

EVALUATING THE REPLICABILITY AND
SPECIFICITY OF EVIDENCE FOR
NATURAL PEDAGOGY THEORY



Priya Silverstein, B.Sc., M.Sc.

This thesis is submitted

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

May 2020

Department of Psychology

Declaration

I hereby declare that the work presented in this thesis has not been submitted, in whole or in part, for the award of a higher degree at this or any other university. I further declare that this thesis is a product of my own work and the intellectual content of this thesis reflects my own thinking. All experimental studies included in this thesis were completed under the supervision of Professor Gert Westermann and Dr Eugenio Parise, with Study 1 being completed under the further supervision of Dr. Teodora Gliga, and Study 3 being completed under the further supervision of Dr. Katherine Twomey. In addition, Study 2 was conducted in collaboration with my colleague Marina Bazhydai. To reflect the collaborative effort of the empirical work and the collective theoretical, methodological and stylistic contributions, I will therefore present this work in the first-person plural voice.

The experimental studies comprising this thesis were submitted to and published in the following academic journals:

Silverstein, P., Gliga, T., Westermann, G., & Parise, E. (2019). Probing communication-induced memory biases in preverbal infants: Two replication attempts of Yoon, Johnson and Csibra (2008). *Infant Behavior and Development*. doi:10.1016/j.infbeh.2019.03.005


Bazhydai, M.*, Silverstein, P.*, Parise, E., & Westermann, G. (2020). Two-year old children preferentially transmit simple actions but not pedagogically demonstrated actions. *Developmental Science*. doi:10.1111/desc.12941

(*equal contributions)

Silverstein, P., Westermann, G., Parise, E., & Twomey, K. (2019). New evidence for learning-based accounts of gaze following: Testing a robotic prediction. *Proceedings of the 2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob) Oslo, Norway, August 19-22, 2019*

Silverstein, P., Feng, J., Westermann, G., Parise, E., Twomey, K. (under review). Infants learn to follow gaze in stages: Evidence confirming a robotic prediction. *Open Mind*.

Priya Silverstein, B.Sc., M.Sc.
Lancaster University, UK

A handwritten signature in black ink, appearing to read 'P. Silverstein', with a stylized, flowing script.

Statement of Authorship

Study 1: Probing communication-induced memory biases in preverbal infants: Two replication attempts of Yoon, Johnson and Csibra (2008)

Conception and design of study: Priya Silverstein, Teodora Gliga, Gert Westermann, Eugenio Parise

Acquisition of data: Priya Silverstein

Data analysis: Priya Silverstein

Data interpretation: Priya Silverstein, Teodora Gliga, Gert Westermann, Eugenio Parise

Drafting the manuscript: Priya Silverstein

Revising the manuscript: Priya Silverstein, Teodora Gliga, Gert Westermann, Eugenio Parise

Contribution of principal author: 85%

Study 2: Two-year old children preferentially transmit simple actions but not pedagogically demonstrated actions

Conception and design of study: Marina Bazhydai, Priya Silverstein, Gert Westermann, Eugenio Parise

Acquisition of data: Marina Bazhydai and Priya Silverstein

Data analysis: Marina Bazhydai and Priya Silverstein

Data interpretation: Marina Bazhydai, Priya Silverstein, Gert Westermann, Eugenio Parise

Drafting the manuscript: Marina Bazhydai and Priya Silverstein

Revising the manuscript: Marina Bazhydai, Priya Silverstein, Eugenio Parise, Gert Westermann

Shared contribution of principal authors (Marina Bazhydai and Priya Silverstein): 85%

Study 3: Infants learn to follow gaze in stages: Evidence confirming a robotic prediction

Conception and design of study: Priya Silverstein, Katherine Twomey, Gert Westermann, Eugenio Parise

Acquisition of data: Priya Silverstein, Jinzhi Feng

Data analysis: Priya Silverstein

Data interpretation: Priya Silverstein, Katherine Twomey, Gert Westermann, Eugenio Parise


Drafting the manuscript: Priya Silverstein

Revising the manuscript: Priya Silverstein, Katherine Twomey, Gert Westermann, Eugenio Parise

Contribution of principal author: 85%


We the undersigned agree with the above stated “proportion of work undertaken” for each of the above chapter contributing to this thesis:


Priya Silverstein: 

Marina Bazhydai: 

Jinzhi Feng: 

Teodora Gliga: 

Katherine Twomey: 

Gert Westermann: 

Eugenio Parise: 

Acknowledgements

Firstly, I would like to thank my supervisors for their excellent supervision and training. Eugenio, thank you for always pushing me past what's comfortable and teaching me how to assert myself. Gert, thank you for providing me with emotional and intellectual support. To both of you, thank you for the uniquely theoretically dichotomised supervision experience. This truly is the biggest gift you gave me, as it forced me to work out where I stood very early on and ensured I didn't make any assumptions about the scientific method. Teea, thank you for your important intellectual contribution to Study 1; I enjoyed working with you and only wish we'd done more studies together. Thank you to my PhD examiners (Dermot and Carina) for their extremely helpful comments on a previous draft of this thesis.

I would like to thank my collaborators for working with me and inspiring me every day. Research is by its nature collaborative, and I can't think of better people to have shared this journey with. Katie, you have been a guardian angel throughout this PhD, from your informal advice and support throughout the first two years, to our work together for Study 3, which is my proudest scientific work and work that I hope to continue to pursue in the future. Marina, I'm so glad you were my PhD twin, starting the same year with the same unique supervision combination. The fact that we came out the other side of Study 2 still friends and enjoying working together speaks volumes, and I look forward to our future collaborations.

I was not expecting to find such a passion for methods and open science when I began my PhD, but I did. I have many people to thank for fostering and encouraging that passion and supporting my efforts. Dermot, thank you for founding Prospr, where my open science fire was first lit, and for your continuing support and encouragement throughout my PhD journey. Beyond this, thank you to the massive (mostly online)

community that has inspired and supported me in my open science endeavours, including, but not limited to: Christina Bergmann, Krista Byers-Heinlein, Chris Chambers, Clare Conry-Murray, Sam Forbes, Mike Frank, Monica Gonzalez-Marquez, Olivia Guest, Hans Rocha IJzerman, and Kirstie Whitaker.

I am so grateful to have done my PhD in such a fantastic Babylab, surrounded by so many great people. Kat, you are the best lab manager in the world. Thank you for doing such a great job of supporting everyone who enters the babylab, and for always making me laugh and giving me pep talks when I was low. Thank you to all of my babylab colleagues for being such lovely people. Christian, thank you for being a great mentor, allowing me to vent, and helping with filming stimuli. Anna, thank you for being my amazing actress for Study 1 and 3, for saving the day when the trains were cancelled, and for procrasti-eating questionable snacks with me. Shirly, thanks for always making me smile.

In terms of keeping me sane, thank you to my counsellor, my gym, and Cafuné café in Preston for ensuring self-care was a priority. Thank you to the people in my life who always cheered me up, including but not limited to Beth, Charlotte, Christian, Christina, Giada, Jenny, Margo, Maya, Olivia, & Sienna. Thank you to my mum for always being proud of me, and for paying my MSc fees. I am aware of how much of a privilege this was, and how I probably wouldn't have got a funded PhD place without this. Thank you to my sister Amy for being my academic inspiration to fight for what you believe in even if it's harder than doing things the way they've always been done – I want to be like you when I grow up. Lastly, thank you to my wife Morna for always believing in me and supporting me, forcing me to take time off, and putting up with the absolute monster that I became during this PhD. I owe you one.

Preface

Communication is one of the most inherent features of what it is to be a living organism. Bacteria use quorum sensing to communicate (Kaiser & Losick, 1993). Plants use volatiles to warn neighbouring plants when under attack (Baldwin & Schultz, 1983). The anthropomorphised way we talk about communication even at this level seems uncomfortable, evoking thoughts of intentionality. Of course, plants do not “warn” neighbouring plants about an incoming attack in the same way we do as humans. They are not worried about the safety of plants they have lived alongside for years, hoping the ones they care about the most will make it out alive. But they do convey information from one entity to another – and so they do indeed communicate.

When we turn to thinking about animals, especially when they communicate in ways we are more familiar with, the question of intentionality becomes a little more blurred. For example, Campbell monkeys use different alarm calls as referential labels to differentiate between different types of predator (Zuberbühler, 2001). It is extremely difficult to measure, or investigate, intentionality. As human adults, we are (seemingly) aware of our own intentions. However, this could equally often be an illusion whereby we act and then justify our actions as we go. When we are asked why we did something, it is easy to answer this question by answering, “Why would someone do that”, rather than genuinely introspectively reflecting on the thoughts that supposedly led to that action. We may indeed do this more often than we realise, and overestimate the role that consciousness plays in our behaviour (Bear & Bloom, 2016). But what does this have to do with communication? The study of what makes human communication unique often comes down to the question of intentionality. Plants warning their neighbours about danger do not *intend* to warn them, and the plants receiving the information do not assume or understand any intention from their

neighbour plant. Although it is clear that adult humans do have intentions to communicate, what is less clear is the developmental trajectory of this behaviour, i.e. whether it is necessary to have and understand intentions in order to be receptive to communication, or whether the intentionality itself is an emergent property of communication. If we consider infants interacting with adults, there is an imbalance of intentions: although caregivers intend to communicate with their infants, it is unclear how much of this intention is understood by the infants, and how much is processed instead through statistical learning of the correspondences between referential cues (like words, pointing, and object-directed gaze) and entities. The field of infancy research is divided, as ever, along the predictable line of “nature vs. nurture”. This thesis was motivated by the desire to contribute towards the resolution of this debate, through both empirical investigation and advocating for a methodological reform that is already beginning to reshape the field of infancy research.

Abstract

Do infants understand that they are being communicated to? This thesis first outlines issues facing the field of infancy research that affect confidence in the literature on this (and any) topic to date. Following this, an introductory chapter evaluates evidence for the three core claims of Natural Pedagogy (NP), and the compatibility of this evidence with alternative theories. This is followed by three experimental chapters. In Study 1, we attempted two replications of the study with the highest theoretical value for NP (Yoon et al., 2008). This study has high stakes theoretically, as it is the only study providing evidence for the most specific claim of NP that is difficult to explain by low-level mechanisms. Therefore, a replication of this result that included a reduction of possible confounds and a more sophisticated measure of attention throughout the task was of great theoretical value. In this study, we were unable to replicate the original findings. In Study 2 we went beyond the evidence for the claims made in the outline of NP, and instead generated a new, specific prediction that we believe NP would make. This is important, as theories are only useful if they can make clear, testable predictions. In this study, we pitted pedagogically demonstrated actions and simple actions against each other and evaluated infants' transmission of these actions to someone else. We found no evidence for NP, finding evidence for preferential transmission of simple actions instead. In Study 3 we went beyond NP, and tested a clear prediction stemming from an alternative low-level theory for how infants develop gaze-following ability. We found evidence that infants learn to gaze-follow through reinforcement. Overall, this thesis contributes to the vast literature on infants as recipients of communication, as well as highlighting methods for conducting open and reproducible infancy research.

List of Figures and Tables

List of Figures

Study 1

- 1.1 Figure 1: Flow chart of stimuli in the test trials.....48
- 1.2 Figure 2: Experiment 1: Mean duration of first look with SE bars for each outcome grouped by communicative context.....52
- 1.3 Figure 3: Experiment 1: Mean duration of total look with SE bars for each outcome grouped by communicative context.....53
- 1.4 Figure 4: Proportion looking to face (top) and hand (bottom) plotted over time (ms). N=40.....55
- 1.5 Figure 5: Experiment 2: Mean duration of total look with SE bars for each outcome grouped by communicative context.....62

Study 2

- 1.1 Figure 1: a) Examples of stimuli (toys used in Experiment 2); b) Schematic of the experimental setup; c) Experimental procedure.....80
- 1.2 Figure 2: a) Experiment 1: distribution of first responses at the transmission phase across the two trials; b) Experiment 2: distribution of first responses at the transmission phase across the two trial.....85
- 1.3 Figure 3: Average count of different types of actions shown at transmission phase during both experiments, including pedagogically demonstrated actions and non-pedagogically demonstrated actions.....92

Study 3

- 1.1 Figure 1: Experimental and control conditions. Central cue (direct gaze or fixation cross) for 1000 milliseconds, followed by a directional cue (averted gaze or

object) for 2000 milliseconds.....	109
1.2 Figure 2: Raincloud plot showing proportion target looking across conditions (gaze vs. control) and planes (horizontal vs. vertical) for 12-month-olds.....	112
1.3 Figure 3: Raincloud plot showing proportion target looking across conditions (gaze vs. control) and planes (horizontal vs. vertical) for 6-month-olds.....	115
1.4 Figure 4: Raincloud plot showing proportion target looking across conditions (gaze vs. control) and planes (horizontal vs. vertical) for 18-month-olds.....	118

List of Tables

Study 1

1.1 Table 1: Studies Investigating Communicatively Induced Memory Biases.....	66
---	----

Table of contents

The Replication Crisis in Infancy Research.....	1
Introduction	7
1. Infant sensitivity to ostensive signals	8
1.1 Direct gaze.....	9
1.2 Infant-directed-speech	12
1.3 Combination of cues.....	16
2. Infant referential expectation in ostensive contexts	17
2.1 Infant following of deictic gestures.....	17
2.2 Additional evidence for infant referential expectation in ostensive contexts	29
3. Infant bias to interpret ostensive-referential communication as conveying information that is kind-relevant and generalisable	31
3.1 Infants encode valence of objects over preference of adults in ostensive- referential contexts	32
3.2 Infants encode identity of objects over location in ostensive-referential contexts.....	33
3.3 Infants do not commit the A not B error in non-ostensive-referential contexts	35
4 General conclusions.....	36
Study 1: Probing communication-induced memory biases in preverbal infants: Two replication attempts of Yoon, Johnson and Csibra (2008)	38
Linking statement	38
Abstract.....	41

Introduction	41
Experiment 1	45
Method.....	45
Results	50
Discussion.....	56
Experiment 2	58
Method.....	59
Results	61
Discussion.....	63
General discussion.....	63
Study 2: Two-year old children preferentially transmit simple actions but not pedagogically demonstrated actions.....	68
Linking statement.....	68
Abstract.....	71
Introduction	72
Experiment 1	79
Method.....	79
Results	84
Experiment 1 discussion.....	87
Experiment 2	88
Method.....	88
Results	90
Experiment 2 discussion.....	93

General discussion.....	95
Study 3: Infants learn to follow gaze in stages: Evidence confirming a robotic prediction.....	100
Linking statement.....	100
Abstract.....	102
Introduction	102
Experiment 1 (12-month-olds).....	106
Method.....	107
Results	111
Discussion.....	112
Experiment 2 (6-month-olds).....	114
Method.....	114
Results	114
Discussion.....	116
Experiment 3 (18-month-olds).....	116
Method.....	116
Results	117
Discussion.....	118
General discussion.....	118
Conclusion.....	121
General discussion.....	122
1. Summary of findings	122
2. Replication.....	125

2.1 ManyBabies	127
3. Future directions	129
3.1 Study extensions	129
3.2 Theory development	132
References:	134

The Replication Crisis in Infancy Research

It is no secret that there is a reproducibility crisis in psychology. It is debated whether it should be called a reproducibility crisis or revolution, but what is clear is that up to 50% of published research findings are false positives (Camerer et al., 2018; Klein et al., 2018; Open Science Collaboration, 2015), due in part to papers with significant results (i.e. p values less than 0.05) being much easier to publish (Fanelli, 2010; Forstmeier et al., 2017). Furthermore, even when “quality” rather than “quantity” of publications is relied on as a metric for scientific contribution, studies in the most “high impact” journals often have the lowest reliability (Brembs, 2018). The impact of this is that currently, many published research findings are not replicable (Camerer et al., 2018; Klein et al., 2018; Open Science Collaboration, 2015). One reason for this state of affairs is the “garden of forking paths”, where decisions are made at every stage of the research project to encourage a p value of less than 0.05 (John et al., 2012; Wicherts et al., 2016). Decisions like which participants are included, which dependent variables are reported, which conditions are reported, etc., are made (often with the best intentions) in the direction of favouring the smallest p values (p -hacking, Simmons et al., 2011) and hypothesising after the results are known (HARKing; Kerr, 1998) to create as clear a “story” as possible for the resulting paper.

Although these problems are prevalent across psychology as a whole, and although a lot of attention regarding the reproducibility crisis has been on social psychology (Świątkowski & Dompnier, 2017), infancy research is likely to be one of the most heavily impacted by this crisis (Davis-Kean & Ellis, 2019; Frank et al., 2017). This is for a variety of reasons related to methodological and analytical

decisions. We can consider two comparable paradigms with infants and adults in order to understand how questionable research practices (QRPs; John et al., 2012) come into play in infancy research without any malicious intent. For the purposes of this explanation, we can compare an adult reaction time study with an infant looking time study.

Until very recently, it has been considered normal in infancy to base sample size decisions on previously published papers (e.g. Study 1 and 2 of this thesis), rather than any form of power analysis. This has resulted in many low N studies, making false positives more likely (Button et al., 2013). Using a small sample size is understandable, considering how costly it is to run infants in even simple behavioural experiments. In terms of time cost, in an adult reaction time study, testing can perhaps take fifteen minutes, with participants being scheduled either consecutively every twenty minutes, or in a lab room with multiple participants taking part at the same time, or even online in their own time, with no official testing time for the researcher. When testing an infant, even if the actual testing takes only five minutes, families need to be called in for an hour, in order to go through the consenting procedure in detail, allow for any hiccups regarding the infant being asleep, or needing to be changed or fed. In terms of actual monetary cost, in the adult study adults will be paid for their time – perhaps £2.50 for a fifteen-minute experiment, but often not even this, as undergraduate participants often take part for course credits, i.e. for free from the researcher's point of view. In infant studies, parents often need to be reimbursed for travel expenses (in our lab - £10 per visit), as well as given a gift for participation (usually a book, perhaps worth another £3 if bought in bulk).

In a simple reaction time experiment, it would be unlikely to have many (if any) participants excluded completely. Perhaps one participant came in despite not

fitting some inclusion criterion (e.g. being a native English speaker). However, it is highly possible even in a simple infant looking time experiment to have around 25% exclusion rate because of fussiness, not looking enough at the screen, etc. On a related note, in the adult experiment, with fifteen minutes of reaction time data, there will be maybe hundreds of trials per condition, making it highly likely that even if a participant misses a few from being distracted (less likely in adult participants in the first place), they will still be able to be included in the final sample minus these trials. In the infant experiment however, depending on the design, there may be as few as *one* trial per sub-condition (e.g. Study 1 of this thesis), meaning that if an infant misses one trial, they will need to be excluded from the entire experiment. For these reasons and more, it is expensive and time consuming to run infant studies. In the end, researchers often have to test double or more than double the amount of eventually included infants, and this testing will take much longer than testing the equivalent number of adults. In an extremely well-functioning lab, with intensive testing every day, it can still take around a month to get a final sample of 24 infants, but in other labs with a smaller recruitment base often much longer, sometimes even up to a year.

For this reason, sample sizes are small – meaning the probability of finding a real effect is small, even if it does exist. So, then why don't all experiments with infants result in inconclusive null results? This is where the garden of forking paths comes in. The desire to have something of “publishable” quality after the high expense of conducting the study, combined with the little knowledge we have of how infants actually “work”, means that the researcher can make a million and one decisions about their methodology post-hoc, from inclusion criteria to analysis methods. While the intention when making these decisions is usually to reduce noise and make sure that “bad” infants aren't included in the final sample, this analytical

flexibility (or *p*-hacking) leads to false positive results being reported more often than would be predicted otherwise (Simmons et al., 2011).

When there are the resources available to run multiple versions of the same experiment, tweaking the paradigm until it is “perfect”, this expands the garden of forking paths to before data collection – to the design of the experiment itself. Piloting is important when designing a new paradigm, but there is a fine or invisible line between piloting and “real” data collection (especially when considering small sample sizes). It is possible that the difference between data being considered pilot data vs. “real” data would be in whether or not the desired result was significant or at least pointing in the desired direction. To stop “tweaking” the paradigm at this point and include all participants who took part in the latest version of the paradigm (without restarting data collection after piloting this exact paradigm) means that the resulting effect is highly likely to be a false positive, considering that the criterion for stopping changes to the paradigm was that it showed the effect that was desired. This issue thrives partially due to the lack of standardisation in infancy research (Benasich & Bejar, 1992; Cristia et al., 2016). Unfortunately the other option – guessing extremely specifically which exact outcome measure will show an effect (total looking time or length of first look or proportion looking time) or which exact criteria mean an infant should be included (the ones who cried or the ones who looked like they might cry or the ones who cried a lot) is a guessing game too, and would very likely result in a lot of false negatives.

Luckily, there are many solutions offered by open science (Munafò et al., 2017). Open science is best described as “an umbrella term used to refer to the concepts of openness, transparency, rigour, reproducibility, replicability, and

accumulation of knowledge, which are considered fundamental features of science” (Crüwell et al., 2019, p. 3). Open science enhances knowledge among the scientific community and permits more informed assessments about credibility (Vazire, 2017). One aspect of open science is to be transparent and honest about everything that was conducted in a study, and this is enacted through transparent manuscript writing, reporting all analyses, and sharing data, code, and materials. It is also important to reduce the likelihood of finding a false positive result, by preventing p-hacking and HARKing. To this end, preregistration involves posting a timestamped outline of the research questions, hypotheses, method, and analysis plan for a specific project prior to data collection (Nosek et al., 2018), not preventing further exploration of the data but making clear the difference between confirmatory and exploratory analyses. When preventing false positives, it is likely we will see many more null results. It is important to know whether these null results are indicative of our data showing support for the null hypothesis, or whether instead the results are just inconclusive. Bayes factor analysis goes some way towards distinguishing between these two interpretations of null results (Dienes, 2014, 2016), by being able to determine whether data show support for the experimental *or* null hypothesis, as opposed to frequentist statistics that cannot determine support for the null hypothesis. In addition, normalising replication studies will allow us as a field to work out which studies are replicable and which are not, in order to adjust theory accordingly, avoid multiple failed research pathways due to assumptions based on the current literature, and ultimately will propel us into a more efficient science, closer to truth and further from storytelling. Our manuscripts may be less concise, our stories may be less clear, but our science will not be a vague collection of arbitrary decisions made in order to get our papers published.

