., science) and learning scenarios (e.g., television watching), with a goal of promoting children's long-term development. Below, I discuss three future directions that can expand this dissertation to a broader context.

First, past research consistently suggests that parent-child dialogic interaction stemming from co-engagement with narrative, such as in joint-book reading, can contribute to children's language development. It is believed that dialogic interactions may support children's learning from videos through the same mechanisms as joint reading (Strouse et al., 2013). As such, conversational agents may serve as an effective dialogic partner not only during storybook reading, but also during television or video watching. Indeed, one of my on-going studies aims to develop and evaluate "conversational videos" in which the main character of a television show asks children questions and provides contingent feedback (Xu & Warschauer, 2020). Nevertheless, conversational agents can also be integrated in many other media, such as digital books and plush toys. This can potentially provide ubiquitous learning opportunities for young children as they engage and dialogue with agents on a variety of platforms.

Second, there is a lot of evidence that children growing up in low socioeconomic status (SES) households enter kindergarten with disadvantages in language and literacy development compared to children of higher SES communities. Conversational agents that are low cost may be a viable means to support the language development of low-SES children. While the current dissertation only included children from a well-educated middle class community, I expect that the potential impact of conversational agents has increased relevance for at-risk children, given that these children may be in greater need of educational support than their high SES counterparts (Morgan et al., 2016). Indeed, the conversational video study mentioned above was exactly designed for this goal. The study was carried out primarily with low-income Hispanic children who are the largest and fastest growing minority population in the United States.

Third, the current study has demonstrated that conversational agents can facilitate learning through conversing with children individually. However, this study is not intended to develop agents that supplant the role of parents in reading to their children. Rather, it paves the way towards a new computing paradigm of "Human-AI Collaboration" (Grudin, 2017; Wang et al., 2019) where conversational agents (or other AI systems) serve as a collaborator for caregivers or teachers to support more involved parental guidance during storybook reading or other learning activities. In fact, a recent study has suggested that conversational agents like smart speakers help augment parent-child interaction as a third-party mediator (Beneteau et al., 2020). Future studies may examine the feasibility of using conversational agents as a training system to model to parents beneficial strategies of guided conversation or of deploying prompts to include parents in the conversation. In addition, future systems may consider how parents, conversational agents, and children form a "conversation triad", where parents and conversational agents collaboratively engage children in discussions during learning processes. This could amplify conversational agents' potential by mobilizing other elements in family contexts that work together to support children's language development.

Taken together, I am optimistic that conversational agents enabled by the rapid development of artificial intelligence can support socially productive early learning. Rather than fearing that such technologies may impede children's fruitful face-to-face interpersonal interactions, I believe that conversational agents will provide additional, and unique, interaction and learning opportunities, complementing children's everyday language experiences. Nevertheless, building conversational agents for young children is a complex endeavor. To make conversational agents truly beneficial, it is vital to consider children's still developing cognitive abilities and specific communication needs. By learning from the well-established research in learning sciences, researchers, developers, and educators will be in a position to take a proactive, theory-driven approach to the development and evaluation of conversational agents as children's social learning partners.

References

Beneteau, E., Boone, A., Wu, Y., Kientz, J. A., Yip, J., & Hiniker, A. (2020). Parenting with Alexa: Exploring the introduction of smart speakers on family dynamics. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher*, *45*(1), 18-35.

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child development*, *90*(3), 985-992.

Grudin, J. (2017). From tool to partner: The evolution of human-computer interaction. *Synthesis Lectures on Human-Centered Informatics*, *10*(1), i–183.

Xu, Y., & Warschauer, M. (2020). "Elinor is talking to me on the screen!" Integrating conversational agents into children's television programming. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. April 25-30, 2020, Honolulu, HI. ACM.