In addition to this, large scale replication projects such as ManyBabies (based on the model of ManyLabs), allow for extremely large sample sizes in the thousands for one study, in order to see whether a result is replicable or not (and the expected effect size) and investigate contextual factors about participant demographics and lab protocols that might mediate this effect with enough power to do so (Frank et al., 2017). They do this by having many labs each contribute a sample of infants to the same study, following a very stringent and transparent protocol. In the case of ManyBabies, they also use all the above outlined solutions to increase transparency, including having all materials, protocols, and code openly available.

In this thesis, there are examples of many of these solutions (pre-registration, Bayes Factor Analysis, transparent reporting, and open materials, data and code) which were implemented bit-by-bit throughout the entire PhD program (for a primer on how to implement open science practices as a postgraduate student, see our tutorial paper: [Kathawalla et al., 2020](#)). My story is less clear – I was unable to continue in one vein with my research due to multiple failed replications of studies in the literature that I wished to build on. But I have learnt that the methods for enhancing the replicability and openness of science can be just as interesting as the theoretical questions themselves.

Introduction

There is a dichotomy within infancy research: those who believe that infants understand everything, and those who believe that infants understand nothing. What I mean by this more specifically, is that when infants exhibit a behaviour in our studies (be it a look, specific neural activity, pointing to an object, etc.) this can be interpreted in two ways: infants have some higher-level understanding of the task at hand, or they are purely influenced by low-level cues. For the development of receptivity to communication, this debate is certainly ongoing. Put simply, high-level theories propose that infants are sensitive to the fact that they're being communicated to (Csibra & Gergely, 2009; 2011; Tomasello, 1999; Tomasello et al., 2005), whereas low-level theories propose that infants do not have this receptivity (Anderson, 2013; Butterworth, 1991; Moore & Corkum, 1994; Triesch et al., 2006).

Planted in this debate is one of the prominent theories of communication development in developmental psychology. Natural Pedagogy (NP) provides an evolutionary account for the uniqueness of human communication, rooted in the idea that infants are born receptive to the intentions of their conspecifics. The first of these intentions is the “communicative intention”. Infants are argued to recognise that an adult has the intention to communicate with them, by being sensitive to cues that suggest this, such as infant-directed-speech and direct gaze. The second of these intentions is “referential intention”. Once they recognise the communicative intention of their conspecific, infants are then argued to recognise any following referential cues (such as pointing and referential gaze) as being due to their referential intention, i.e. that the conspecific would like to refer to whatever they are attempting to draw the infant's attention to (Sperber & Wilson, 1986).

NP suggests that humans have evolved the uniquely human ability to allow transmission of generic knowledge through teaching (Csibra & Gergely, 2009; 2011). This communication system means that humans are able to learn cognitively opaque cultural knowledge quickly, the type of knowledge that would be difficult and slow to learn purely through observation. It is suggested that infants are prepared to receive this kind of knowledge through an innate receptivity to natural pedagogy. The arguments for NP are threefold, of increasing specificity: (i) infants are sensitive to ostensive signals, (ii) these ostensive signals trigger a referential expectation in infants, and (iii) infants expect the information that is communicated to them in ostensive-referential contexts to be kind-relevant and generalisable.

When assessing the evidence for the different levels of these claims, it is important to note that all evidence should not be weighted equally. Although ideally in science the accumulating results from several studies in the literature lead to theory refinement (Kuhn, 1962; Popper, 1963), many of the issues surrounding reproducibility outlined in the first chapter of this thesis prevent this (Ioannidis, 2005, 2012). For this reason, rather than assuming the current state of the published literature must be taken for granted and integrated in order to tell a coherent story, it is reasonable to be sceptical about much of the literature to date (Ioannidis, 2005; Smaldino & McElreath, 2016).

1. Infant sensitivity to ostensive signals

Ostensive signals inform the infant that they are going to be the recipient of communication. Ostensive signals include direct gaze, infant-directed-speech (IDS), contingency, and the infant's own name (Gergely Csibra & Gergely, 2009). This thesis will focus specifically on direct gaze and IDS. There is ample evidence of

infants' sensitivity to ostensive signals, however what is more contested is whether this alone can be considered evidence for NP, or whether there are alternative explanations for this sensitivity.

1.1 Direct gaze

Direct gaze (eye contact) is a key ostensive signal in the visual domain. Adults preferentially orient to direct gaze automatically and without awareness (Mares et al., 2016; Rothkirch et al., 2015), and this effect is modulated by how it is shown (symbolic, photo, avatar, live) and the context (duration, interlocutor, purpose, emotion, action, social context) of the cue (Hamilton, 2016). Non-human primates also show a preference for direct gaze (Emery, 2000), sometimes at as young as 1 week old (Mendelson et al., 1982; Myowa-Yamakoshi & Tomonaga, 2001). When perceiving direct gaze, the same neural structures have been shown to be activated in adults (Kampe et al., 2003) and in 4-month-old infants (Grossmann et al., 2008). Infants are sensitive to direct gaze, in that even newborns preferentially orient to direct gaze (Batki et al., 2000; Farroni et al., 2002; 2006, although see Vecera & Johnson, 1995) and show differential neural activation for direct and averted gaze (Urakawa et al., 2015). However, this seems specific to gaze directed at the self, as infants only show sensitivity to mutual gaze between two interlocutors after they are 10 months old (Beier & Spelke, 2012).

There is ample evidence that this preference for direct gaze is highly specific. This newborn preference disappears with an upside down face (Farroni et al., 2006), and this inversion effect only holds for human faces (Itier et al., 2011). The preference for direct gaze is also eliminated when the contrast polarity of the eye is switched, that is, when the sclera of the schematic eye is black, and the iris is white (Farroni et

al., 2005; Jessen & Grossmann, 2014), and this is the case not only in human infants but in face-naïve chicks (Rosa Salva et al., 2012). The fact that the preference is eliminated in this case suggests that this preference is specific to the morphology of the eye, and not merely any directional cue. The fact that this effect of contrast-polarity is seen not only in infants but in face-naïve chicks suggests that this is likely to be an innate preference, although whether it is specific to eyes or presents a broader perceptual bias remains an open question.

Infant sensitivity to direct gaze has also been studied in more naturalistic scenarios. Increased alpha band desynchronization (a neural signal found previously to be elicited in adult-adult joint attention scenarios) has been found when 9-month-olds engage in joint attention about an object with an adult while engaging in eye contact compared to when no eye contact is made (Hoehl et al., 2014). This result was interpreted in the framework of NP, suggesting activation of a generic semantic knowledge system, however this is not the only interpretation that follows from these results, with an alternative explanation being simply increased attention because of the saliency of direct gaze, something not specific to ostensive signals.

1.1.1 Alternative theories for infant sensitivity to direct gaze

If we consider infant sensitivity to direct gaze, although this is necessary for and compatible with NP, there are also several other explanations for this phenomenon. These can be grouped into low-level arousal theories, high-level intention detector theories, cultural pedagogy theory, and two-stage models.

Low-level arousal theories suggest that sensitivity to direct gaze is a non-specific effect on general arousal (Anderson, 2013; Hood et al., 2003; Kawashima et al., 1999). This is supported by several studies finding increased arousal when

perceiving direct gaze (Gale et al., 1975; Helminen et al., 2011; Kleinke & Pohlen, 1971; Nichols & Champness, 1971; Williams & Kleinke, 1993). These theories suggest that sensitivity to direct gaze is not innate, and that postnatal experience determines the reward value of eye contact and therefore how arousal should be mediated accordingly. These theories are at odds with findings of newborn infants' sensitivity to direct gaze. Additionally, these theories could be explaining the resulting effect of a more high-level theory. For example, it makes sense that if infants were to interpret direct gaze as communicative, this would heighten arousal. It is difficult to disentangle these explanations for increased arousal when perceiving direct gaze.

High-level communicative intention detector theories suggest that sensitivity to direct gaze is due to innate detection of intent from another interlocutor (Baron-Cohen, 1995; Perrett & Emery, 1994). Although these theories share something in common with NP, in that they all assume innate detection of intent from another interlocutor, they differ in what this intent is. Whereas NP assumes detection of *communicative* intent, the Eye Direction Detector (EDD) proposed by Baron-Cohen (1995) is proposed to have evolved due to the evolutionary benefit of knowing whether someone is going to try to eat you or seduce you. These occurrences are not necessarily communicative, and so this is a crucial difference. The EDD is an eye-specific mechanism, which is at odds with evidence that different communicative cues are processed similarly (see section 1.3).

Two-stage models suggest that an initial detection of eyes is followed by decoding of the informational content (Grossmann, 2017; Johnson et al., 2015; Senju & Johnson, 2009). These theories suggest that an initial innate bias to orient to eyes interacts with postnatal experience, which then results in the decoding of the

informational content, including the emotional, attentional, and mental states of the interlocutor. Evidence for this comes from the finding that these processes seem to emerge at during different times during infant development. For example, although newborns are sensitive to direct gaze, they do not begin to be comforted by it until 1 month of age (Zeifman et al., 1996), suggesting that over this time infants have learned the contingency between direct gaze from caregivers and being comforted. Further evidence comes from the finding that overlapping but distinct brain areas are implicated in automatic detection of gaze and further inferences about the informational content of the gaze. Specifically, although the amygdala is implicated in automatic detection of gaze (e.g. Adolphs et al., 2005), areas involved in mental state reasoning (the posterior superior temporal sulcus and the medial prefrontal cortex) are involved in responding to these gaze cues (e.g. Baron-Cohen et al., 1999). Although these theories allow for detection of communicative intent at the second stage, NP suggests instead that infants are sensitive to direct gaze *because* of its communicative intention, making them incompatible.

It seems that human infants (and possibly some other animals, e.g. chicks) have an innate sensitivity to direct gaze. The bigger question however is the mechanism underlying this sensitivity, and whether or not there is substantial evidence for the further claims of NP. Importantly, it seems that this sensitivity is not a uniquely human trait, and therefore is not specific evidence of a uniquely human adaptation for cultural knowledge transmission, which is a key feature of NP.

1.2 Infant-directed-speech

IDS is a key ostensive signal in the auditory domain. IDS is a term to describe the characteristic speech that is directed towards infants in many cultures. In terms of

acoustic characteristics compared to adult-directed speech (ADS), IDS has a higher pitch and greater pitch excursions (Fernald, 1989). Infants are sensitive to IDS, in that they prefer to listen to IDS than ADS (for a meta-analysis see Dunst et al., 2012). Newborn infants exhibit a preference for IDS (Cooper & Aslin, 1990). This preference is also reflected in differential neural activation for IDS than ADS (Peter et al., 2016; Zangl & Mills, 2007), even in newborns (Háden et al., 2020), but only for female voices (Sulpizio et al., 2018), and this preference is mediated by the familiarity of the speaker (Naoi et al., 2012). IDS is also associated with heightened cerebral blood flow in sleeping newborns (Saito et al., 2007).

A recent large multi-lab collaboration to investigate the replicability of infant IDS preference found that indeed, this effect was replicable across the world in 3- to 15-month-olds (The ManyBabies Consortium, 2019) with North American English stimuli. However, there was substantial variability in the effect size, with the strongest effect for North American infants and for older infants (contrasting with the previous finding of a decreased preference with age: Newman & Hussain, 2006). These findings about the variability in IDS preference suggest that this preference is affected by experience: having more experience (i.e., older infants having heard more accumulative IDS) and more *specific* experience of IDS (i.e. being familiar with the type of IDS used in the stimuli) results in a higher preference. This could suggest that rather than infants having an innate preference for IDS because of its role as an ostensive signal, this preference is higher the more chance that infants have had to experience the positive social interactions that co-occur with IDS. This is feasible, as infants receive vast input of speech (and specifically IDS) from birth. Nonetheless, this is substantial evidence that infants are indeed sensitive to IDS.

1.2.1 Alternative theories for infant sensitivity to IDS

If we consider infant sensitivity to IDS, although this is necessary for and compatible with NP, it is also possible that this preference is driven by perceptual characteristics of the IDS that have nothing to do with understanding communicative intent. Alternative theories for this phenomenon can be grouped into perceptual, linguistic, and social theories.

Two key suggestions have been made for how perceptual differences could drive the infant IDS preference. The first is its unpredictability. It has been shown that F0 contours in IDS are less predictable than those in ADS (Räsänen et al., 2018), and this could in effect mean that IDS is perceived by infants to be more novel (statistically unpredictable) and therefore attractive to infants. Another perceptual difference between IDS and ADS is pitch. IDS has a higher pitch than ADS, and it has been found that infants' auditory system develops first for higher pitched sounds (Schneider & Trehub, 1992). This could effectively mean that infants can simply hear IDS better than they can hear ADS (Trainor et al., 2000). This theory (as with many low-level theories) is not necessarily at odds with NP, as it could be true that infants detect a communicative intention from IDS while also hearing it better. However, it is important to acknowledge the possibility that being able to hear IDS better than ADS could be *enough* to result in an IDS preference even without any understanding of communicative intention.

The linguistic theory for infant preference for IDS suggests that IDS emphasizes the lexical and grammatical structure in speech. Snow & Ferguson (1977) suggest that some speech characteristics of IDS may make speech segments more salient. For example, rising sentence-final pitch terminals in imperatives in IDS could function to mark sentence boundaries clearly, helping with the parsing of speech.

However, the authors do acknowledge that some of the speech characteristics of IDS (e.g. rising sentence-final pitch terminals in imperatives) may *also* serve a social function, for example to indicate to the child that a response is required. Infants may therefore attend preferentially to IDS for multiple co-occurring reasons. Trainor and Desjardins (2002) emphasized this point, when finding that although some speech characteristics of IDS improved infant vowel discrimination (the larger pitch contours of IDS), other features actually hindered this discrimination (high pitch). They suggest that although some characteristics of IDS may emphasize the lexical and grammatical structure in speech, high pitch must serve some other function, for example to attract infants' attention or aid emotional communication. Alternatively, as infants' auditory system develops first for higher pitch sounds (as discussed above), this could explain why high pitch, despite not emphasizing lexical and grammatical structure in speech, attracts infants' attention.

Leading on from this, the social theory for IDS preference proposes that having a preference for IDS promotes affiliation with caregivers. This theory suggests that the prosody and pitch of IDS communicates a benevolent intention to infants, and therefore enables affiliation between infants and caregivers before the infant can understand the content of the speech (Fernald, 1991, 1993; Papoušek et al., 1987; Trainor et al., 2000; Trainor & Desjardins, 2002). It is proposed that infants may be attracted to IDS due to the positive affect that it conveys, independent of the content (Trainor et al., 2000). As mentioned above, multiple factors could contribute to the infant preference for IDS, and one trajectory suggested by Trainor & Desjardins (2002) is that first IDS evolved for its social function, and later was co-opted to fulfil the linguistic function as that evolutionary pressure arose. Their suggestion for this

evolutionary order is that the social function seems to “win”, in that one of the main features of IDS is high pitch *despite* its detrimental effect on vowel discrimination.

It seems like there is a complex interplay of different both competing and complementary features and functions of IDS, and it is unclear how each of these either evolved or are learnt. It seems that the theories discussed here focus on the *function* of IDS – i.e. what purpose does it serve. Therefore, rather than asking the question of “Why would infants attend preferentially to IDS?” they ask, “Why would it be useful for infants to attend to IDS?”. One could argue that any function that makes it beneficial to attend to IDS could be selected for evolutionarily, and so it is difficult to distinguish between these theories as they could all predict newborn sensitivity to IDS. Therefore, as with infant sensitivity to direct gaze, infant sensitivity to IDS is consistent with NP, but this finding alone cannot differentiate the likelihood of NP versus competing theories.

1.3 Combination of cues

A key piece of evidence for the claim that infant sensitivity to both direct gaze and IDS is due to a common mechanism (i.e., a preference for ostensive signals in general) rather than both separately due to learning from repeated input of both these signals is that they produce a similar neural response (Kampe et al., 2003; Parise & Csibra, 2013). Interestingly, in 6-month-olds only the combination of both direct gaze and IDS is enough to show enhanced brain activity in a naturalistic scenario (Lloyd-Fox et al., 2015). In adults, it has been found that functional brain activity when perceiving these two types of signal is correlated (and that one signal can affect the perception of the other, e.g. Stoyanova, Ewbank, & Calder, 2010), but it has been found that non-overlapping brain regions process these two signals in 5-month-olds

(Grossmann et al., 2010). From this it is unclear how direct gaze and infant directed speech interact, which is key to understanding whether the infant perceives these signals similarly.

2. Infant referential expectation in ostensive contexts

The second level claim of NP is that infants develop a referential expectation in ostensive contexts. This is to say that infants expect the adult to specify which referent they are trying to draw the infant's attention to (in order to later learn something about it). Csibra & Gergely (2009) argue that the key way that this is done is through deictic gestures, including pointing and object-directed gaze. These gestures are directional, social cues to a referent, and involve the agent expressing a referential intention (although the key topic of debate is whether this intention is understood by the infant or needs to be in order to result in the infant following these cues).

2.1 Infant following of deictic gestures

There is much controversy regarding both the age that infants begin to follow deictic gestures, and what this means about the mechanism for how they do so. NP suggests that this is a very early behaviour that we share with other non-human animals (who use it in order to be able to look where others have deemed worthy of attending to), but that serves a further function in human infants. Investigating the developmental trajectory of infants' following of deictic gestures, as well as the contexts under which it occurs, is key in establishing empirical support for different theories of how this skill comes to be.

2.1.1 Gaze-following

There is considerable controversy with regards to at what age infants begin to gaze-follow. Some suggest that it is an innate ability present from birth, which is evidenced by the finding that newborn infants make faster saccades to peripheral targets cued by the direction of eye movements of a schematic face (Teresa Farroni et al., 2004). This *gaze-cueing* was interpreted by the authors to be a rudimentary form of gaze following. *Gaze following* itself has been shown at as young as 3 (D'Entremont et al., 1997) to 6 (Gredebäck et al., 2008; 2018; Szufnarowska et al., 2014) months of age. Gaze following is also found in many non-human species including ravens (Bugnyar et al., 2004), apes (Okamoto-Barth et al., 2007; Michael Tomasello et al., 2007), and chimpanzees (Povinelli et al., 1996), suggesting that there could be a common evolutionary origin for gaze following ability. Although Csibra & Gergely (2009) acknowledge non-human animals' gaze following abilities, they suggest that there are different mechanisms and usages for this common ability. It is important to note, however, that to make this claim, it must be shown how human infants respond differentially to gaze than other non-human animals.

The ability of infants to gaze follow is affected heavily by the context and paradigm used. Gaze following is affected by age (Brooks & Meltzoff, 2005; Gredebäck et al., 2008; Scaife & Bruner, 1975), whether stimulus presentation is live or computerised (Frischen et al., 2007), whether there is perceived movement (Farroni et al., 2004), whether the eye direction is accompanied by a head turn (Moore, Angelopoulos, & Bennett, 1997), whether a communicative or attention directing cue is used before the gaze shift (Gredebäck et al., 2018; Hernik & Broesch, 2018; Senju & Csibra, 2008; Szufnarowska et al., 2014), and what measure is used (Frischen et al.,

2007). Overall, results are very mixed, which suggests that gaze following may not be an all-or-nothing response that is explicitly understood at some particular point, but more context dependent.

2.1.2 Point-following

As with gaze following, there is much controversy surrounding the age at which infants begin to follow pointing. Historically, it was established that infants could not point-follow until the end of their first year (Blake et al., 1994; Butterworth, 2003; Caron et al., 2002; Daum et al., 2013; Deak et al., 2000; Franco, 2005; Lempers, 1979; Leung & Rheingold, 1981; von Hofsten et al, 2005), and that infants younger than this either found it difficult to disengage from looking at the pointing hand itself (Butterworth & Grover, 1990; Churcher & Scaife, 1982) or could follow pointing only to close but not to distant targets (Flom et al., 2004; Leung & Rheingold, 1981; Morissette et al., 1995; Murphy & Messer, 1977). In these studies, a live actor pointed to a target, and an active head response from the infant was required. However, the development of the point-cueing paradigm allowed a different way of investigating point following that did not require an active head response from the infant. In this paradigm a central pointing hand is displayed and following this a target object appears in a location congruent or incongruent with the direction in which the hand is pointing, and the speed at which the infant looks at the target is compared across congruent and incongruent trials.

Rohlfing et al. (2012) found that even 4- and 6-month-old infants oriented more quickly to the target object when the pointing hand was congruent than when it was incongruent with target location, but only for a dynamic point (i.e. a video of a hand moving into a pointing formation) not a static point (i.e. an image of a pointing

hand), making it possible that infants were following only the motion and not the pointing gesture itself. Following this, Daum et al. (2013) found that a static pointing hand was able to cue 12- but not 10-month-old infants to a target object. However, neither of these studies provided a sufficient control condition whereby pointing was compared to a non-social directional control. Additionally, in both studies the hand disappeared before the appearance of the target, meaning that infants did not have to disengage from the cue in order to look at the target – arguably a key component of point following, at least in a naturalistic setting.

Therefore, Bertenthal et al. (2014) added two control conditions to the paradigm: an arrow, and a foil non-social directional cue shaped similarly to a pointing hand. They found that neither 4- nor 6-month-olds followed the arrow, and only 4- but not 6-month-olds followed the foil, whereas both age groups followed the pointing hand. In a more naturalistic version of the paradigm, Schmitow et al. (2016) found that 6-month-olds only followed a pointing hand if the arm and hand angle were the same, but if these were at odds, infants were at chance. This suggests that rather than understanding the vector of a pointing gesture itself, infants are responding to the prototypical version that they are most often exposed to.

There seem to be contradictory results in the literature with respect to the age at which infants can follow different (especially more and less prototypical) pointing gestures. As with gaze following, motion seems to be an important cue that infants rely on for early point-following, so it is still unclear whether at younger ages infants are truly responding to the vector of the point or only the direction of motion itself.

2.1.3 The importance of ostensive context

A large debate has recently centred on the importance of ostensive context for infant following of referential cues. NP suggests that preceding ostensive cues are necessary for young infants to be able to follow referential cues, because they create a referential expectation. In the core study that forms the evidence for this claim (Senju & Csibra, 2008), 6-month-old infants were shown videos of an actress first either establishing eye contact or no eye contact, and then the actress made a head turn and gazed either left or right towards one of two objects. Infants were more likely to look first at the gazed-at object if this was preceded by eye contact than if a moving cartoon image of a flower was superimposed over the face of the actress. In a second experiment, the same was found for IDS – infants were more likely to look first at the gazed-at object if this was preceded by audio of IDS than if it were preceded by audio of ADS. In an attempt to determine whether this response was a function of experience with gaze, Hernik and Broesch (2018) used a similar paradigm in Vanuatu, where face-to-face parent-infant interactions are more rare than in WEIRD (Westernised, Educated, Industrialised, Rich, and Democratic) communities. They replicated Senju and Csibra (2008), finding that infants only gaze-followed when gaze shifts were preceded with audio of IDS, but not if they were preceded by ADS. The authors of both studies suggest that infants gaze-follow because of the referential expectation induced by the presence of ostensive signals. However, an alternative explanation is possible, that instead, any attention-grabbing stimulus preceding the gaze shift would result in better gaze following. Although the inclusion of a cartoon flower superimposed over the head was in an attempt to control for the attention-grabbing quality of direct gaze, it is unclear whether this would be as attention-grabbing as a human action.

Szufnarowska et al. (2014) investigated precisely this possibility, by comparing the original two conditions (ostensive and non-ostensive) to two additional conditions: shivering and nodding, whereby an actress shivers or nods while looking down prior to the gaze shift. These new conditions served as human actions that were proposed to be attention-grabbing but that would have no reason to induce a referential expectation. The authors found gaze following in both the ostensive and shivering conditions, but not in the non-ostensive or nodding conditions. From this result they concluded that 6-month-olds do not need an ostensive context in order to gaze follow, and instead any attention-grabbing preceding stimulus will help enable them to do so. However, there were some criticisms of this study from the original authors (Csibra et al., 2015). The within-subjects design left open the question of whether there could be a carry-over effect from the ostensive trials, whereby as all infants saw some ostensive trials, other trials could be contaminated in that they were still affected by this ostension.

Following this debate, Gredebäck et al. (2018) conducted another version of this study, in order to respond to the criticisms presented. In this version, the same conditions were used, in a much higher powered, between-subjects study (with almost 100 infants). In this way, this version of the study could serve both as a conceptual replication of the study by Szufnarowska et al. (2014), while also responding to these criticisms and therefore providing a more conceptually sound design. In this study, gaze following was found in all three contexts. This again shows support for the idea that 6-month-olds can gaze-follow without a referential expectation being induced by the presence of ostensive cues. One question may be why only in this version of the study did infants follow gaze even in the non-ostensive condition. One possibility is that because this study was higher powered than either of the other two studies, it was

able to capture smaller effects, including the gaze following effect in the non-ostensive condition. This highlights the importance of studies with higher power than is currently standard in infancy research, as we could be obscuring possible smaller effects and incorrectly concluding that they are not present. Taken together, as both studies providing support for NP (Hernik & Broesch, 2018; Senju & Csibra, 2008) could also be explained by attentional mechanisms, and two studies attempting to tease apart these explanations found support for an attentional mechanism, it seems unlikely that infants gaze follow only when a referential expectation is induced by an ostensive context.

Although less prominently, this hypothesis has also been studied with regards to point following. Daum et al. (2013) used a point cueing paradigm to investigate infant point following, but also manipulated the communicative content of an accompanying acoustic stimulus, which ranged from being human and communicative to being artificial. They found that 12-month-olds were only cued to the correct location if the pointing hand had been accompanied by the communicative acoustic signal. However, Bertenthal et al. (2014) found that even 4- to 6-month-olds were able to follow the pointing hand without any accompanying acoustic stimulus. It seems that, as with the gaze following literature, it is unclear whether a signal is needed before the referential cue in order for young infants to follow. Regardless, even if a preceding signal is necessary, the mechanism for how this would affect infant following of the referential cue is unclear, as all that could be necessary is that the signal is attention-grabbing, regardless of its communicativeness.

2.1.4 Alternative theories for infants' following of deictic gestures

2.1.4.1 Gaze following

If we consider infant gaze following, as discussed above, NP suggests that infants are initially only able to do this with the presence of ostensive cues, which create a referential expectation (Csibra & Gergely, 2009). However, we now know that infants are able to gaze-follow after viewing other non-ostensive signals (Szufnarowska et al., 2014) or even no signals at all (Gredebäck et al., 2018), suggesting that it isn't this referential expectation that drives early gaze following. There are several other theories for how gaze following initially develops, which can be grouped into those suggesting an innate ability, a socio-cognitive process, or social reinforcement as the basis for infant gaze following.

Theories suggesting a completely innate gaze-following ability assume that humans have developed an evolutionary adaptation to the detection of gaze direction. The most famous of these theories has been proposed by Baron-Cohen (1995), who suggests that humans are born with an "Eye Direction Detector" (EDD). However, although (as discussed previously) it is highly probable that infants are sensitive to direct gaze from either birth or very early, this does not seem to be the case for gaze directed in other directions, as the age of onset of gaze following continues to be highly debated but is certainly not from birth. Evidence for an innate *precursor* to gaze following comes from Farroni et al. (2004), who found that newborn infants were faster to make saccades to peripheral targets cued by the direction of gaze than to un-cued targets. It is possible that some rudimentary precursor to gaze following may be present from birth; however, this result needs replication as currently it is the only study showing such an early precursor to gaze following. NP is included within

the group of theories suggesting a socio-cognitive process for the development of gaze following. The main theory other than NP suggesting a socio-cognitive basis is the 'like-me' theory (Meltzoff, 2008). The 'like-me' theory suggests that infants have the innate ability to recognise parallels between their own behaviours and the behaviours of others, and so as they go on to learn about links between their own behaviours and mental states (in this case, their gaze being towards something and their visual perception of that thing), they can then infer this about others also.

The largest but least prominent group of alternative theories for the development of gaze following in infants suggests that social reinforcement is behind this skill. This theory has a long history and has seen a lot of evolution over time. This theory originated with Butterworth (1991) who proposed that infants' attention is driven by intrinsic, attention-capturing properties of objects in the environment combined with change in caregiver's gaze, leading to the eventual learning of gaze-following. However, this theory is unclear on why caregiver gaze should be significant for the infant in the first place, making the argument circular. The attentional cueing account (Corkum & Moore, 1998; Moore et al., 1997) built on this, suggesting that the gaze-following scenario itself is an attentional cueing paradigm, where infants learn that a central stimulus can predict the appearance of a target on either side. Therefore, if the central stimulus is directional, infant attention will match the direction of this central stimulus. Moore and Corkum speculate about exactly what part of the cue is used to determine the direction, suggesting that it may be more the dynamic movement than the eye or head direction itself. Based upon this series of work by Moore and Corkum, Triesch et al. (2006) developed a model of how infants may learn to gaze follow in this way. The authors propose a model of the 'Basic Set'

of structures and mechanisms sufficient for gaze following to emerge. This includes the infant's perceptual skills and preferences, habituation and reward-driven learning, and a structured social environment featuring a caregiver who tends to look at things the infant will find interesting. However, this model didn't take into account how the learner develops its perceptual and motor skills as it is developing its gaze following ability.

Following this, Nagai et al. (2006) successfully taught a developmental robot with a simple neural network cognitive architecture to gaze follow through low-level, supervised associative learning. The robot was equipped with a camera which fed images of a human experimenter to a visual system consisting of a connectionist map, which encoded these images, and a retinal smoothing layer, simulating the development of infants' visual acuity. In the experimental set-up the experimenter held up an object in the robot's visual field. After processing the visual input from the camera, the robot generated a motor command, adjusting the joint angles in its head and neck, resulting in a head turn and a change in its visual field. The robot was then given feedback based on the output error between the location of the object in the visual field and its gaze direction: if the object was centred in a predetermined location in the visual field, gaze following was considered successful and no adjustments to the neural network were made. When the object was outside the visual field or off-centre, gaze following was considered unsuccessful, and random noise was added to the connection weights in the robot's neural network. Across training, therefore, head movements resulting in incorrect gaze following were less likely to be produced, increasing the relative strength of connections which produced correct gaze following. Following training, testing with previously untrained images demonstrated that the robot could successfully follow the experimenter's gaze. However, testing at

intervals in training revealed that it did so in stages: early in learning, the robot initially learned to follow gaze in the horizontal plane, and only later in the vertical plane, making the empirically testable prediction that infants too should initially follow gaze shifts more successfully in the horizontal than the vertical direction (tested in Study 3 of this thesis).

For social reinforcement to be a plausible mechanism for infants to learn how to gaze-follow, there is an assumption that adults tend to look at things that infants find interesting. Deák et al. (2014) investigated whether this is the case in naturalistic play. They found a strong preference in 3- to 11-month-olds for watching caregivers manipulating objects. This finding highlights that in naturalistic interactions, the input is there for infants to plausibly learn how to gaze follow from their interactions with caregivers. Although there is a large body of evidence for the social reinforcement theory of gaze following, one piece of evidence at odds with this is the finding that gaze following emerges even in children of blind parents (Senju et al., 2015). However, it is likely that these infants still have interactions with other adults that may lead to them to developing a normal gaze following ability.

Lastly, it is possible to attempt to integrate existing theoretical accounts when developing a new theory of how gaze following develops. Del Bianco et al. (2019) have attempted just that. They propose the Perceptual Narrowing Account of Gaze Following Emergence, whereby initially infants have an attentional bias for faces in general, which is then refined through experience to show a bias towards head and gaze cues, which eventually is refined further to be based only on gaze cues. This could explain a variety of seemingly contradictory results that vary depending on context, paradigm, and age of infants. For example, this could explain why although

newborn infants are sensitive to gaze shifts independent of a face/head movement, older infants seem to need this additional head movement in order to follow gaze.

2.1.4.2 Point following

Interestingly, although there has been a vast amount of theoretical work on the development of pointing *production*, much less has been done on point following. NP suggests that infants are initially only able to follow pointing (like gaze) with the presence of ostensive cues, which create a referential expectation (Csibra & Gergely, 2009). Despite being an understudied area, many of the alternative theories of gaze following discussed could be applied also to point following. However, pointing is different from referential gaze in some key ways that may impact on the development of following this cue. Firstly, the cue itself is a far clearer vector than gaze. With object-directed gaze, the vector must be extrapolated entirely (as there is no line to follow), whereas with pointing, the line has already been partially made by the hand and finger and needs only to be lengthened. For this reason, it could be hypothesised that an innate vector extrapolation ability could render infants able to follow pointing from birth. However, the evidence discussed shows that this is not the case, as very young infants are unable to disengage from the hand itself. Secondly, pointing is always communicative, whereas gaze is not. One can look at something “for themselves”, i.e. not to direct attention but instead to attend to it for their own benefit. This could mean that reinforcement theories of point following (e.g. Butterworth, 1991) are even more likely, as carers are likely to point almost exclusively to interesting things, resulting in more opportunities to learn the correspondence between pointing and interesting things and therefore reinforce the following of this cue. Lastly, infants are born with the ability to gaze in different directions, whereas

they are not born with the ability to point. This means that if we consider theories that require the infant to compare their own actions with those of their interlocutors (e.g. Meltzoff, 2008), they would only be able to do this once they already begin to point themselves. This seems unlikely, as we have seen that young infants are able to detect the congruence between a specific location and a pointing hand. More theoretical work is needed that integrates both gaze following and point following, without ignoring the key differences between these cues, in order to make predictions about the development of these abilities.

2.2 Additional evidence for infant referential expectation in ostensive contexts

As well as the alleged importance of ostensive context for infant gaze following, other paradigms have also shown infant referential expectation indexed in a different way. Senju et al. (2008) looked not at gaze following, but at whether 9-month-olds preferred to look at object-directed over non object-directed gaze shifts in ostensive and non-ostensive contexts. They found that only when there was a period of eye contact before the gaze shift did infants prefer the object-directed gaze. They showed that when gaze direction and movement of the entire face were at odds, infants still preferred to look at the object-directed than non object-directed gaze shifts, suggesting that their result is not due only to motion direction in relation to an object. Csibra & Volein (2008) found that when 8- to 12-month-olds saw a model look to one of two locations preceded by ostensive cues, they then looked longer at the empty location when the model had previously looked at it than when the model had previously looked at the other location. They suggested that infants expect an object at a location that has been previously gazed at because they have a referential

expectation induced by the ostensive context and are therefore surprised when there is no referent, which has also been found to be the case with older children in a live context (Moll & Tomasello, 2004). However, it could equally be the case that infants have learned that caregivers tend to look more often at things that are not empty than things that are empty. Data on naturalistic input to infants could help clarify this.

In a similar paradigm, Gliga & Csibra (2009) found that 13-month-olds will also show longer overall looking when an object is revealed at a location not indicated by an ostensive model (who also labels the object) than when it is revealed at the location indicated by the model. Interestingly, they found that when the source of the speech (including the label) is not the same as the model indicating the location, this is no longer the case. This result was interpreted as infants having an expectation of co-reference, i.e. that a label will be accompanied by a referential cue in order to identify the referent of the label. In order to ensure infants knew the speech was coming from someone else, in the condition where speech and reference did not co-occur, a male model was used in combination with the female speech from the condition where speech and reference did co-occur. However, this suggests a confound whereby a male model was used in the non-occurring condition, and a female model in the co-occurring condition. This is not a minor issue, as usually female models are used in studies into ostensive cues. It is unclear whether we would expect a difference, but without a counterbalanced version of the study we cannot rule out this possibility.

Another issue is the different variables used to index the infant's response in Gliga and Csibra (2009) and that by Csibra and Volein (2008). In the study by Gliga and Csibra (2009), they found a difference in overall looking time, whereas in the study by Csibra and Volein (2008), they measured overall looking time, looking time

to the object location, and looking to the empty location, and only found a significant difference in looking time to the empty location. It is unclear why a different dependent variable would index the same cognitive process in two different studies, therefore the possibility remains that having too many possible variables that could be looked at and finding a significant difference in the predicted direction for only one of them (and not a consistent one) is due to the use of multiple dependent measures resulting in a false positive result. It is of course possible that there's some difference in the studies that would be the reason for this discrepancy, and so direct replications of these studies could help differentiate which of these reasons for the discrepancy is more likely.

Although this group of studies could be taken as evidence of ostensive context inducing a referential expectation in infants, all of these studies would be susceptible to the same criticisms as Senju & Csibra (2008), whereby no appropriate control was used in place of the ostensive cues. Without knowing how infants would respond in an identical paradigm but with another salient attention-getter instead of ostensive cues, we cannot conclude that ostensive cues induce a referential expectation.

3. Infant bias to interpret ostensive-referential communication as conveying information that is kind-relevant and generalisable

The last of the claims made by NP is that, when infants encounter ostensive-referential communication, they interpret this as conveying information that is kind-relevant and generalisable, as opposed to information that is specific about the current

situation. Importantly, it is this claim that truly differentiates NP from other theories of communication development (e.g. Tomasello, 2008).

3.1 Infants encode valence of objects over preference of adults in ostensive-referential contexts

One of the pieces of evidence for this claim (that, when infants encounter ostensive-referential communication, they interpret this as conveying information that is kind-relevant and generalisable, as opposed to information that is specific about the current situation) comes from the finding that in an ostensive context, infants expect positive and negative emotional displays by actors to reflect on the valence of the objects referred to as opposed to an individual attitude of the actor (Gergely et al., 2007). In this study, 14-month-olds were familiarised to Experimenter 1 repeatedly showing positive emotion to object A and negative emotion to object B, and Experimenter 2 repeatedly showing the opposite pattern. Half the infants saw the same number of both of these types of trial, and half the infants saw 75% of one actress and 25% of the other. At test, both actresses showed the congruent and incongruent emotion with the familiarisation trials for the two objects. If infants were sensitive to the individual preferences of the actors, at test they should show a violation of expectation (VOE) response to the actors showing the incongruent emotion towards the objects. Instead, what was found was that infants only had a VOE in the condition where the number of trials of each type was asymmetric, i.e., if object A was shown positively more times than it was shown negatively, they had a VOE response to it being shown negatively, regardless of actor, and vice versa. This was interpreted as being because the ostensive context of the familiarisation trials biased infants to encode the emotion information as being about the valence of the

object rather than the individual actors' attitudes. However, as there was no non-ostensive control condition, it isn't clear that this is what was driving this effect. It is equally possible that infants made low-level associations between the emotions and the objects and weighed these up depending on how many trials they saw with each association. Additionally, the interpretation of results is mainly based on p values larger than or very close to 0.05, which suggests the values close to but slightly smaller than 0.05 could be false positives, and the others should be interpreted as negative results. The interpretation of the effect of ostensive context is weak without a direct manipulation of ostensive context. In the paper it is mentioned that a version of the study manipulating exactly this is underway, however to our knowledge this was never published.

3.2 Infants encode identity of objects over location in ostensive-referential contexts

One of the most prominent pieces of evidence for this claim (that, when infants encounter ostensive-referential communication, they interpret this as conveying information that is kind-relevant and generalisable, as opposed to information that is specific about the current situation) stems from a study finding that infants were biased to encode identity information about novel objects after viewing them in a ostensive-referential context, and location information after viewing novel objects in a non-ostensive-referential context (Yoon et al., 2008). In this study, the authors presented 9-month-olds with videos of communicative and non-communicative scenarios. The communicative videos included IDS, direct eye contact, and pointing, whereas the non-communicative videos included ADS, no direct eye contact, and reaching. In communicative scenes, an actress said 'Hey

baby!' in IDS, while engaging in direct eye contact, and then pointed towards a novel object out of reach on the left or right side of the scene. In non-communicative scenes, an actress said 'What's this?' in ADS, while looking at the object, and then reached towards the object. Screens then occluded the object and actress. After a few seconds, the occluders opened to reveal the object again. At the point of reveal, either the identity or location of the object had changed, or no change occurred. Infants looked longer (longer looking indicating a violation of expectation, Onishi & Baillargeon, 2005; Woodward, 1998) at the identity change in the communicative condition and at the location change in the non-communicative condition. This finding was key for NP, as it could not be explained by attentional differences, as the information retained was qualitatively different in the two contexts. However, this is the only published study to date that has found this pattern, and we were unable to replicate this finding in Study 1 of this thesis (Silverstein et al., 2019). Other research on this topic in infants has found that an ostensive-referential context either only improves identity memory in 4-month-olds (Hirai & Kanakogi, 2018), or impairs location memory in 9-month-olds (Okumura et al., 2016). Research with adults mirrors this, finding impaired location memory in ostensive-referential contexts (Marno et al., 2014, 2016). All of these outcomes could be explained by the ostensive-referential context either heightening attention to the objects themselves or distracting attention away from object location, and so without the double dissociation found in the study by Yoon et al. (2008), this is not evidence for the claim that the ostensive-referential context biases infants to encode kind-generalisable information. This is discussed in more detail in Study 1 of this thesis.

3.3 Infants do not commit the A not B error in non-ostensive-referential contexts

Another piece of evidence for this claim (that, when infants encounter ostensive-referential communication, they interpret this as conveying information that is kind-relevant and generalisable, as opposed to information that is specific about the current situation) is the finding that removing the ostensive-referential context from the A not B task reduces infants' perseverative errors (Topal et al., 2008). In the A not B task, infants are repeatedly shown an object being hidden in box A, and then an object being hidden in box B. After every trial, the infant has the opportunity to try to retrieve the object from the box. Previously it had been found that once the object starts being hidden in box B, infants continue looking in box A. Topal and colleagues found that when the hiding is not done in an ostensive-referential context (no eye contact, no communication with the infant), 10-month-olds are better at choosing the correct box once the object starts being hidden in box B. This is interpreted as being because the ostensive context biases infants to think they're learning generalisable information about where objects are found, and therefore they persevere at the previous location. However, this seems at odds with the claim that infants are biased to encode identity over location information for objects seen in an ostensive-referential context (Yoon et al., 2008). If this were true, the context may instead cause them to be bad at the task throughout (as they have trouble encoding which box the object has been put in in both A and B trials). Additionally, it is unclear why the ostensive context would bias infants to think they're learning something about the boxes rather than the objects themselves. This is a case where, although the results seem to be consistent with NP, without preregistered predictions we cannot be sure

whether this was just one of many outcomes that could have been concluded to be consistent with NP.