Wang, D., Weisz, J. D., Muller, M., Ram, P., Geyer, W., Dugan, C., ... & Gray, A. (2019). Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of

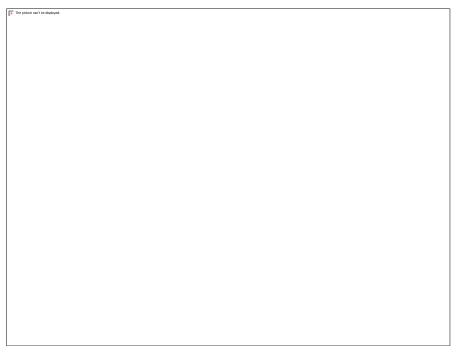
Automated AI. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), 1-24.

Appendix A

Story Comprehension

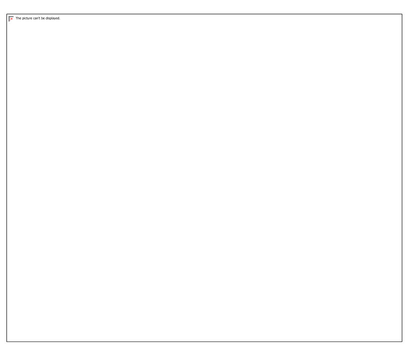
1. The bears broke their mother's beautiful blue seashell when they tried to get the honey. How did they plan to get out of trouble?
   CUE: Alright! I'm gonna give you three different options:
   Did they plan to find another seashell to replace the broken one? Did they run away and hid from their Mama, OR did they just wait at home for their Mama to return?
   1. Find another seashell to replace the broken one
   2. Run away and hide from their Mama
   3. Wait at home for their Mama to return.

2. How did the bears feel when they first began their boat ride to find a new seashell? **(show picture Q2)**



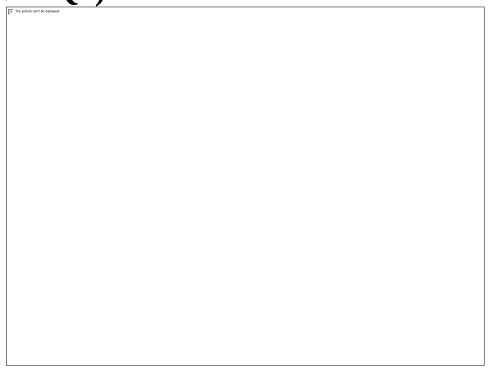
   CUE: I'm gonna give you three different options:
   Did they feel excited, or scared, or confused?
   1. Excited
   2. Scared
   3. Confused

3. Why did the bears believe that they could find the blue seashell on an island shaped like a lumpy hat? **(show picture Q3)**



   CUE: I'm gonna give you three different options:
   1. Because their Mama told them.
   2. Because an old salty bear told them.
   3. Because they saw a blue seashell there before.

4. The bears sailed past an island that looked like a lot of fun. Why didn't they stop at that island to have some fun? **(show picture Q4)**

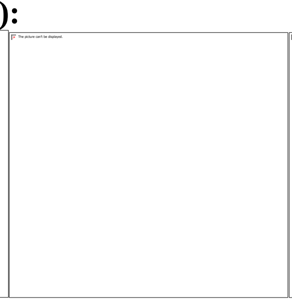
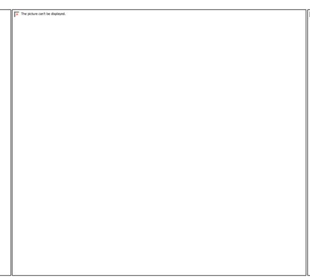
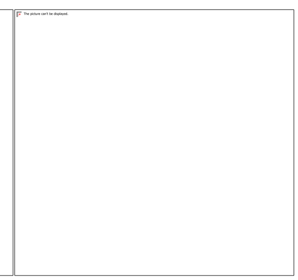


CUE: I'm gonna give you three different options:

1. Because they didn't like riding ferris wheels.
2. Because they wanted to keep searching for the seashell.
3. Because there were too many bears already on the island.

5. The bears stopped at an island shaped like a lumpy hat and searched for the seashell. Do you remember what places the bears searched on that island?

CUE (show pictures):

6. The bears couldn't find the seashell on the island. On their way back home, why didn't the bears notice that a storm was coming?

CUE: I'm gonna give you three different options:

Is it because they were playing, or because they were arguing, or because they were sleeping?

a. They were playing.
b. They were arguing.
c. They were sleeping.