In a comment on the original paper, Spencer, Dineva, and Smith (2009) show through modelling that this pattern of results could also be explained by a low-level model whereby an ostensive-referential context distracts infants from the location information, which could also explain other results previously discussed showing impaired location memory in ostensive-referential contexts (Marno et al., 2014, 2016; Okumura et al., 2016; Yoon et al., 2008). They stress how it is important to evaluate whether results consistent with a high-level theory could also be explained by unified theories of known neural and behavioural processes.

4 General conclusions & open questions

Overall there is a large body of evidence consistent with NP. Infants are sensitive to ostensive cues (in that they're attracted by them), however the mechanism for this is unclear is also consistent with several other theories. Ostensive cues seem to improve infants' gaze- and point-following ability, however other non-social cues seem to work equally well, and so it could be that any attention directing cue improves attention to the following referential cue, or indeed that infants are able to follow referential cues without any attention directing cue in highly-powered experiments.

The most specific of the claims of NP, that differentiates it from competing theories, is that an ostensive-referential context biases infants to encode kind-relevant, generalisable information. There seems to be a smaller body of evidence for this, some with low replicability, and some that can also be explained instead by low-level mechanisms. As it is unclear whether there is currently support in the literature for the

claim that an ostensive-referential context biases infants to encode kind-generalisable information, Study 1 of this thesis attempts to replicate the key study showing evidence for the claim (Yoon et al., 2008).

Following this, instead of directly assessing a claim of NP, we attempt to ask a question that would follow from the claims of NP, even if not explicitly stated. Specifically, NP is about the uniquely human ability to allow transmission of generic knowledge through teaching, and yet the theoretical claims stated and discussed are all in relation to the recipient of the teaching, with no mention of how this is related to the subsequent teaching of information that has been learned in an ostensive-referential context. However, if the reason kind-generalisable information is retained in an ostensive-referential context is because it is culturally relevant, then this information should also be preferentially transmitted to another person (over information that is deemed to be less culturally relevant due to not being learned in an ostensive-referential context). In Study 2, we test this.

In Study 3, after not finding support for NP in either Study 1 or 2, we instead move on to testing a different theory for how infants may learn to follow referential cues (specifically, referential gaze) without the need for a receptivity to communicative intent. Instead, we test whether infants can learn to follow gaze through social reinforcement, testing a prediction from Nagai et al. (2006) that infants should learn to follow gaze in the horizontal plane before the vertical plane due to increased perceptual variability in the horizontal plane during gaze interactions with their caregivers.

Study 1: Probing communication-induced memory biases in preverbal infants: Two replication attempts of Yoon, Johnson and Csibra (2008)

Linking statement

Replication is pivotal to the advancement of science. In a time where it has been repeatedly found that around 50% of published research findings are false (Camerer et al., 2018; Klein et al., 2018; Open Science Collaboration, 2015), it is no longer good practice to continue a research line based on the assumption that previously published findings can be assumed to be true. This is particularly the case when an area has been understudied, or a large amount of emphasis has been placed on a single study to hold the burden of a particular theoretical claim. This study was conceptualised for this exact reason: Yoon et al. (2008) is a seminal paper in developmental psychology, being cited over 150 times to evidence the power of ostensive cues to bias infant memory towards the identity of objects.

As outlined in the introduction of this thesis, this paper serves as one of the key pieces of evidence for the specific claim made by NP that the presence of ostensive cues biases infants to encode kind-generalisable information. In order to build on this result, and investigate the mechanism for this effect, it was important to first replicate this finding. Due to the original study being published before preregistration was prevalent in psychology, it was particularly important to

preregister this attempted replication in line with decisions made in the original paper, in order to ensure that the same design and analytical decisions lead to the same outcomes, rather than being the product of the “garden of forking paths”.

Additionally, it was important to contact the original authors in order to clarify decisions made that were not clear in their original manuscript. Again, the original study was published before open materials were prevalent, and so some aspects of the design of the study were unclear (and misleading) from the manuscript itself. Having two versions of the replication (one before, and one after communication with the original senior author) highlighted how more than a traditional manuscript is needed in order to adequately replicate a study.

This study also provided the opportunity to look at how infant looking unfolds over time in a study with naturalistic videos. As the original study was conducted before eye-tracking was prevalent in infancy research, this was a technological advance that could now provide new data on infant attention in this type of paradigm. It is important to know how infant attention may differ between conditions that we are comparing, especially when trying to probe higher-level processes, as it is possible that low-level attentional differences may instead be responsible for any differences we observe in our dependent variable. Plotting the time course of infant looking also more generally provides a rich picture of what infants are doing during throughout the entire study, as opposed to only the “test phase”.

For these reasons, this set of experiments aimed to be a fair, preregistered replication of Yoon et al. (2008), with all methodological changes and additions well justified, and all data and materials fully open. It must be noted that the sample size for this study was decided based on the sample size of Yoon et al. (2008), which, while a common practice in infancy research to this date, leaves open the possibility

that if there is a true effect being measured but it is smaller than that reported in Yoon et al. (2008), we would have been unable to detect it. If redesigning this study, we would have used a power analysis to determine an appropriate sample size.

Abstract

In a seminal study, Yoon et al. (2008) showed that nine-month-old infants retained qualitatively different information about novel objects in communicative and non-communicative contexts. In a communicative context, the infants encoded the identity of novel objects at the expense of encoding their location, which was preferentially retained in non-communicative contexts. This result had not yet been replicated. Here we attempted two replications, while also including a measure of eye-tracking to obtain more detail of infants' attention allocation during stimulus presentation. Experiment 1 was designed following the methods described in the original paper. After discussion with one of the original authors, some key changes were made to the methodology in Experiment 2. Neither experiment replicated the results of the original study, with Bayes Factor Analysis suggesting moderate support for the null hypothesis. Both experiments found differential attention allocation in communicative and non-communicative contexts, with more looking to the face in communicative than non-communicative contexts, and more looking to the hand in non-communicative than communicative contexts. High- and low-level accounts of these attentional differences are discussed.

Introduction

Humans are expert learners. We learn implicitly, through mechanisms like statistical learning (Fiser & Aslin, 2001; Kirkham et al., 2002; Newport & Aslin, 2004), and explicitly from others through social learning (Csibra & Gergely, 2009; 2011; Tomasello et al., 2005). Social learning can occur either through observation (Meltzoff, 1988b; 1988a; Meltzoff & Moore, 1989), or through pedagogy, or explicit

teaching (Csibra, 2007; Csibra & Gergely, 2009; Tomasello et al., 2005). Although teaching usually involves language, knowledge transfer can also occur in its absence. Two types of communicative cues have been suggested to be key to information transmission through teaching: ostensive cues such as direct eye contact or infant directed speech (IDS) convey the intention of communication, and referential cues such as pointing or gaze shifts direct attention to the source of the information to be learned (Csibra, 2010; Csibra & Gergely, 2009). According to the Natural Pedagogy theory (Csibra & Gergely, 2009) ostensive communicative cues signal to infants when to learn culturally relevant kind-generalizable information about an object. In the presence of these cues, infants would be biased to encode surface features, which support learning about object kinds, over spatio-temporal information.

One method to investigate how infants encode object properties is the violation of expectation paradigm (VoE). The VoE paradigm is based on the assumption that infants look longer at events that violate their expectations (Onishi & Baillargeon, 2005; Téglás et al., 2011; Woodward, 1998) including when features of an object change (Krøjgaard, 2009; Mareschal & Johnson, 2003). Yoon et al. (2008) used a VoE paradigm to test the hypothesis that being communicated to should bias infants to encode surface features. In their study, the authors presented 9-month-old infants with videos of communicative and non-communicative scenarios. The communicative videos included IDS, direct eye contact, and pointing, whereas the non-communicative videos included adult-directed speech (ADS), no direct eye contact, and reaching. In communicative scenes, an actress said ‘Hey baby!’ in IDS, while engaging in direct eye contact, and then pointed towards a novel object out of reach on the left or right side of the scene. In non-communicative scenes, an actress said ‘What’s this?’ in ADS, while looking at the object, and then reached towards the

object. Screens then occluded the object and actress. After a few seconds, the occluders opened to reveal the object again. At the point of reveal, either the identity or location of the object had changed, or no change occurred.

Infants looked longer at the identity change in the communicative condition and at the location change in the non-communicative condition (both in terms of first look and total look length). The authors concluded that infants encoded the identity of the object after being communicated to, as this was relevant to kind-generalizable learning. In contrast, infants encoded the location of the object in the non-communicative condition, due to this being the default, or perhaps the attempted reach enhancing the perceived graspability of the object. This double dissociation in the encoding of identity and location information suggested that the communicative cues did not merely increase overall memory but elicited a specific memory bias towards identity information.

There have been few papers so far attempting to replicate or extend this finding. [Okumura et al. \(2016\)](#) found that in a live study, 9-month-old infants showed an identity bias in a direct gaze condition. However, the authors did not replicate the finding of a location bias for the condition with no direct gaze, instead finding encoding of both identity and location. The authors suggest that, due to the video deficit effect ([Anderson & Pempek, 2005](#)), infants may have performed better in their study than in the original study, therefore managing to encode both identity and location in the non-communicative condition. Their findings suggest that instead of identity being preferentially encoded by infants after viewing communicative scenes, it may be that infants are able to encode both spatiotemporal and recognition-relevant features, but ostensive signals disrupt location encoding. Two studies following up on this result in adults drew the same conclusions ([Marno et al., 2014; 2016](#)) finding

encoding of both identity and location information in the non-communicative condition, and only encoding of identity in the communicative condition. As there are only two studies investigating communicatively induced memory biases in infants, we felt it was necessary to replicate the original finding before extending this research ourselves.

After being sent example videos from one of the original authors we noticed that in the communicative videos, the actress pointed for longer than she reached (6.8 s compared to 4.5 s). This difference raises the possibility that a longer duration of having the hand on screen could be responsible for inducing an identity memory bias. Therefore, in our stimuli both types of actions were performed twice for the same duration. Also, in the original study, the actress had bars stopping her from being able to reach the object. We did not use bars in our stimuli, instead having the actress be too far away from the object to reach it, in order to conceptually replicate the idea that the actress was unable to reach the objects, but without obscuring her. Lastly, we used eye tracking to investigate infants' distribution of attention while observing the communicative or non-communicative scenes. Although Yoon et al. (2008) compared overall looking to the action scenes and reported no difference between conditions, this does not speak to where the infants were looking while viewing these scenes. A difference in attention allocation in the two contexts could be responsible for the memory biases observed. For example, as it has been argued that preverbal infants tend to follow referential cues only when these are preceded by ostensive cues (Senju & Csibra, 2008; but see Gredebäck et al., 2018; Szufnarowska et al., 2014), we thought that perhaps in the communicative context infants would look more directly towards the objects, and that this could enhance the encoding of features. As this part of the study was exploratory, we did not pre-specify specific hypotheses relating to

looking during the pre-occlusion section of the videos. We report two replication attempts. The first replication, Experiment 1 (bar the changes outlined above), followed the methodological details reported in Yoon et al. (2008). In the second replication, Experiment 2, we made slight changes to the method and exclusion criteria following guidelines provided by Csibra (personal communication, November 2017). Both replication attempts were preregistered, and all data, code, supplementary results and materials are openly available on the Open Science Framework (OSF; osf.io/77gpt).

Experiment 1

Method

Experiment 1 was conducted following the method section of Yoon et al. (2008), bar some changes outlined in the introduction. Participants were recruited from a database of families at an infancy lab of a UK university, and were given a book as a gift for participation and £10 travel reimbursement. Parents gave informed, written consent before participation, and were free to withdraw their consent. All data were kept confidential. Both experiments were approved by the university ethics committee and adhered to the British Psychological Society guidelines.

Participants

Replication. Forty-two normally developing 9-month-old infants took part in the experiment. Of these twenty-four were included in the replication analysis (mean age: 274 days; range: 258 days to 286 days; 9 female; 23 Caucasian; 20 monolingual English). Exclusion criteria matched those used in Yoon et al. (2008): Infants were

excluded for ceiling looking time for all trials ($n = 4$), experimenter error ($n = 4$), fussiness ($n = 7$), and not looking during one or more occlusion events ($n = 3$). An occlusion event in this experiment was defined as the time between the first frame of the clip when the occluder starts closing, and the first frame when the occluder is fully closed.

Scene analysis. We were able to use less stringent exclusion criteria for this analysis as our exclusion criterion of watching the entire occlusion event was irrelevant when looking at infant looking before the occlusion. Of the forty-two infants who took part in the experiment, 40 were included in the scene analysis (mean age: 276 days; range: 258 days to 322 days; 17 female; 39 Caucasian; 20 monolingual English). Two infants were excluded due to experimenter error.

Stimuli

We created the video stimuli and digitally added objects (on the left or right side of screen) and occluders on each clip. Infants first saw two familiarization trials. These familiarization videos were 29 s long and consisted of the actress moving around slightly to upbeat music, while looking either at the infant with direct gaze and smiling (communicative condition), or at the object with intrigue (non-communicative condition) (6 s). After this, yellow screens occluded the object and actress (3 s), there was a short break where the occluders stayed closed (5 s), the object screens revealed the object again (with no change) (2 s), and the object remained on screen for a maximum of 15 s (less if the infant looked away for two seconds, in which case the next trial was advanced). We always used the same two objects, but counterbalanced across participants for actress, side and order.

In the test videos (Fig. 1), there was first an introductory sentence produced by the actress (“Hey baby!” with direct gaze for the communicative videos, and “What’s that?” with gaze to the object for the non-communicative videos) (4 s). This was followed by the action being executed once towards an object to the front of the actress on the left or right side (a point or a reach towards the object as if to try and grasp it) (4 s). The actions were completed at an equal distance away from the object and were of the same duration. After completing the action once, the actress returned to the resting position and said either “Wow” (with direct gaze) while waving at the infant (communicative) or “Hmm” (without direct gaze) with a hand on her chin (non-communicative) (4 s). After this, she produced the pointing or reaching action again (4 s). Following this, screens moved to occlude both the object and the actress (2 s). The occluders stayed on the screen (5 s), after which the object screens reopened (2 s) to reveal the object. There had either been no change to the object, or it had changed in either identity or in location. The objects stayed on screen for a maximum of 15 s, or until infants looked away for 2 seconds. Test videos were 37 s long. Occluders produced sounds when opening and closing to direct infant attention to these events.

We obtained the objects from the Noun database (Horst & Hout, 2016), and chose 6 pairs with medium similarity ratings. We used the first of each of these pairs as the initial object, and the paired objects for the identity change condition (i.e., when the object changed identity, the second object in the pair was what it changed into). Every infant saw all 6 of the objects at test. Which object was shown for which condition combination was pseudo-randomised into 8 trial orders (with 3 infants viewing each order). Which actress played which role (communicative or non-communicative) was counterbalanced across infants. Object, side of action, condition and outcome were pseudo-randomised in the 8 possible trial orders. Additionally,

infants never saw more than two trials similar on any factor (e.g. no more than two left actions) in a row. All videos are openly available on the OSF.

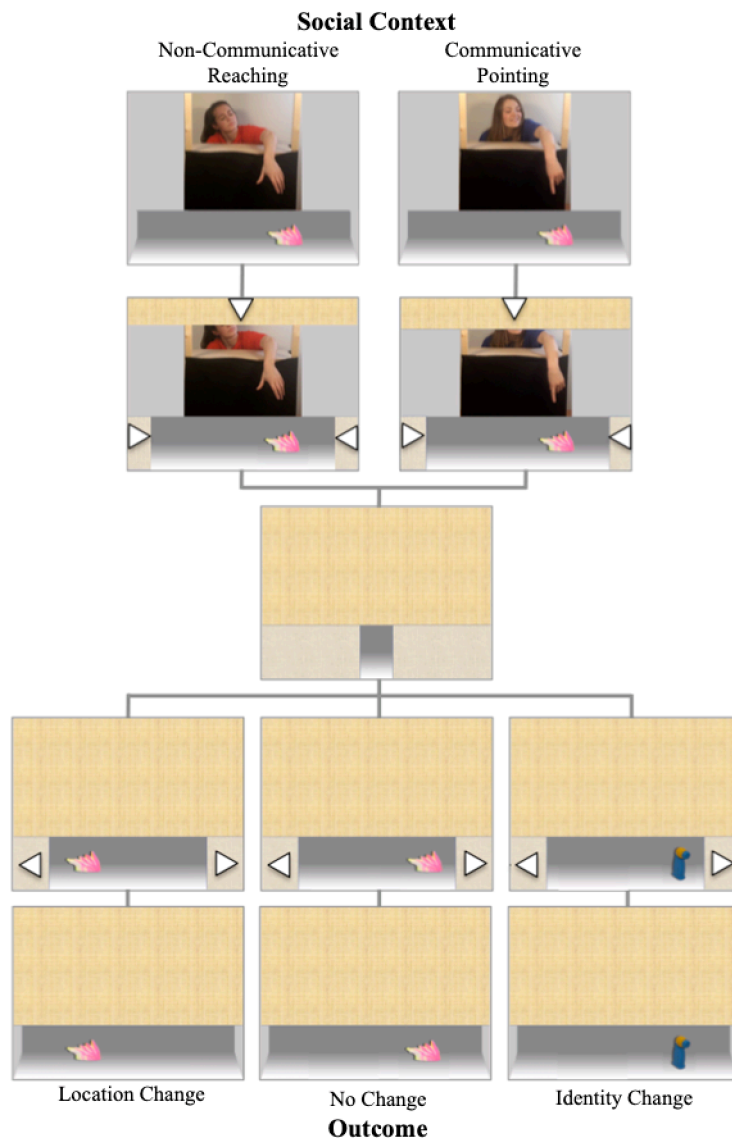


Figure 1. Flow chart of stimuli in the test trials. Full videos are available on the OSF.

Procedure

Infants sat on their caregiver's lap during the experiment in front of a 23-inch screen (seated approximately 0.6 meters away). An eye-tracker (Tobii X120) captured infant looking times and locations on screen for the action scene analysis. We used Tobii studio 3.3.1 to present stimuli and gather eye-tracking data. A camera placed

above the screen fed input directly into Tobii studio, from which videos of infants were later exported for offline looking time coding. We performed a 9-point calibration for all infants before beginning the experiment. After this calibration, we instructed parents not to talk to or interact with their infant, and the experiment began.

Infants saw 2 familiarization videos (1 communicative, 1 non-communicative) followed by 6 test videos (for each of the communicative or non-communicative familiarization video there was one no change, one identity change, and one location change test video). Therefore, each infant contributed one trial per sub-condition to the replication analysis, and three trials per condition to the scene analysis.

Analysis

Replication. We performed all analyses on raw data in R (R Core Team, 2017). All code is openly available on the OSF. Videos of infants were exported from Tobii studio and infant looking was blind coded in Elan (Version 5.0.0, 2019). A second blind coder performed secondary coding on 20% of videos. We found a high degree of reliability between coders: the average measure ICC was 0.97 with a 95% confidence interval from 0.95 to 0.99. The length of the first look was defined as the time between the infant looking towards the screen and their first look away, beginning at the first frame of the occluders opening. The total looking time was defined as the cumulative length of time of all looks towards the screen, beginning at the first frame of the occluders opening, and ending at the first frame of the next attention getter.

We replicated the analysis by Yoon et al. (2008). We carried out a 2x3 repeated ANOVA on the length of first look to screen after object reveal as a function of action (communicative vs. non-communicative) and outcome (identity change,

location change, no change), with follow up one-way ANOVAs and parametric and non-parametric pairwise tests. The same analyses were conducted for total looking times.

Scene analysis. We performed all analyses on raw data in R. All code is openly available on the OSF. We exported raw data from Tobii Studio 3.3.1, and visualized and analysed them in R using the *eyetrackingR* package (Dink & Ferguson, 2015). First, Areas of Interest (AOIs) were created for face (630 X 310 pixels), hand (385 X 340 pixels), and object (380 X 340 pixels) areas of the videos. The data were then visualized as a timecourse. We performed a bootstrapped cluster based permutation analysis (Maris & Oostenveld, 2007) to establish during which time-points conditions differed significantly. This involved running a test on each time bin (17 ms) that quantified a significant difference between actions (communicative vs. non-communicative). We grouped into clusters the adjacent bins that showed a significant difference. We then shuffled the data and performed this same test on one thousand iterations of the shuffled data. This produced a table of the probability of each cluster appearing under the null hypothesis. Clusters that had a probability of less than 5% of appearing under the null hypothesis (i.e., $p < 0.05$) were considered to be significant. This test accounts for both Type 1 and Type 2 errors, by controlling the false-alarm rate while sacrificing little sensitivity (Maris & Oostenveld, 2007).

Results

Replication: First look length

We compared the length of first look in a 2 X 3 ANOVA with action (communicative vs. non-communicative) and outcome (location change, identity change, no change) as within-subject factors (Fig. 2). There was a significant main

effect of outcome [$F(2,46) = 4.15, p = 0.02, \eta_p^2 = 0.15$]. There was no significant main effect of action [$F(1,23) = 0.2, p = 0.66$] or significant interaction between action and outcome [$F(2,46) = 1.32, p = 0.28$]. Paired comparisons by parametric (Student's *t*) and nonparametric (Wilcoxon rank-sum) tests were carried out to assess the effect of each outcome on the looking times. Infants looked longer to the identity change than the no change outcome, regardless of communicative context. Looking was significantly longer for the identity change than the no change outcome [$t(23) = 2.79, p = 0.008, \eta_p^2 = 0.26$; Wilcoxon's $Z = -2.45, p = 0.01$], whereas looking time for the identity change and location change [$t(23) = 1.47, p = 0.15$] and location change and no change [$t(23) = 0.95, p = 0.35$] outcomes did not differ significantly. A nonparametric McNemar test found that this difference could be attributed to some of the infants rather than the entire group. Six of the 24 infants showed the behaviour reported in Yoon et al.'s study (longer looking at identity change than no-change after communicative scenes and longer looking at location change than no-change after non-communicative scenes), 9 infants showed only the identity bias (longer looking at identity change than no-change after communicative scenes), 5 infants only showed the location bias (longer looking at location change than no change after non-communicative scenes), and 4 infants showed the opposite pattern in both contexts (longer looking at the identity change than no-change after non-communicative scenes and longer looking at location change than no change after communicative scenes). This distribution was not significantly different from chance (McNemar's $p = 0.42$).

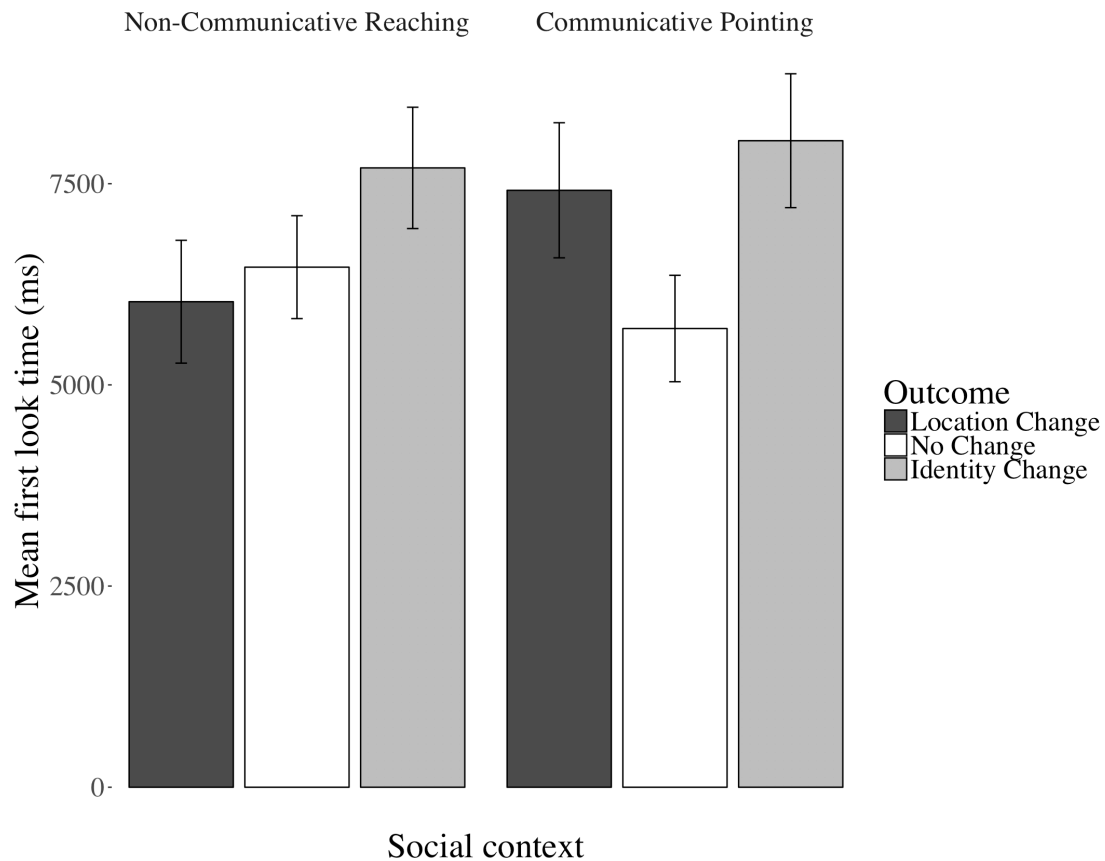


Figure 2. Experiment 1: Mean duration of first look with SE bars for each outcome grouped by communicative context.

Replication: Total look length

Total looking length was compared in a 2 X 3 ANOVA with action (communicative or non-communicative) and outcome (location change, identity change, no change) as within-subject factors (Fig. 3). There was no main effect of outcome [$F(2,46) = 0.66, p = 0.52$] or action [$F(1,23) = 0.01, p = 0.93$], showing that overall infants did not look longer at test in either the communicative or non-communicative condition. There was a significant interaction between action and outcome [$F(2,46) = 3.5, p = 0.04, \eta_p^2 = 0.13$]. We carried out separate 3-level one-way repeated-measures ANOVAs followed by paired comparisons by parametric (Student's *t*) and nonparametric (Wilcoxon rank-sum) tests to assess the effect of each

outcome on the looking times. In communicative context trials, there was no difference in looking time between the three outcomes [$F(2,46) = 1.06, p = 0.35$]. In non-communicative context trials, there was a main effect of outcome [$F(2,46) = 3.91, p = 0.03, \eta_p^2 = 0.15$]. Further analyses revealed significantly shorter looking for location change compared to both identity change [$t(23) = 2.36, p = 0.03, \eta_p^2 = 0.19$; Wilcoxon's $Z = -2.34, p = 0.02$] and no change [$t(23) = 2.21, p = 0.04, \eta_p^2 = 0.18$; Wilcoxon's $Z = -1.76, p = 0.08$] (note: when using a non-parametric test the difference between looking time to location change compared to no change was not significant). There was no difference between identity change and no change outcomes [$t(23) = 0.61, p = 0.55$].

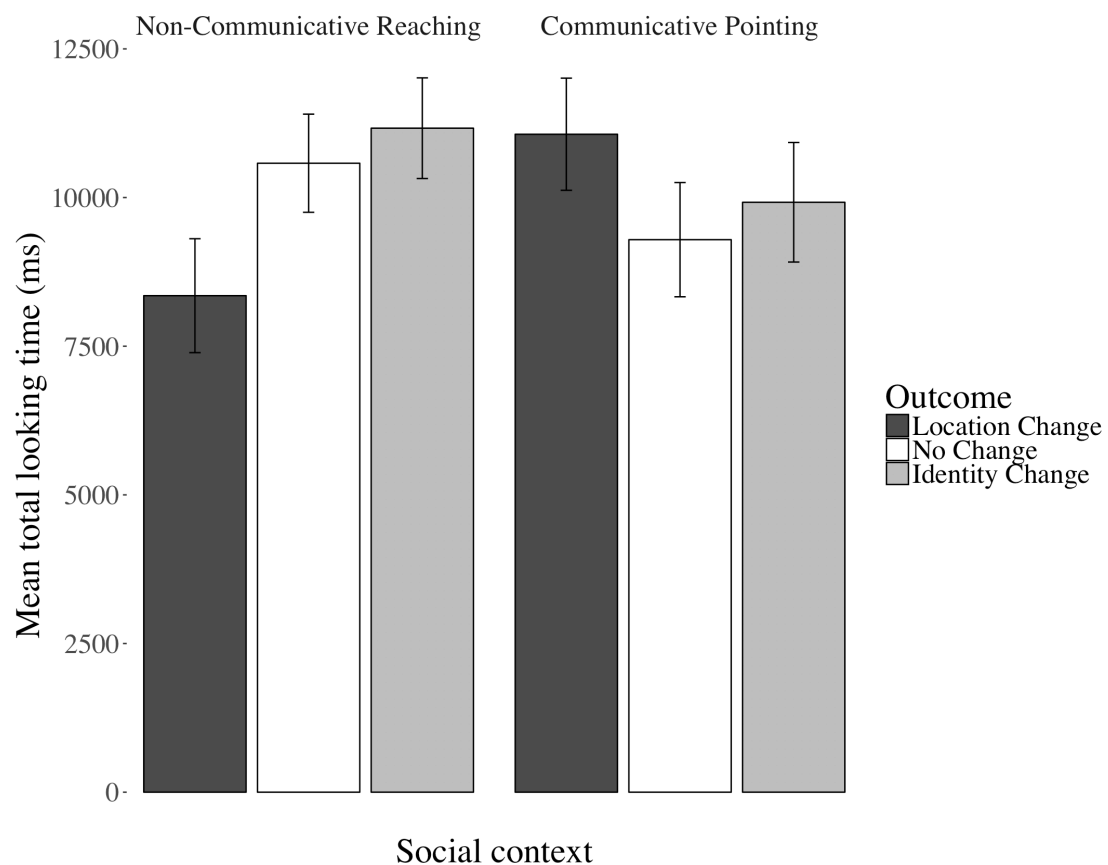


Figure 3. Experiment 1: Mean duration of total look with SE bars for each outcome grouped by communicative context.

Scene analysis

Figures 3, 4 and S5 (see supplementary materials on the OSF) show proportion looking to the face, hand, and object AOIs, respectively, for communicative and non-communicative scenes. Overall, infants showed similar looking patterns when viewing both types of scenes (looking towards the face the most, especially when the actress was speaking, looking towards the hand when the point/reach was being performed, and very little looking towards the object at any time point). A bootstrapped cluster-based permutation analysis found that during both of the periods where the hand action (point/reach) was not being performed (0-3000 ms, 6000-9000 ms), looking towards the face was significantly higher in the communicative condition ($p = 0.027$ and $p = 0.048$ respectively), and conversely, during the first time the action (point/reach) was performed (3000-6000 ms), looking towards the hand was significantly higher in the non-communicative condition ($p = 0.024$) (Fig. 4). Looking towards the object was very low overall and did not differ significantly between conditions. Overall looking time towards the scene did not differ between conditions [$t(39) = 0.03$, $p = 0.98$].

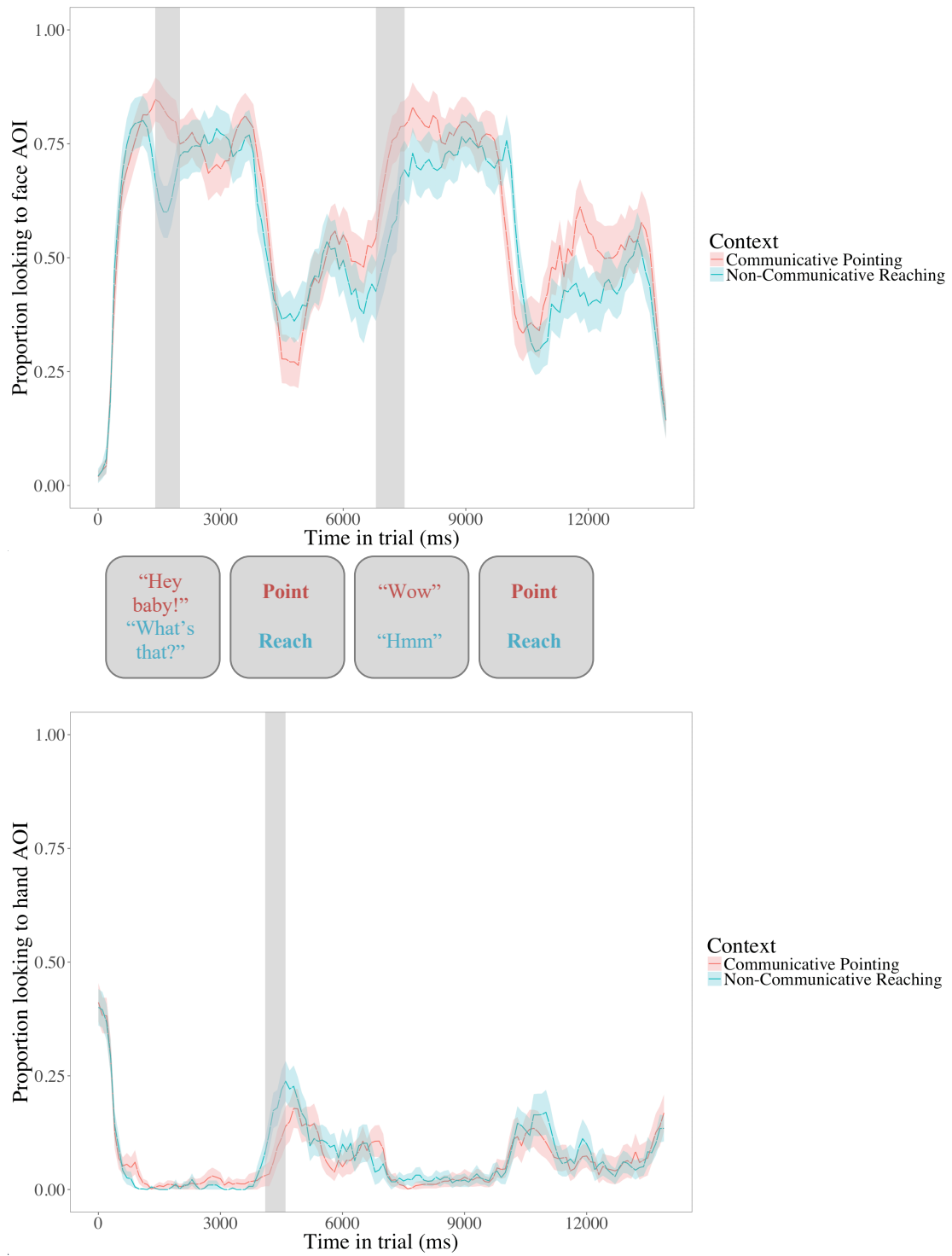


Figure 4. Proportion looking to face (top) and hand (bottom) plotted over time (ms).

N=40.

Discussion

Replication

Our results do not replicate the finding by Yoon et al. (2008) that infants show a memory bias for identity information after viewing communicative scenes, and a memory bias for location information after viewing non-communicative scenes. Neither do we find that communicative scenes disrupt the encoding of location information, which is preserved when viewing non-communicative scenes (Marno et al., 2014, 2016; Okumura et al., 2016). Instead, our results suggest that infants show longer looking to identity changes regardless of communicative context (measured by length of first look). However, we found that a minority of infants drove this effect, and that we do not find the results in the same direction for total look, weakening our belief that this truly indicates an identity memory bias. If we do interpret this result as an identity memory bias, this is surprising, given that the default for preverbal infants seems to be to encode location over surface features (Carey & Xu, 2001; Haun et al., 2006; Mareschal & Johnson, 2003; Xu & Carey, 1996). After this first replication attempt, we contacted Csibra, who generously provided detailed comments that led to some key methodological changes in Experiment 2.

Scene analysis

Our results showed that there are differences in where infants allocate their attention when viewing communicative and non-communicative scenes. Infants looked more towards the face when direct gaze and infant-directed speech were displayed, and more towards the hand when a reach was performed than when a point was performed. This shows that even before infants are actively communicating themselves, they are responding differently when they are being communicated to,

compared to when they are not. As both communicative and non-communicative scenes involved speech and the same sequence of actions, we can reasonably assume that differences are due to more specific features of the two types of scene. This experiment cannot clarify whether these differences are due to low-level perceptual differences or a higher-level understanding of being communicated to.

We found that infants looked more towards the face when direct gaze and infant-directed speech were used. However, in these scenes, this was also when the actress was waving to the infant. We could hypothesize that merely the movement of waving the hand is more salient than the moving of the head to look at the object in the non-communicative scenes, purely because there is more motion involved. Alternatively, consistent with Natural Pedagogy account (Csibra & Gergely, 2009), infants might look more towards the face when ostensive signals are present because they are prepared to learn from the interlocutor. Again, our results cannot differentiate between these two interpretations.

We also found that infants looked more towards the hand when it was reaching than when it was pointing. This difference in looking occurred at a different time-point to where we saw differences in looking towards the face, suggesting that this is not merely the other side of the coin (i.e. when infants are not looking at the face they are instead looking at the hand), and is in fact a different process at play. One low-level interpretation for this result could be that the hand occupies more space when it is a reaching hand than when it is a pointing hand, which could draw the infants' attention. Also, the reaching hand moves around a little, to show that the actress is unsuccessfully trying to reach the object, whereas the pointing hand does not move. Like the waving in the communicative videos, enhanced hand looking in this case could simply be the product of motion drawing the infants' attention.

Alternatively, as the goal for the reach is to grasp the object, infants might fixate on the hand in order to see what happens next (i.e. whether the person manages to reach the object), whereas in the communicative condition, the goal of the point (to communicate) has been reached as soon as the infant perceives and understands it themselves. Again, these data do not speak to which of these interpretations is more likely. Our question was not the mechanisms behind any attentional differences, but instead whether attentional differences could be driving memory biases, and so further research should investigate the reasons for these differences. However, as we find an identity bias regardless of action condition, these differences in where infants allocate their attention cannot be responsible for any memory biases in our experiment. It is also worth noting that infants are not looking at the object in either context, suggesting that they are not fully understanding the goal, as in both cases the goal is either to share attention about an object, or to reach the object. We know that 12-month-olds anticipate the goal of reaching actions by looking at the object, but 6-month-olds do not (Falck-Ytter et al., 2006), so perhaps the infants in our experiment are too young to fully comprehend these goal directed actions (but see Kanakogi & Itakura, 2011).

Experiment 2

After communication with Gergely Csibra (personal communication, November 2017), we were made aware of some important methodological differences between our replication attempt and the original study. Most importantly, we found that our interpretation of ‘occlusion event’ was not the same as theirs. We had interpreted the ‘occlusion event’ as the time where the occluders were currently closing, whereas they had meant it to mean the time where the occluders were

closing, *plus* the entire period where the object is currently occluded. As the original exclusion criteria specified that infants should be excluded if they had not watched all of the occlusion events without looking away, this meant that we had essentially used a different exclusion criterion. When we went back to our data to check which infants would still be included with this new, very strict criterion, we found that none of them would be. Further discussion with Csibra made it apparent that in Yoon, Johnson and Csibra (2008) the occlusion time had been wrongly reported as five seconds, when in fact, in the original stimuli this was actually only three seconds. This longer occlusion time, and the fact that we did not include any music during the occlusion period, may have been responsible for none of our infants showing continuous looking at the screen during the whole 5 s occlusion event for all six trials. Therefore, in the second replication attempt, we shortened the occlusion time to 3 s, added music to the occlusion period, and changed our exclusion criterion to match that of the original paper. Additionally, this second replication could serve as a confirmation for our action scene findings, as this analysis was exploratory in the first attempt.

Method

The method remained overall the same as in Experiment 1, but with the changes to occlusion duration and exclusion criteria discussed with Csibra.

Participants

Replication. Seventy-nine typically developing 9-month-old infants took part in the experiment, and of these twenty-four were included in the replication analysis (mean age: 273 days; range: 261 days to 287 days; 13 female; 22 Caucasian; 23 monolingual English). Infants were excluded for falling asleep ($n = 1$), ceiling looking

time for all trials (n = 7), fussiness (n = 14), experimenter error (n = 1), and not looking during one or more occlusion events (n = 32). An occlusion event in this experiment was defined as the time between the first frame of the occluder beginning to close and the first frame of the occluder being fully open (this was a crucial difference to Experiment 1).

Scene analysis. Of the seventy-nine infants who took part in the experiment, sixty were included in the scene analysis (mean age: 272 days; range: 260 days to 289 days; 31 female; 55 Caucasian; 57 monolingual English). Infants were excluded for falling asleep (n = 1), fussiness (n = 11), less than 40% good eye-tracking data (n = 3), and experimenter error (n = 1).

Stimuli

Stimuli were the same as in Experiment 1, except for three changes. First (outlined above), the objects were occluded for three seconds instead of five. Second, there was a larger gap (roughly 5 times wider) between occluders when objects were fully occluded. This was to ensure infants could see that the object had not moved from one side to the other (and thus, a location change would be genuinely surprising). In order to make such gap larger, the third change was that the objects were made slightly smaller, and moved further apart, which also made the object spacing more comparable to the original study.

Procedure

The procedure was identical to that of Experiment 1.

Analysis

Replication. The analyses were identical to those of Experiment 1. Secondary coding was performed on 20% of videos by a second blind coder. A high degree of reliability was found between coders. The average measure ICC was 0.95 with a 95% confidence interval from 0.90 to 0.98.

Scene analysis. The analysis was identical to that of Experiment 1.

Results

Replication: First look length

The 2 x 3 ANOVA revealed no significant main effect of outcome [$F(2,46) = 0.13, p = 0.88$] or action [$F(1,23) = 0.62, p = 0.44$] or significant interaction between outcome and action [$F(2,46) = 1.32, p = 0.28$] (Fig. S4). No statistical inference can be derived from this non-significant result (Lakens et al., 2018). To determine whether the current data provides evidence for the null hypothesis (H0) relative to the alternative hypothesis (H1), a Bayes factor was conducted. Bayes factors (BF01) provide a measure of how likely the data are assuming H0 is true relative to how likely the data are assuming H1 is true. For the current analyses, a default Bayes factor with a wide cauchy distribution (scale of effect = 0.707) was calculated using the BayesFactor R package (Morey & Rouder, 2015), and yielded $BF_{01} = 4.12$. Thus, we can conclude that the data constitutes moderate evidence for the null hypothesis.

Replication: Total look length.

The 2 x 3 ANOVA revealed a significant main effect of action [$F(1,23) = 4.33, p = 0.05, \eta_p^2 = 0.16$], with infants looking longer to all outcome conditions after

viewing the communicative videos. There was no main effect of outcome [$F(2,46) = 0.09, p = 0.92$], or interaction between action and outcome [$F(2,46) = 0.01, p = 0.99$] (Fig. S5).