7. Where did they end up finding the seashell?

CUE: I'm gonna give you three different options:

a. On an island that looks like a lumpy hat
b. On an island very far away from their home
c. On the shore of their own island

8. How did Mama bear feel when the bears returned?

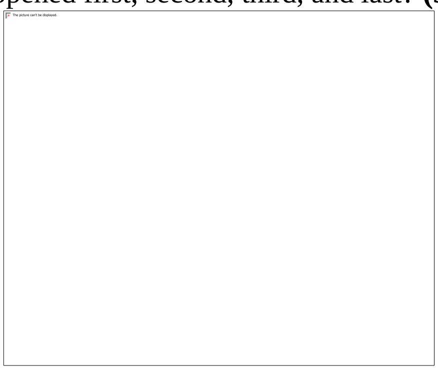
CUE: I'm gonna give you three different options:

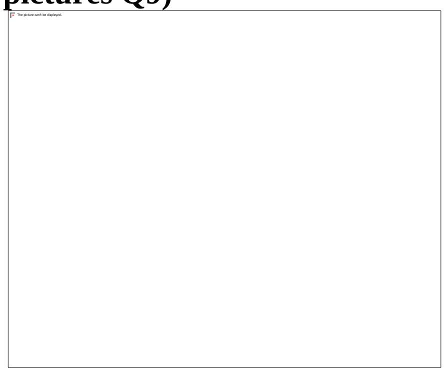
Did Mama bear forgive them? Was she sad? Or was she surprised?

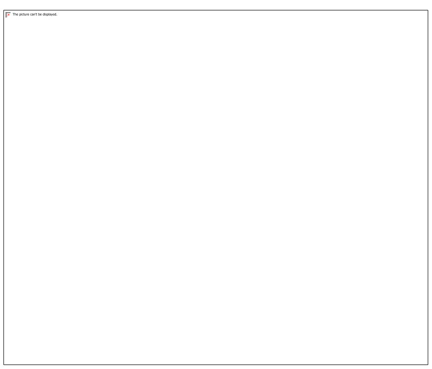
   a. Forgiving
   b. Sad
   c. Surprised

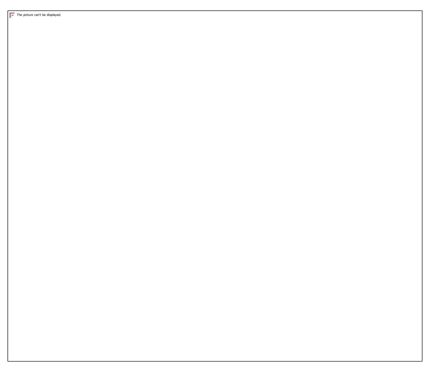
9. Why didn't Mama bear give the bears dessert after dinner?
   CUE: Alright! I'm gonna give you three different options:
   a. Because they did not finish their dinner that Mama gave them
   b. Because they fought with each other on the boat on their way back home
   c. Because the bears tried to secretly eat the honey when their Mama was away and broke their Mama's seashell

10. I'm going to give you 4 pictures of scenes from the story. Put them in order for me. What happened first, second, third, and last? **(show pictures Q9)**









Appendix B

Perception Measures for CA Reading Partner

1. General description
    1) Who were you talking to?/who was reading the story to you?
    2) What is XX (repeat the child's answer)?
    3) Is it a real person?
        a. Yes/no -> why

2. Cognitive attributes
    1) Do you think Google is smart?
        a. Yes/no -> why
    2) Do you think Google is a good listener?
        a. Yes/no -> why
    3) Do you think Google still remembers how old you are and what your favorite color is?
        a. Yes/no -> why

3. Psychological attributes
    1) Can Google like others as a friend?
        a. Yes/no -> why
    2) Can Google feel sad?
        a. Yes/no -> why

4. Behavioral attributes
    1) Can Google see you?
        a. Yes/no -> why
    2) Can Google hear you?
        a. Yes/no -> why
    3) How can it talk?
        a. Yes/no -> why