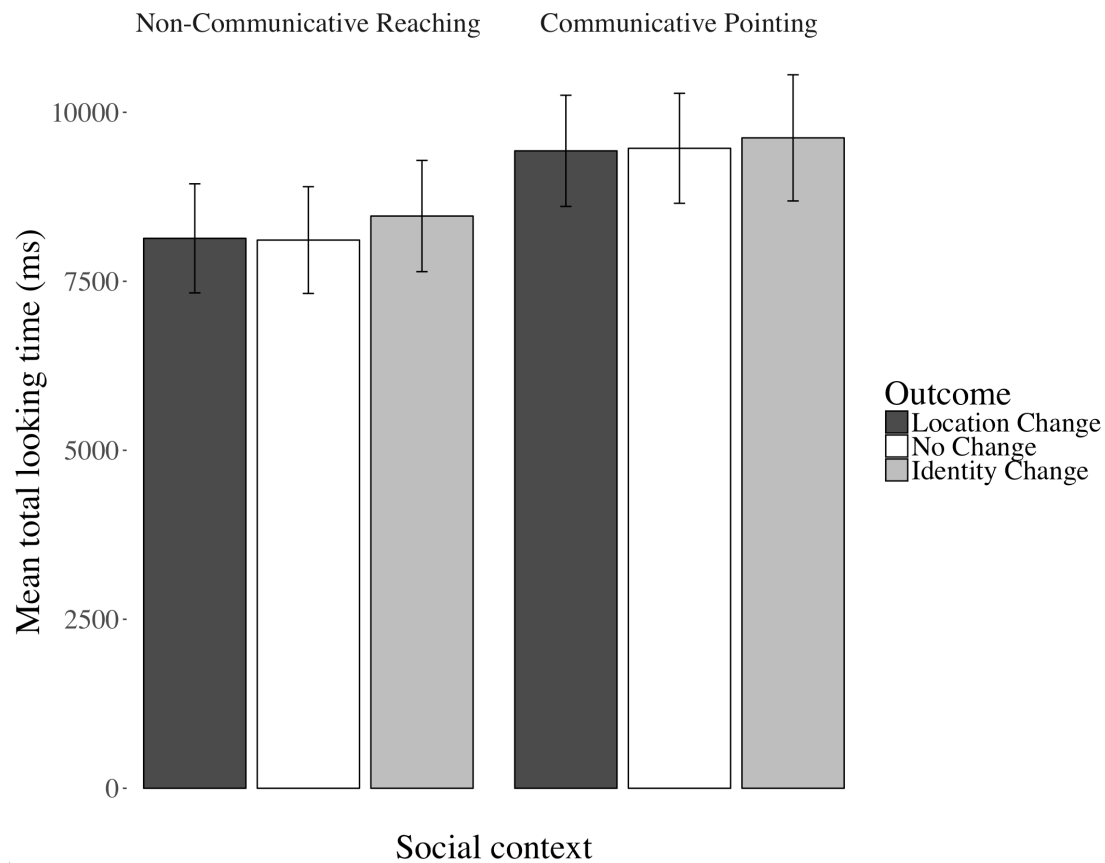


Figure 5. Experiment 2: Mean duration of total look with SE bars for each outcome grouped by communicative context.

Scene analysis

Our findings from Experiment 1 for the scene analysis were exactly replicated (Fig. S6-8). During both of the periods where the hand action (point/reach) was not being performed, looking towards the face was significantly higher in the communicative condition ($p = 0.002$ and $p = 0.025$), and conversely, during the first time the action (point/reach) was performed, looking towards the hand was

significantly higher in the non-communicative condition ($p = 0.038$). There were no significant differences in proportion looking to the object. Overall looking time towards the scene did not differ between conditions [$t(59) = 0.86, p = 0.39$].

Discussion

Our results show no evidence for memory biases, instead showing overall increased attention to the screen at test after viewing communicative videos. Our scene analysis results completely replicate findings from Experiment 1, suggesting that the attention allocation differences found are reliable.

General discussion

We have described two attempts to replicate the results of Yoon et al. (2008). In the original study, infants looked longer at an identity change following a communicative context, and longer at a location change following a non-communicative context. These results were interpreted as a preferential encoding of object identity in a communicative context. In our first replication attempt, we found longer looking at identity changes regardless of context. In our second attempt (which, following communication with Csibra (personal communication November 2017), was better matched to the original study in terms of stimuli and exclusion criteria), we found no memory biases at all, and instead just higher increased overall attention to the screen after viewing communicative videos compared to non-communicative videos. It is important to note the discrepancy in the results from Experiment 1 and Experiment 2. If we assume that the changes in stimuli and exclusion criteria in the two experiments did not have a meaningful impact on infant

looking, these results may be due to random fluctuations due to small sample sizes, as when the two Experiments are combined, we get moderate to high support for the null hypothesis (see supplementary materials). Alternatively, if we assume that there were key differences in the stimuli for the two experiments, we should compare results from Experiment 2 to those found by Yoon et al. (2008), as these are better matched. These results replicate neither the original study showing a double dissociation of identity and location memory biases, nor do they replicate other findings of impaired location memory in communicative contexts (Marno et al., 2014, 2016; Okumura et al., 2016) (table 1) or even any memory effects in general (Blaser & Kaldy, 2010; Kibbe & Leslie, 2011). It may be that the method itself is insensitive to measuring infant object memory in this specific paradigm.

Our results could differ from those found by Yoon et al. (2008) because of small differences in our stimuli such as the size of the faces, the use of superimposed objects as opposed to real ones, or the distance between the two objects (video examples from Yoon et al. (2008) are available on the OSF for comparison). It is also possible that some of these differences had an effect on the higher exclusion rate in our study compared to the original (70% vs. 57%) due to our videos being potentially less engaging. We did also purposefully make the methodological changes of the absence of the bars, and of the matching of the duration of actions in both conditions. However, we have no reason to assume that any of these factors would affect the postulated creation of memory biases through the presence or absence of communication. Despite this, it is still impossible to know whether these (or other) changes could be responsible for the difference in results. Regardless, these small changes are not accounted for in current theory, which would predict that with our setup we would find the same result as Yoon et al. (2008). The original study is a key

piece of evidence for the claim of Natural Pedagogy that ostensive signals not only enhance attention, but also specifically induce an expectation to learn kind-generalizable information. If small differences to a paradigm can disrupt this, then this might question the generality of the claims made by this theory, as the changes that we made should not affect the hypothesized mechanism.

It is possible that the current study is a Type 2 error. This seems unlikely, given that we didn't replicate in either of the two attempts, and find moderate support for the null hypothesis using Bayes Factor Analysis (and moderate to high support for the null hypothesis if we combine the results from Experiments 1 and 2 – see supplementary materials on the OSF). However, it does remain possible that the true effect size is smaller than that observed by Yoon et al. (2008), and that a higher-powered study is needed in order to find an effect. Further studies on communicatively induced memory biases in infants could also investigate small methodological changes, in order to see under what specific scenarios the original finding holds and advance theory about what information infants encode in communicative and non-communicative contexts. However, due to the extremely high exclusion rate in the second experiment (70% of infants tested excluded from the replication analysis), this may be a very demanding and resource consuming challenge. As out of six experiments (table 1) only one has shown a specific identity memory bias induced by communication (as opposed to the loss of location memory), we believe that within this paradigm, the evidence against the experimental hypothesis is stronger than evidence for it, and we must consider the possibility that the original result may be a Type 1 error. In order to further study this hypothesis in a way that doesn't require more than 50% data loss, we suggest the development of a new paradigm.

Our attention allocation results are robust, with Experiment 2 completely replicating the results from Experiment 1. We found that at certain time points infants looked more to the face in communicative contexts than in non-communicative contexts, and more to the hand when it was reaching than when it was pointing. These results could be due to low-level perceptual differences between the two types of scene (e.g. with infants allocating their attention to where there is more movement), or to a high-level mentalistic interpretation of why infants would pay more attention to these areas (awaiting communication from the face in the communicative context, and awaiting the outcome of the reach in the non-communicative context). We believe there should be caution in attributing rich interpretations to phenomena that could also be explained by lean, attention-based interpretations (Haith, 1998; Heyes, 2016; Newcombe, 2002). We originally wished to investigate infant attention allocation in order to relate it to the observed memory biases. As we do not observe any differential memory biases for different contexts, we cannot relate these attention allocation differences to memory. What we can say is that attention allocation differences do not have an impact on what information is encoded or retained in our studies. Nonetheless, we feel that, when possible, eye-tracking data should be used in looking time studies to rule out attention allocation differences driving effects.

Table 1

Studies Investigating Communicatively Induced Memory Biases

Study	<i>N</i>	Effect size	Non-communicative	Communicative
-------	----------	-------------	-------------------	---------------

Yoon, Johnson & Csibra, 2008 – infants	24	$\square \eta p^2 = 0.41$	Location encoding	Identity encoding
Okumura, Kobayashi & Itakura, 2016 – infants	28	$\eta p^2 = 0.15$	Location & identity encoding	Identity encoding
Marno, Davelaar & Csibra, 2014 – adults	24	$\eta p^2 = 0.16$	Location & identity encoding	Identity encoding
Marno, Davelaar & Csibra, 2016 – adults	80	$\eta p^2 = 0.20$	Location & identity encoding	Identity encoding
The current paper: Experiment 1	24	$\eta p^2 = 0.15$	Identity encoding	Identity encoding
The current paper: Experiment 2	24	NA, but $BF01 = 5$	No encoding	No encoding

Study 2: Two-year old children preferentially transmit simple actions but not pedagogically demonstrated actions

Linking statement

In this study we attempted to explore the effect of ostensive cues on information transmission. Learning and teaching are two sides of the same coin, and it is likely that factors that affect learning in infants and young children also affect their own transmission of learned information. In fact, [Vredenburg et al. \(2015\)](#) investigated exactly this assumption. They found that 2-year-olds preferentially transmitted information that had been taught to them in the presence of ostensive signals. However, crucially, this could be the case for multiple reasons not specific to Natural Pedagogy theory. As emphasised in the introduction of this thesis, there is much evidence that infants preferentially attend to ostensive cues, with, however, less empirical evidence for more complex claims such as that this is because they enhance a referential expectation or that they prime infants to encode culturally relevant information. In this study, we pushed the boundaries of the predictions made by Natural Pedagogy.

We speculated that, if infants preferentially encode culturally relevant information following the presence of ostensive cues, and if this extends also to the transmission of information, then infants should preferentially transmit actions that have been taught to them in the presence of ostensive cues even if those actions are more complex than other actions taught to them without these cues. This is because

the infants should encode these actions as culturally relevant, and therefore worth teaching to a conspecific. When we instead found that infants preferentially transmitted the simple actions, we realised that without a direct replication of the original study, we would be unable to know what this result meant. As outlined in Study 1, replication is essential in order to build on a finding.

The study by Vredenburg and colleagues is a good example of a study requiring a replication effort: to our knowledge no studies have attempted to build on this finding until ours. Therefore, in the second experiment we conducted a direct replication of the original study and were unable to replicate the original finding. This study serves as another example of why, although it can be tempting to build on existing research, it is crucial to first be confident that the effect under investigation exists and is reliable. It is also important here to note how small changes to the design and method of a replication study do not per se disadvantage a replication effort, and instead can be combined with original evidence to suggest that perhaps after controlling for possible confounds, the effect is no longer present.

This study can be conceptualised as testing a related prediction to that seen in Study 1: the unique prediction stemming from Natural Pedagogy that infants preferentially encode culturally relevant, kind-generalisable knowledge. It is worth noting that Natural Pedagogy does not explicitly make the claim that factors affecting learning affect information transmission, but it is however a reasonable assumption, as any theory attempting to explain cultural information transmission must explain both how information is learnt and transmitted. It must be noted that the sample size for this study was decided based on the sample size of Vredenburg et al. (2015), which, while a common practice in developmental research to this date, leaves open the possibility that if there is a true effect being measured but it is smaller than that

reported in Vredenburg et al. (2015), we would have been unable to detect it. If redesigning this study, we would have used a power analysis to determine an appropriate sample size.

Abstract

Children are sensitive to both social and non-social aspects of the learning environment. Among social cues, pedagogical communication has been shown to not only play a role in children's learning, but also in their own active transmission of knowledge. Vredenburg et al. (2015) showed that 2-year-olds are more likely to demonstrate an action to a naive adult after learning it in a pedagogical than in a non-pedagogical context. This finding was interpreted as evidence that pedagogically transmitted information has a special status as culturally relevant. Here we test the limits of this claim by setting it in contrast with an explanation in which the relevance of information is the outcome of multiple interacting social (e.g., pedagogical demonstration) and non-social properties (e.g., action complexity). To test these competing hypotheses, we varied both pedagogical cues and action complexity in an information transmission paradigm with 2-year-old children. In Experiment 1, children preferentially transmitted simple non-pedagogically demonstrated actions over pedagogically demonstrated more complex actions. In Experiment 2, when both actions were matched for complexity, we found no evidence of preferential transmission of pedagogically demonstrated actions. We discuss possible reasons for the discrepancy between our results and previous literature showing an effect of pedagogical cues on cultural transmission and conclude that our results are compatible with the view that pedagogical and other cues interact, but incompatible with the theory of a privileged role for pedagogical cues.

Introduction

Social transmission of information is achieved through observation, imitation, and explicit teaching. The human capacity to flexibly engage learners in pedagogical contexts is at the core of transmission of complex cumulative human culture (Burdett et al., 2018; Caldwell et al., 2018; Kline, 2015). Although socially-mediated learning in early childhood has been studied extensively, little is known about child-initiated teaching (Nakao & Andrews, 2014; Sobel & Letourneau, 2015; Ziv & Frye, 2004) and its cognitive mechanisms (Corriveau et al., 2017; Paulus et al., 2015). Active transmission of information from children to others starts in infancy (Liszkowski et al., 2008) but the majority of studies on child-initiated teaching focus on preschool to primary school age children (Flynn et al., 2016; Flynn & Whiten, 2012; Whiten & Flynn, 2010), leaving a gap in understanding the developmental trajectory of information transmission. This emerging body of research shows children's sensitivity to what and whom they are teaching (Gweon & Schulz, 2019; Kim et al., 2016; Rhodes et al., 2015; Ronfard et al., 2016).

While research on information transmission in early childhood is limited, as they acquire knowledge, children are sensitive to both non-social and social aspects of the learning environment. Non-social factors include, among others, perceptual salience (Pruden et al., 2006), novelty (Mather & Plunkett, 2012), perceived causality (Burdett et al., 2018; Ronfard et al., 2016), efficiency (Pinkham & Jaswal, 2011), and complexity or difficulty (Bannard et al., 2017; Flynn et al., 2016; Kidd et al., 2012). Such salient factors guide integration of knowledge about objects, causal relations, statistical patterns, and updating and rejecting of hypotheses in light of acquired evidence (Bonawitz et al., 2012; Gopnik et al., 2017; Waismeyer & Meltzoff, 2017).

Social factors affecting learning include perceived intentions, competence, past accuracy, and social status of the informant or teacher (Harris, 2012; Over & Carpenter, 2012; Tomasello, 1999). Furthermore, pedagogical cues, such as infant-directed speech (Eaves et al., 2016), mutual gaze and joint attention (Striano et al., 2006), and explicit linguistic cues (Butler & Tomasello, 2016; Gelman et al., 2013), have been shown to selectively affect children's learning (Sage & Baldwin, 2011) in a way that is argued to not be able to be explained by heightened attention (but see Gredebäck et al., 2018; Szufnarowska et al., 2014).

Pedagogical cues have been shown to not only play a role in children's learning, but also in children's own active transmission of knowledge. Vredenburgh et al. (2015) showed that 2-year-olds were more likely to demonstrate an action to a naïve experimenter not present during the demonstrations after learning it in a pedagogical than in a non-pedagogical context. In their study, children were taught two possible actions on one toy. One action was shown in an explicitly pedagogical manner (involving infant-directed-speech and direct eye contact), whereas the other action was shown in an intentional but non-pedagogical way. Children were equally likely to learn both of these actions immediately following the demonstrations, but preferentially demonstrated the action that had been shown in a pedagogical context.

A variety of social learning theories propose that a human predisposition for learning from others, mediated by pedagogical cues, enables effective cultural transmission, thereby reducing the costs of unconstrained trial-and-error exploration. These theories predict that knowledge is most likely to be effectively transmitted when it was learnt as a cultural convention (Butler & Tomasello, 2016; Csibra & Gergely, 2009; Harris, 2012; Tomasello, 1999; Tomasello, 2016; Vygotsky, 1978). For instance, Tomasello (2016) argues for pedagogical instruction being one of the

pillars of cultural learning, with evidence accumulating that pedagogical contexts selectively enable children to generalize such information (e.g. Butler & Markman, 2012; Gelman et al., 2013).

While most pedagogical learning theories merely highlight the importance of pedagogical cues and are not specific about the relationship between these cues and other, non-pedagogical cues, the arguably most prominent such theory, Natural Pedagogy (hereafter NP: Csibra & Gergely, 2009; 2011), argues for a qualitative distinction between pedagogical and non-pedagogical cues. According to NP, humans have evolved to be sensitive to a circumscribed set of pedagogical cues (direct eye gaze, child directed speech, contingent reactivity), which directly create the expectation of communicative intention in the learner, circumventing the need to infer such an intention from other behaviours (Csibra, 2010). In this way pedagogical cues are thought to be a potent driver of infants' learning especially of culturally relevant knowledge that would be hard to acquire relying on observation alone, and their presence leads to the infant's effective encoding of such information. Nevertheless, as a privileged status for pedagogical cues in knowledge transmission is common to all described theories, we refer to them here collectively as *Privileged Pedagogy (PP)* views.

An interpretation of pedagogical cues that is in contrast to views ascribing a qualitatively distinct role to them is that, through their salience, they merely enhance attention to the learning situation without assuming infants' understanding of communicative and referential intentions behind them (Heyes, 2016, 2017). In this way, pedagogical cues interact with other situational factors to affect the likelihood of successful learning. In such a *Cue Combination (CC)* framework, each cue, weighted by its saliency, has an effect on the likelihood of some information being learned or

transmitted, and combined weights of the cues favor learning of one piece of information over another. Such an approach resembles, for instance, the emergentist coalition model of word learning (Hirsh-Pasek et al., 2000), which suggests that children attend to, weigh, and integrate a variety (coalition) of developmentally accessible cues (social, perceptual, cognitive, and purely linguistic) in learning new words. A combination of social and non-social cues has been shown to guide infants' attention, leading to effective learning outcomes in different settings (Barry et al., 2015; Smith et al., 2010; Wu et al., 2014; Wu & Kirkham, 2010). Extending this account beyond gradual weighted cue combination and presenting the social and non-social cue combination as a complex framework (Yurovsky & Frank, 2017), it has been proposed that domain general cognitive processes such as attention and speed of information processing explain developmental change in infants' use of social cues during word learning. Relatedly, rational constructivist approaches to social learning also argue for an integrative view, namely accounting for children's existing knowledge (Sobel & Kushnir, 2013) and statistical inference (Xu & Kushnir, 2013). Together, these approaches argue for embracing the complexity arising from real life learning situations, and for designing experiments which manipulate several competing cues while taking into account the different weightings of these cues.

Critically, while both *PP* and *CC* accounts assume a strong role for pedagogical cues in learning and transmission, they differ in how they explain the mechanisms underlying the effect of pedagogical cues. Whereas, for example, *NP* assumes that pedagogical cues have an evolved privileged and specific status in learning that is not based on their salience (Csibra & Gergely, 2006; 2009; 2011; Senju & Csibra, 2008; Yoon et al., 2008), others have argued that indeed the role of pedagogical cues is salience-based and have shown that non-pedagogical salient cues

may lead to the same effects (Gredebäck et al., 2018; Szufnarowska et al., 2014). Other work found that previous evidence for a privileged status of pedagogical cues may be difficult to replicate (Silverstein et al., 2019), and that both independently accumulated statistical evidence and an explicit pedagogical stance that highlights the importance of information affect the likelihood of successful learning (Buchsbaum et al., 2011).

Studies that are aimed at specifically evaluating the function of pedagogical cues typically contrast a pedagogical and a non-pedagogical condition with other cues held constant (e.g. Senju & Csibra, 2008; Vredenburgh et al., 2015), and can therefore be accounted for both within the *PP* and *CC* frameworks. In order to disentangle these two accounts, it is necessary to manipulate both pedagogical and non-pedagogical aspects of the learning situation, but few studies have taken this approach. Previous findings from studies pitting pedagogical against other cues have focused on the efficiency of an action for achieving a specific goal, and have been inconclusive: different studies found selective preference for pedagogical cues (Marno & Csibra, 2015), equal rates of imitation of pedagogically cued and efficient actions (Brugger et al., 2007), imitation of inefficient actions only after pedagogical demonstration but not after independent exploration (Pinkham & Jaswal, 2011), or effects of pedagogical cues in overriding children's propensity to over-imitate inefficient actions (Hoehl et al., 2014).

Building on Vredenburgh et al.'s (2015) finding that 2-year-olds are more likely to demonstrate an action to an adult after learning it in a pedagogical rather than non-pedagogical context, in the present study we set to investigate which of two potentially competing cues - action complexity as a non-social cue and pedagogical demonstration as a social cue - selectively affects the likelihood of an action being

shown to an ignorant adult. Although Vredenburgh et al. interpreted their results within a *PP* framework, we here aimed to directly evaluate *PP* vs. *CC* accounts by pitting an experimenter's pedagogically demonstrated complex action against another's non-pedagogically demonstrated simple action and observing which action was then preferentially transmitted by the child to a third, naïve experimenter. We focused on 2-year-old children to enable direct comparison between our results and those found by Vredenburgh et al.

While largely adopting Vredenburgh et al.'s (2015) experimental paradigm, we made several crucial changes to the procedure and some of the phrasing to provide a more stringent test of the pedagogical manipulation, by stressing the experimenter's direct, explicit teaching intention. Specifically, we changed the verbal prompt used by the experimenter performing the pedagogical demonstration from "Look! Do you see this?" to "Look at this! This is how you do it!", and the prompt used in the non-pedagogical demonstration from "I like this! Nice!" to the accidental discovery prompt: "What's this? Oh, that's how you do it". Both phrases could (and should) lead to social learning, but the difference lies in demonstrators taking the intentional and norm inducing pedagogical as opposed to intentional but non-directive non-pedagogical stances. We ensured that both demonstrators did not interact with children prior to the experiment to control for potential preferences for one over the other, while the third experimenter acting as an ignorant person, on the contrary, was familiar to them and strongly evoked the conventional, normative context by expecting action transmission from children (saying "Can you show me how to play with it?", which is a change from "What does this do? Can you show me?"). Collectively, these modifications were made to provide a more controlled manipulation and a more specific test of the effect of pedagogy above and beyond

normativity inferences made by children from observing demonstrations (Casler et al., 2009; Schmidt et al., 2016), while maintaining interactive engagement in the aim of ecological validity.

For the non-social cue, we manipulated action complexity, or difficulty of execution. Simple actions are faster to perform, require less effort, and have a higher probability of achieving the desirable outcome, and thereby, if we assume *CC*, action simplicity may outweigh the enhanced attention to more complex actions afforded by pedagogical cues. This reasoning is also in line with recent findings that older children transmit information to others in accordance with principle of utility calculus, making rational and efficient decisions based on the complexity of information and the perceived needs of the naive learner (Bridgers et al., 2019; Gweon & Schulz, 2019). Note that this approach is in contrast to studies manipulating the efficiency cue, where different manners of executing an action result in the same outcome (Marno & Csibra, 2015; Pinkham & Jaswal, 2011). We chose to instead manipulate complexity in order to ensure there were qualitatively different ‘pieces’ of information (i.e. actions with different outcomes) to learn and transmit.

In Experiment 1, we directly contrasted *PP* and *CC* explanations for children’s choice of action transmission by manipulating pedagogical cues and action complexity simultaneously. *PP* would argue that pedagogically transmitted actions should be transmitted preferentially as pedagogical cues hold a privileged status over other cues (Csibra, 2010; Csibra & Gergely, 2006). In contrast, from a *CC* perspective, pedagogical cues and action complexity should interact to lead to children’s preferred choice of action based on the relative weighting of the cues, so that pedagogical cues can be outweighed by information complexity. Then, in order to clarify the results of Experiment 1, in Experiment 2 we conducted a replication of

Vredenburg et al.'s (2015) manipulation of pedagogical cues while matching action complexity, retaining other methodological changes outlined in Experiment 1. Data, example videos, and supplementary results can be found on the Open Science Framework (OSF; osf.io/e2hvj).

Experiment 1

Method

Participants

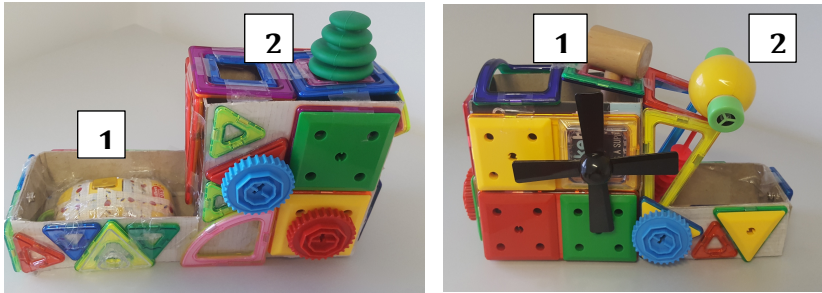
Thirty-one 24-month-old children from a local community in a small city in Northwestern England participated in the experiment (15 females, $M_{age} = 24$ months 2 days, $SD = .49$, range 23.28-25.25 months). Two more children were tested but not included in the final sample due to failure to perform any actions.

Stimuli

Two unfamiliar toys served as stimuli in two trials (see Figure 1a). Each toy had two target functions, each producing a unique attractive sound. The first toy included two different buttons. One of the buttons was visibly located on the surface of the toy and could be pressed with a finger (simple action). The second button was hidden inside the toy but could be pressed with a hammer-like tool slotted into the toy (complex action). The second toy included a spiral feature, which made a sound when squished down (simple action), and a shaker feature that needed to be removed from the object before shaking to produce the sound (complex action). Toys were designed specifically so that the actions were not obvious, with many distracting non-functional

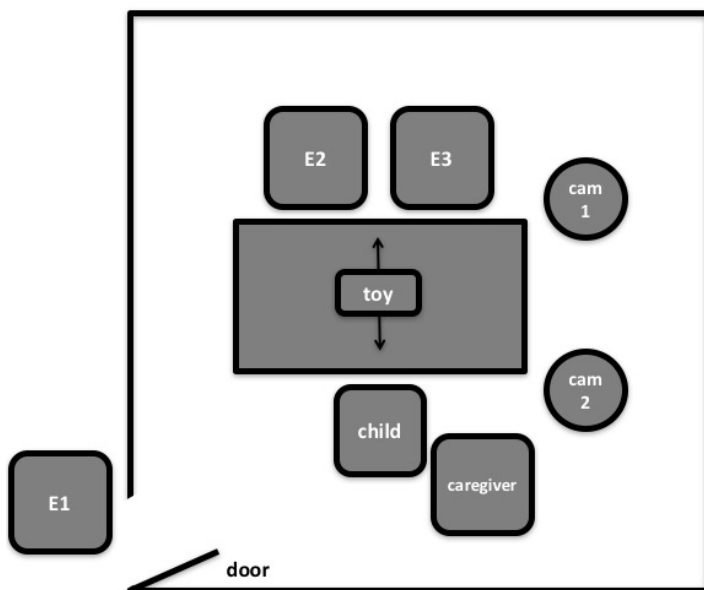
elements. All action outcomes produced comparable sounds that differed in the complexity required to elicit them.

1a: Examples of stimuli (toys used in Experiment 2)



Toy1: two simple actions (button press on yellow button at the bottom (1) and the green button on the top (2) to elicit the sound); Toy 2: two complex actions (wooden toy (1) is pushed through the opening at the top of the toy to press the button to elicit the sound; plastic toy (2) is pulled out and shaken to elicit the sound)

1b: Schematic of the experimental setup



1c: Experimental procedure

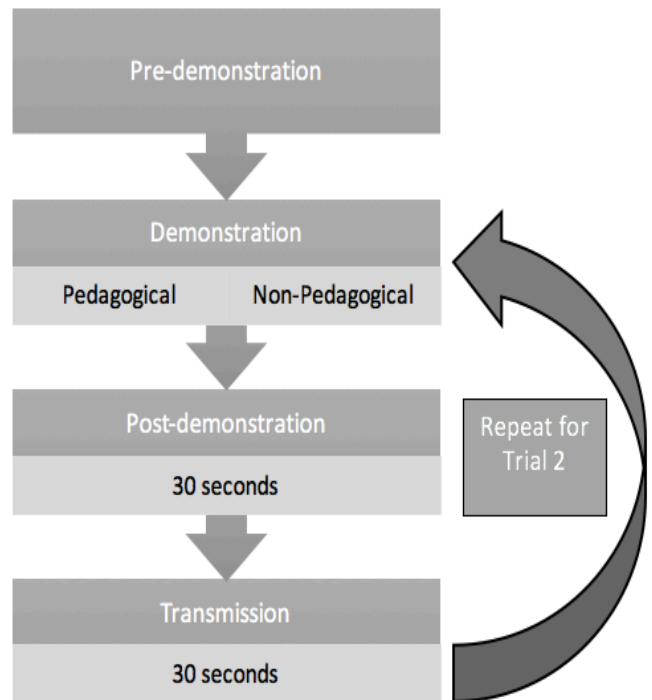


Figure 1. a) Examples of stimuli; b) Schematic of the experimental setup; c) Experimental procedure

Procedure

The experiment was divided into two phases: a Demonstration (including a Pedagogical and a Non-pedagogical demonstration and a post-demonstration) and a Transmission phase. Experimenter 1 (E1) briefly interacted with the child during a warmup before the Demonstration started. The child was seated in a highchair at a rectangular table, with the caregiver sitting slightly behind and to the side of the child (see Figure 1b). The caregiver was instructed not to interfere in any way. Following the warmup phase, E1 told the child that she would now leave the room and that her friends would like to come in with a new toy that she herself had never seen. She then left the room and closed the door. There were two consecutive trials, one for each toy. Each trial included all phases. The entire procedure was filmed using two video cameras.

Demonstration phase

Experimenters 2 and 3 (E2 and E3), previously unseen by the child, were assigned to either the pedagogical or non-pedagogical demonstration, with order and condition counterbalanced across trials and participants. Upon E1 leaving the room, E2 and E3 entered, bringing in the first toy. They sat side by side across the table, facing the child. E2 and E3 each demonstrated one action on the toy three times, either pedagogically (always the complex action) or non-pedagogically (always the simple action). Then (Post-demonstration) the experimenter who had just performed the action placed the toy in front of the child and asked: “Can you do it?” If the child

performed the demonstrated action within 15 seconds, the phase was finished, the toy returned to the other end of the table, the other experimenter demonstrated the other action and asked the child to perform it. If the child did not act after 15 seconds, the experimenter asked: “Can you show me?”. Throughout the demonstration, the other experimenter gazed downwards. Upon completing the demonstrations, both experimenters left the table and sat in the far corner of the room behind the child.

Pedagogical demonstration. The experimenter made frequent eye contact with the child throughout the demonstration and spoke in child-directed speech. After taking a look at the toy, the demonstrator looked at the child, establishing joint attention, and said excitedly: “Look at this! This is how you do it!”. She then demonstrated the complex function in a slow, deliberate manner, repeating it 3 times, alternating her gaze between the child and the object.

Non-pedagogical demonstration. The experimenter maintained attention on the toy, making no eye contact with the child. She looked at the toy and uttered in a self-addressed manner in adult-directed speech: “What’s this? Oh, that’s how you do it”. The demonstrator performed the simple function in a slow, deliberate manner and repeated it 3 times.

Transmission phase. E1 knocked at the door and re-entered the testing room, greeted the child and showed excitement upon noticing the toy on the table, saying: “Wow! Is that your new toy? That’s an interesting toy! I’ve never seen that before! Can you show me how to play with it?” If the child did not perform any actions after 15 seconds, the experimenter asked: “Can you show me what this toy does?” After 30 seconds, the experimenter thanked the child and took the toy away, leaving the testing room.

Coding

In order to be included in the final sample, children had to contribute at least one of the two experimental trials. Twenty-five children contributed both trials, and 6 children contributed one.

Trial exclusion. Six individual trials were excluded from final analyses due to children's failure to perform any actions ($N = 4$), experimenter error ($N = 1$), and parental interference ($N = 1$).

Post-demonstration. For each of the four post-demonstrations (two actions by two objects), we coded whether children performed the target action (hereafter called 'manipulation check') and whether this resulted in achieving the outcome, i.e. sound ('achieving the action outcome'). Note, we have used the terminology "performed" for actions that were attempted by the children regardless of whether they achieved the desired outcome, in contrast to "achieving the action outcome". We created two additional dichotomous variables ('exploratory behaviours'): accidentally discovering the non-target action during the first action's post-demonstration and performing both target actions during the second action's post-demonstration phase.

Transmission phase. The exact sequence of actions performed in each transmission phase (one per toy/trial) was coded, resulting in two measures: which action was shown first to the experimenter ('first action') and number of switches to each action type ('number of actions').

Inter-rater reliability. An independent blind coder performed the offline coding. Another researcher double-coded 20% of the data. We used Cronbach's α (for continuous variables) and Cohen's κ (for dichotomous variables) to assess coders' agreement, who were found to be highly reliable, $\alpha = .89$ and $\kappa = .84$. All disagreements were resolved in a joint discussion. Where the two coders differed in

their coding of continuous variables by over 50%, the video was jointly recoded and a final code produced. When the two coders' responses differed by less than 50%, an average was used.

Results

Statistical tests used were consistent with those by Vredenburg et al. (2015).

Post-demonstration

Trial order (and toy) did not have an effect on manipulation check (Pearson's $\chi^2 = 2.8, p = .09$) and on achieving the outcome of the action ($\chi^2 = .85, p = .36$), we therefore combined across these factors for further analyses.

Manipulation check. Children passed the manipulation check on 86% of trials (i.e. performed the action after demonstration), and this did not differ between simple non-pedagogically taught actions and pedagogically taught complex actions (McNemar's $p = .69$).

Achieving the action outcome. Children achieved the action outcome on 57% of trials, and were significantly more likely to achieve the outcome of simple action (this happened in 98% of the trials) than complex action (only 15% of trials), Pearson's $\chi^2 = 46.56, p < .001$, despite successfully passing the manipulation check in both types of actions.

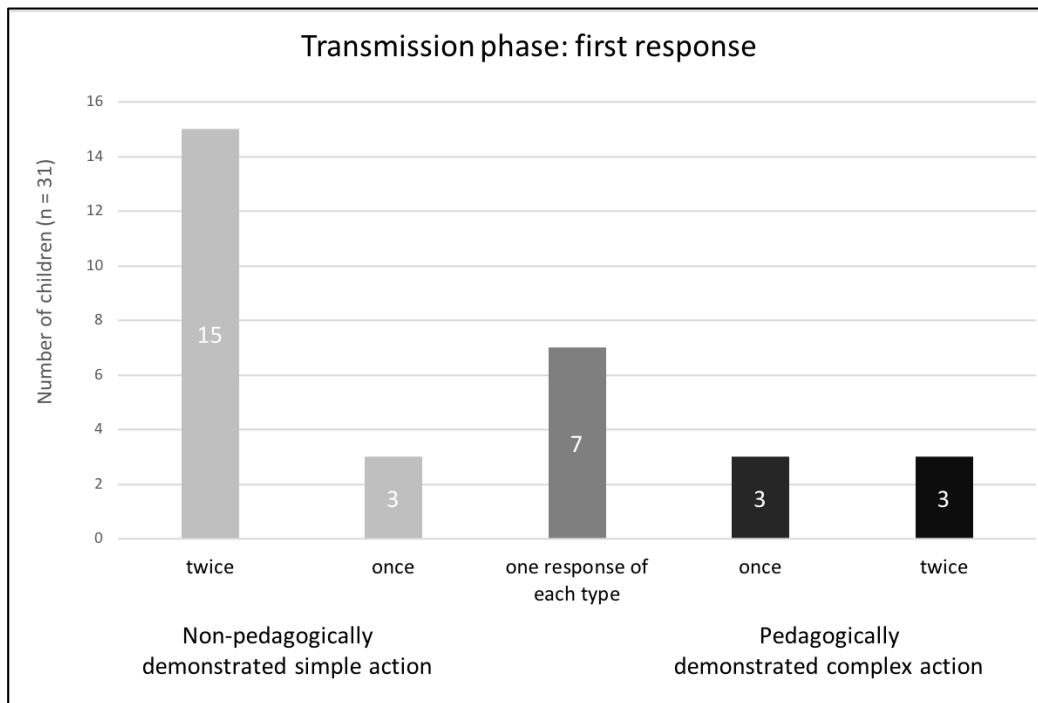
Exploratory behaviours. In 48% of post-demonstration trials, children accidentally discovered the non-target action, both simple ($N = 31$) and complex ($N = 22$), Pearson's $\chi^2 = 3.9, p = .05$.

Transmission

To test which action the children preferentially selected during this phase we examined two main variables: the choice of the first action and the number of actions.

First action. First actions from both trials were converted into scores: +1 (only demonstrated complex action first), 0 (one demonstration of each action first), and -1 (only demonstrated simple action first). These scores were compared to chance (0). Children performed the simple non-pedagogically demonstrated action first significantly more than they performed the complex pedagogically demonstrated action first ($t(30) = 2.68, p = .01, 95\% \text{ CI } [-.68, -.09], \text{ Cohen's } d = .49, \text{ paired t-test}$), see Figure 2a.

2a: Experiment 1: distribution of first responses at the transmission phase across the two trials



2b: Experiment 2: distribution of first responses at the transmission phase across the two trial

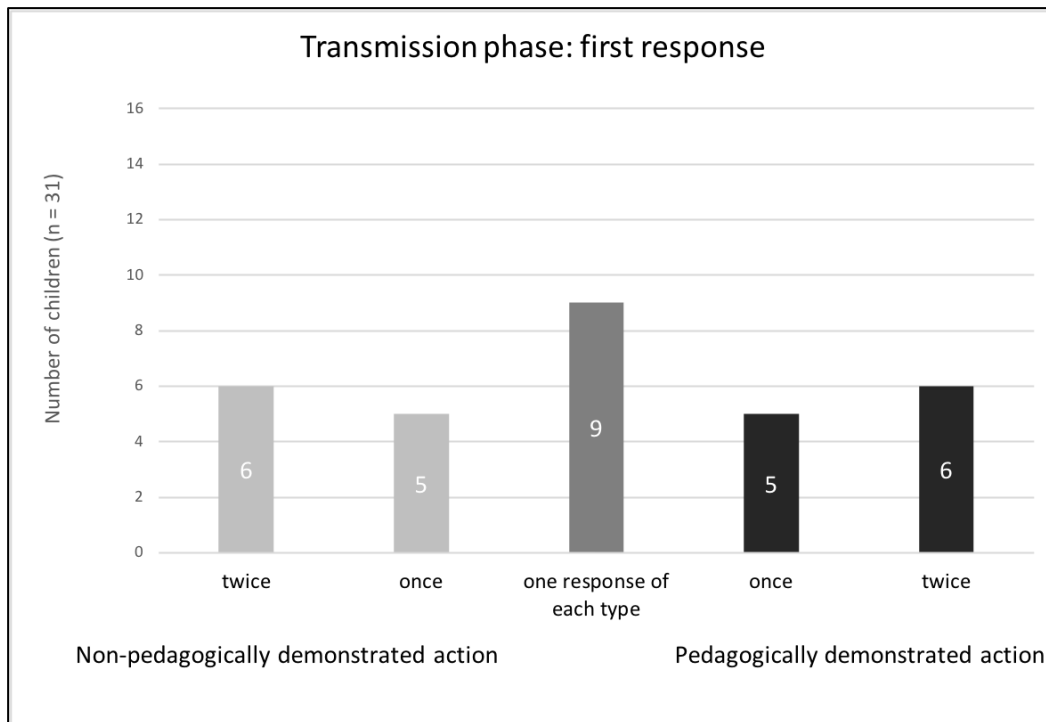


Figure 2. Distribution of first responses at the transmission phase across the two trials in Experiment 1 (a) and Experiment 2 (b). From left to right, the responses are ordered by frequency and exclusivity; the number of children whose first response was only the non-pedagogically demonstrated action (both times or once only), to those whose first response was mixed (one of each type), to those whose first response was the only pedagogically demonstrated action.

Number of actions. The majority of children (79%) performed both actions in both trials. Children performed a higher number of simple ($M = 1.61$, $SD = .84$) than complex ($M = 1.23$, $SD = .5$) actions, $t(30) = -2.47$, $p = .02$, 95% CI $[-.71, -.07]$, Cohen's $d = .55$, paired t-test, see Figure 3.

Exploratory behaviour analyses. The choice of the action transmitted first was not affected by accidental discovery of the non-target action not shown in the demonstration phase (trial 1: Pearson's $\chi^2 = 1.81$, $p = .18$; trial 2: Pearson's $\chi^2 =$

1.54, $p = .21$), nor by whether the child performed both actions at the transmission phase (trial 1: Pearson's $\chi^2 = .05$, $p = .82$; trial 2: Pearson's $\chi^2 = 1.17$, $p = .28$).

Experiment 1 discussion

We found support for children's preferential transmission of the simple non-pedagogically demonstrated actions over the pedagogically demonstrated complex actions with two converging measures. Children transmitted the simple non-pedagogically demonstrated action first more often, and they performed more of these actions. Our complexity manipulation was successful, in that children found it difficult to achieve the outcome of the complex actions, which is considered a good manipulation of complexity in other studies of information transmission (Whiten & Flynn, 2010). This does, however, make it possible that the actions were so complex that children did not preferentially transmit them due to developmental motor skill deficiencies; it has been previously found that toddlers will not imitate an action that they cannot motorically achieve (Paulus et al., 2011, but see Nielsen, 2006). Crucially, however, in our study, overwhelmingly children transmitted both actions despite preferentially (i.e., first) transmitting the simple action. Likewise, there was no relationship between infants' achieving the action at post-demonstration and later transmitting it: children's preferential transmission of the complex action was not predicted by their successfully achieving the outcome: of the 16 children who transmitted the complex action first, 4 had achieved its outcome at post-demonstration but 12 had not.

These results (in combination with those by Vredenburg et al., 2015) are best understood within a *CC* account. When pedagogical cues were manipulated alone, actions demonstrated pedagogically rather than non-pedagogically 'won'

(Vredenburgh et al., 2015). However, as shown here, when pedagogically demonstrated complex actions were pitted against non-pedagogically demonstrated simple actions, simple actions ‘won’, suggesting that enhanced salience resulting from pedagogical cues is weighted against action simplicity in children’s choice of which action to transmit preferentially. This is not to say that it is impossible that even under a CC framework, pedagogical cues could not outweigh simplicity.

However, with our design we could not tell whether our results showed that the properties of the simple action *overrode* the effect of pedagogical demonstration, or instead whether we, in contrast to Vredenburgh et al. (2015), might find that in our set-up pedagogical teaching does *not* lead to preferential transmission in the first place. To this end, we decided to conduct a replication of the study by Vredenburgh et al. (2015). We used the same procedure as in Experiment 1, except that action complexity was now matched for both pedagogically and non-pedagogically demonstrated actions in each trial. We thus had one trial with two simple actions (as in Vredenburgh et al., 2015) and the second trial with two complex actions, allowing us to compare how children perform on both pairs at the transmission phase (note that the trial with the simple actions always occurred first in order to ease children into the procedure).

Experiment 2

Method

Participants

Thirty-one 24-month-old children participated in this experiment (14 females, $M_{\text{age}} = 24$ months 4 days, $SD = 0.36$, range 23.28-24.69 months). Three more children

were tested but not included in the final sample due to failure to perform any action at post-demonstration or transmission phases ($N = 1$) and parental interference ($N = 2$).

Stimuli

Stimuli were the modified toys from the Experiment 1, where the two functions of each toy were now matched for complexity, with the toy used in the first trial always operated with two simple actions, and the second toy (second trial) operated with a pair of complex actions (see Figure 1a).

Procedure and coding

The procedure was identical to Experiment 1, bar the difference in manipulation (i.e. now actions were matched for complexity rather than contrasted by complexity in each trial). We intentionally did not counterbalance the order of the simple and complex trials so that children would be eased into the procedure with the simple trial. For this reason, we did not compare simple and complex trials statistically, as this manipulation was confounded with trial order. Twenty-two children contributed both trials, and 9 children contributed one of the trials.

Trial exclusions. Nine individual trials were excluded from final analyses due to failure to perform any actions ($N = 5$) and experimenter error ($N = 4$).

Coding. The experimental procedure and coding were identical to Experiment 1, except for the addition of action duration as an outcome measure previously shown to be sensitive to this manipulation (Vredenburgh et al., 2015). Thus, the duration of actions performed by the children was manually coded offline frame by frame, from the first frame of the child touching the functional part of the toy. This measure was not reported in Experiment 1 due to the inherent difference in time needed to

complete complex and simple actions (these results can be found in Supplementary Materials).

Inter-rater reliability. The coders achieved high reliability, with continuous variables producing Cronbach's $\alpha = .87$ and dichotomous variables resulting in Cohen's $\kappa = .90$.

Results

Post-demonstration

Manipulation check. Children passed the manipulation check on 97% of trials, and the pass rate did not differ between the two actions (McNemar's $p = .69$; trial 1/simple: 93% of children passed the manipulation check; trial 2/complex: 86% of children).

Achieving the action outcome. Children were equally able to achieve the outcome of the pedagogically and the non-pedagogically demonstrated actions (McNemar's $p = 1$; trial 1/simple: 96% of children achieved the simple actions outcomes; trial 2/complex: 31% achieved the complex action outcomes).

Duration. There were no significant differences in how long children spent executing the target actions in both conditions ($t(51) = .33, p = .74$, paired t-test).

Exploratory behaviours. In 40% of post-demonstration trials, children accidentally discovered the non-target action, but this did not differ between pedagogically demonstrated actions and non-pedagogically demonstrated actions in terms of both prevalence (McNemar's $p = .52$), and duration of time spent on these actions ($t(50) = .78, p = .44$, paired t-test).

Transmission

First action. Children did not perform the pedagogically demonstrated action first significantly more than the non-pedagogically demonstrated action overall ($t(30) = .00, p = 1.0$, paired t-test; Figure 2b) or separately for either trial (trial 1/simple: $t(30) = .00, p = 1.0$; trial 2/complex: $t(30) = .00, p = 1.0$). No statistical inference can be derived from this non-significant result (Lakens et al., 2018). Collapsed across trials, a default Bayes factor (Rouder et al., 2012) with a wide Cauchy distribution (scale of effect = 0.707) yielded $BF_{01} = 3.86$. We can conclude that the data constitute moderate evidence for the null hypothesis (Jeffreys, 1961).

Number of actions. The majority of children (81%) performed both actions in both trials. The number of pedagogically and non-pedagogically demonstrated actions did not significantly differ overall ($t(30) = .40, p = .69$, paired t-test), or separately for either trial (trial 1/simple: $t(27) = -.25, p = .80$; trial 2/complex: $t(27) = .24, p = .81$), see Figure 3. Collapsed across trials, a default Bayes factor with a wide Cauchy distribution (scale of effect = 0.707) yielded $BF_{01} = 3.69$, suggesting moderate evidence for the null hypothesis. Overall, children performed more actions during trial 1 (simple) than during trial 2 (complex): $t(21) = 3.91, p < .001$; 95% CI (.81, 2.65).

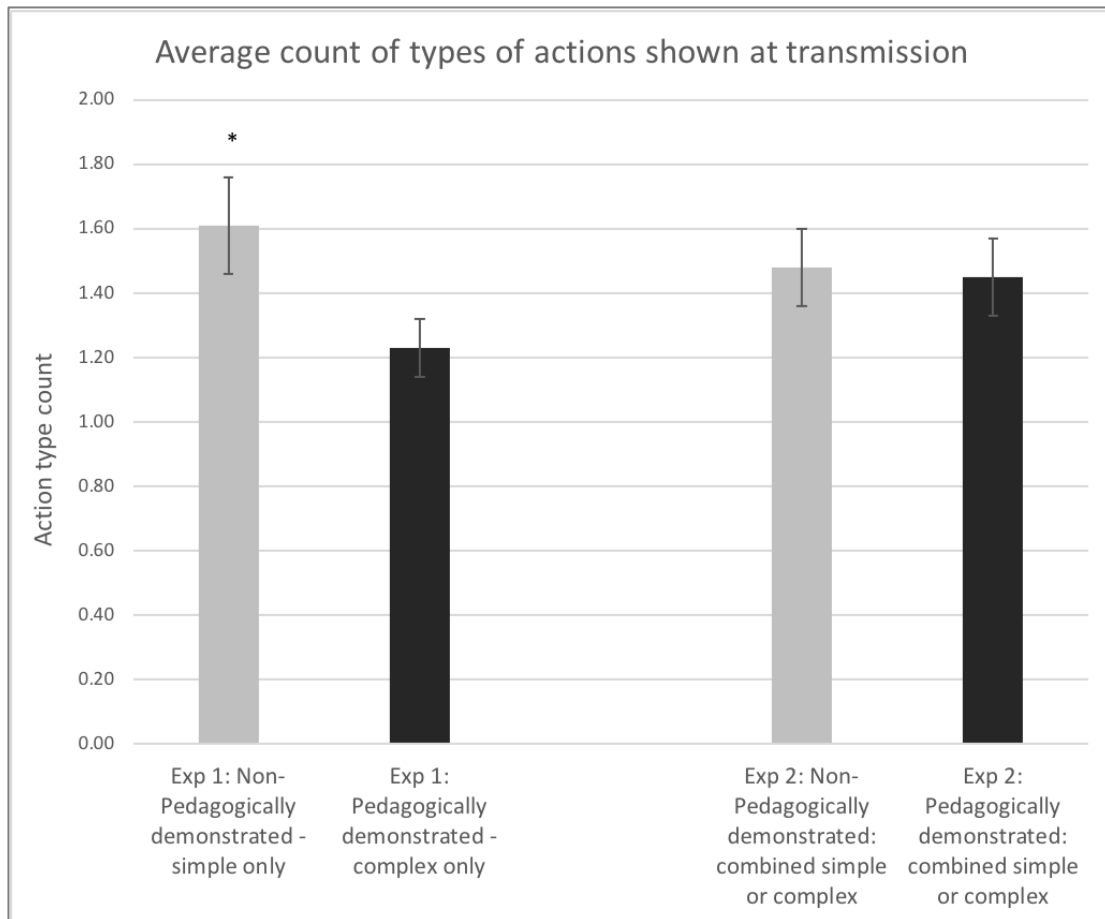


Figure 3: Average count of different types of actions shown at transmission phase during both experiments, including pedagogically demonstrated actions and non-pedagogically demonstrated actions. *Note:* Bars indicate standard errors; * $p < .05$

Duration of actions. Children did not spend significantly longer performing either of the actions overall ($t(30) = .19, p = .85$, paired t-test), or separately for either trial (trial 1/simple: $t(27) = -.50, p = .62$; trial 2/complex: $t(24) = .22, p = .83$).

Collapsed across trials, a default Bayes factor with a wide Cauchy distribution (scale of effect = 0.707) yielded moderate support for the null hypothesis, $BF_{01} = 3.79$.

Exploratory behaviour analyses. Whether or not a child accidentally discovered the action that was not shown to them in the demonstration phase (trial 1: Pearson's $\chi^2 = 0.70, p = .40$; trial 2: Pearson's $\chi^2 = 0.71, p = .79$), and whether or

not a child performed both actions at the transmission phase did not affect which action they chose to transmit first (trial 1: Pearson's $\chi^2 = 0.00$, $p = 1.00$; trial 2: Pearson's $\chi^2 = .52$, $p = .47$).

Experiment 2 discussion

Contrary to the results by Vredenburg et al. (2015), we found no evidence of preferential transmission of pedagogically demonstrated actions even when both actions were matched for complexity, indexed through first action, duration of actions, or number of actions. This overall result was also found separately for trials with two matched simple actions and two complex actions. This finding is supported by Bayes Factor Analysis, which shows moderate support for the null hypothesis.

Critically, even in the second trial, where both actions were complex, children still transmitted both actions, despite the outcomes only being achieved by 31% of children at post-demonstration and 40% of children at transmission phase.

Importantly, ignoring all possible effects of our complexity manipulation, even in the simple trial (which is comparable to the actions used in Vredenburg et al.), we saw no preferential transmission of the pedagogically demonstrated action.

There are several possibilities for why we did not replicate Vredenburg et al.'s (2015) finding. First, it is possible that our finding is a Type 2 error. However, this is unlikely given Bayes Factor Analysis showing moderate support for the null hypothesis across the range of measures used, and very similar results across the simple and complex action trials. We stress that the changes to the pedagogy manipulation that we made in our design, as compared to Vredenburg et al., aimed for an even stronger and more stringent comparison of pedagogical versus non-pedagogical demonstration, making us more likely to detect the differences between

the manipulations. It is also possible that the original finding was a Type 1 error. As no other studies to date have replicated this finding, this remains a possibility. Future research could investigate the subtleties required for reproducing this effect, and whether the underlying theory should be modified accordingly to include these specific constraints.

Another possibility is that there are key differences between Vredenburg et al.'s (2015) and our experiment. For example, it is possible that our verbal prompts in effect meant that both conditions were considered normative by the children, in that both prompts contained "this/that is how you do it". In our attempt to create a more controlled manipulation of pedagogical cues, whereby the only difference between prompts was the pedagogical nature, we may have inadvertently created a more normative context for both conditions. Nevertheless, normativity is not the defining feature of pedagogical context, and in *PP* a pedagogical transmission of normative information would still be preferred over non-pedagogical transmission of equally normative information (Csibra, 2010; Csibra & Gergely, 2009). The defining features of pedagogical context (which we did manipulate) are the presence of ostensive cues and explicit teaching. Future research can disentangle the effect of pedagogical context and normativity in information transmission.

Within the proposed *CC* framework, it is also possible that differences in the specific pedagogical cues and specific actions used in the two studies alter the specific weighting of cues. In our Experiment 2, the inherent interest of the actions may have outweighed the salience difference introduced by manipulating pedagogical cues, whereas in the study by Vredenburg et al. this was not the case. An analogy to illustrate this possibility is a child receiving two presents for their birthday, one accompanied by 'pedagogical' cues ('Wow, look at this, a present for you!'), and the

other just placed in front of the child. The child might be so excited about the presents per se that the manner (social context) in which they were delivered makes no discernible difference to their interest in them.

Finally, it is possible that, in contrast to Vredenburgh et al.'s (2015) finding, pedagogical cues do not modulate preferential transmission of actions. In light of strong evidence that explicit pedagogical cues are of special importance to children in a variety of learning situations (Butler & Markman, 2012, 2014, 2016; Csibra & Gergely, 2009; Marno & Csibra, 2015) this would be surprising. Nevertheless, Vredenburgh et al.'s study was the first to extend the notion of the importance of pedagogical cues from learning to a child's active transmission of information, and it is possible that factors that are important in learning do not necessarily translate to information transmission, or do so at a later age than currently investigated. Future research should therefore focus on examining the contexts under which different cues combine to affect transmission.

General discussion

In this study we investigated how two types of cues (pedagogical demonstration and action complexity) affect children's transmission of recently demonstrated actions to a naïve adult, testing the prediction of a Privileged Pedagogy (*PP*) account that pedagogical cues would ensure preferential encoding of the action even despite its higher complexity, against a Cue Combination (*CC*) view that both pedagogical and complexity cues would be weighted to determine which action was preferentially transmitted. We report evidence that children preferentially transmitted simple non-pedagogically taught actions over pedagogically taught complex actions (Experiment 1), and no evidence of preferential transmission of pedagogically

demonstrated actions when both actions were matched for complexity (Experiment 2), with moderate support for the hypothesis that pedagogical cues have no effect on preferential transmission.

Our results are compatible with the *CC* view, and incompatible with the *PP* view. The *CC* framework can explain this pattern of results, and the difference of our results compared to those of Vredenburg et al. (2015), if as well as the cues that we purposefully manipulated (action complexity and pedagogical demonstration) we also take into account the intrinsic properties of the toy. It is possible that compared to Vredenburg et al. (2015), the inherent salience of our toys and our actions was higher. In this way, even if pedagogical demonstration did have some weighting, this did not incur a big enough difference to affect children's preference for action transmission. This is in contrast to Experiment 1, where there are intrinsic action properties that enhance the likelihood of transmission of the simple action. However, we note that although our results are *compatible* with the *CC* view in this way, future research should quantify and examine in more detail the relative weightings of different cues in determining outcomes for learning and transmission.

Several features of our study and that by Vredenburg and colleagues warrant discussion and further research. First, it is possible that children were not sensitive to the pedagogy manipulation and instead considered the whole demonstration phase as a uniform learning episode, where one demonstrator is ostensibly more communicative than the other. Although we manipulated pedagogy explicitly, using various established cues, a more extreme manipulation of pedagogical and non-pedagogical cues may help resolve this consideration (e.g. Marno & Csibra, 2015). However, this may be at the expense of ecological validity, and would add further

confounds that could be responsible for any differences found (as opposed to the presence of pedagogical cues themselves).

As we chose to manipulate the complexity of the actions by varying their ease and transparency of execution, this resulted in the complex actions being harder to achieve for some of the participants. Although this raises the possibility that it discouraged some children from choosing to transmit complex actions regardless of the manner in which they were demonstrated, three main points speak against this interpretation. In Experiment 1, in trials where children transmitted the complex action preferentially (16), they were no more likely to have been able to achieve the outcome at post-demonstration than not (4 achieved, 12 did not). In Experiment 2, the results for the trial in which both actions were complex showed that even with low achievability, infants still transmitted both actions. In the trial where both actions were simple (comparable to Vredenburg et al., 2015), children showed the same pattern: there was no difference in their transmission of pedagogically and non-pedagogically demonstrated actions. Hence, there is no evidence that being unable to achieve the outcome of the complex action affected children's transmission choice. Future research should investigate the relationship between the ability to achieve the outcome of an action and its transmission, with a large enough sample to statistically examine these individual differences.

Another promising avenue for future research is cultural transmission of different types of information (Corriveau et al., 2017; Flynn & Whiten, 2012; Paulus et al., 2015; Ronfard & Harris, 2018). In the present study, we investigated easy to acquire, developmentally appropriate, rewarding, causally unambiguous, accurate, and easily verifiable information. It is an open question whether the mechanisms for transmission of this type of information would be the same as for information that is

causally opaque, inefficient or irrelevant (e.g. Burdett et al., 2018; Corriveau et al., 2017; Lyons et al., 2011; Ronfard et al., 2016), socially conforming and more frequently endorsed (e.g. Morgan et al., 2015), or tabooed (e.g. Seehagen et al., 2017). It would also be important to further investigate what is meant by transmission in this paradigm, i.e. whether the children are demonstrating what they can do, or that they know the rules, to show what they themselves have learnt, or to explicitly teach someone else. These are among the exciting questions open for investigation.

Like Vredenburg et al. (2015), we focused on 2-year-olds as the youngest group previously shown to engage in active cultural information transmission, yet a rarely studied population compared to a large body of research with preschool and elementary school children. We still know very little about the developmental trajectory of cultural information transmission in children as they navigate the social world, as well as the extent of the claims made by the *PP* accounts in toddlers, rather than infants or preschoolers, which should be addressed in further research. As we set out to investigate children's selectivity in social learning, we would like to bear in mind that within a complex, dynamic, self-organizing developmental system (Smith & Thelen, 2003), mechanisms are constantly changing. It is plausible that while we showed the parsimonious outcome of the present study's manipulation with two-year-olds, later in development this may change. Understanding when and why such a change occurs is an avenue open for further research.

Overall, our study provided an ecologically valid manipulation of pedagogical cues and evaluated two competing accounts of their role in the learning and transmission of information in toddlers. In our paradigm, children were not influenced by the context of the demonstration, but instead by the properties of the actions themselves, favouring actions that are easy to perform, require less effort, have a

higher probability of achieving the desirable outcome, and result in a rewarding sound most readily. What we can conclude is that we do not have support for *PP*, as this view would assume that pedagogical cues should not only have a positive effect on transmission when manipulated alone, but that they should also override other cues, neither of which we were able to show in our two experiments. This study contributes to our understanding of the factors affecting information transmission in early childhood.

Study 3: Infants learn to follow gaze in stages:

Evidence confirming a robotic prediction

Linking statement

We have seen in the previous two chapters that we have not found any evidence supporting the third claim of NP – that infants expect information that is communicated to them in ostensive-referential contexts to be kind-relevant and generalisable. From our review of the literature in the Introduction chapter, we are confident in the first claim of NP – that infants are sensitive to ostensive signals. However, the second claim – that ostensive signals trigger a referential expectation in infants – is still up for debate. After two studies failing to show support for NP, we speculated about alternative possibilities for how infants may observe and respond to ostensive referential cues. In this paper, we explored a different avenue for theory development and testing. The previous papers in this thesis attempted to directly test NP (Study 1), or pit NP against alternative theories in the developmental literature (Study 2). Our predictions for the following study came from a very different direction. In this study, we looked to the developmental robotics literature, positing that if a specific behaviour has been able to be successfully instantiated in a developmental robotic architecture, it is possible that the mechanism implemented in the robot is the same as the mechanism in human infants.

Specifically, we investigated gaze-following ability – one of the key elements of joint attention and social cognition. NP argues that gaze following in young infants occurs because of the referential expectation induced by observing ostensive cues

(Senju & Csibra, 2008), however recently this has been shown to be able to be explained through low-level processes, as discussed in the Introduction chapter (Gredebäck et al., 2018; Szufnarowska et al., 2014). Alternative psychological accounts (Butterworth, 1991; Corkum & Moore, 1998; Moore et al., 1997; Moore & Corkum, 1994) and the computational modelling literature (Nagai et al., 2006; Triesch et al., 2006) instead suggest that even without understanding or expectation, infants could learn to gaze-follow through social reinforcement. When implementing a social reinforcement model of gaze following in a developmental robot, the robot produces previously unexpected behaviours that can then be tested in infants, as we do so in this paper. When using computational modelling to generate and refine theory, verbal theories are converted into tentative formal models, which are implemented in a developmental robotic architecture. If the robot displays human-infant-like learning, it is possible that the formal model is a good approximation of the way that humans learn this same behaviour. The behaviour of the robot is used to make predictions for what we should see in human infants if this is the case. The behaviour of the infant is then in turn evaluated against the theory: does the human behave like the robot? If so, it is possible that this formal model is sufficient for learning the behaviour. If not, the model needs to be refined and perhaps modules added in order to accurately reflect human learning. Importantly, it is possible that the behaviours that the infant and the robot produce, although looking similar, have different origins. However, if we are to assume this to be the case, it is necessary that an alternative theory be developed and tested for why infants show this behaviour.

Abstract

Gaze following is an early-emerging skill in infancy argued to be fundamental to joint attention and later language development. However, how gaze following emerges is a topic of great debate. Representational theories assume that in order to follow adults' gaze, infants must have a rich, innately specified understanding of adults' communicative intention. In contrast, learning-based theories argue that infants may learn to gaze follow based on low-level social reinforcement, without the need to understand others' mental states. [Nagai et al. \(2006\)](#) successfully taught a robot to gaze follow through social reinforcement and found that the robot learned in stages: first in the horizontal plane, and later in the vertical plane – a prediction that does not follow from representational theories. In the current study, we tested this prediction in an eye-tracking paradigm. Six-month-olds were unable to follow gaze in either the horizontal or vertical plane, whereas 12-month-olds and 18-month-olds only followed gaze in the horizontal plane. These results confirm the core prediction of the robot model, suggesting that children may also learn to gaze follow through social reinforcement coupled with a structured learning environment. This study was preregistered and data, code, and materials are available on the Open Science Framework (OSF; osf.io/fqp8z).

Introduction

Gaze following, or the ability to look where a social partner is looking, is a critical milestone in human sociocognitive development (Tomasello, 1995). Although gaze following has been found in many non-human species (e.g. Bugnyar et al., 2004; Okamoto-Barth et al., 2007; Povinelli et al., 1996; Tomasello et al., 2007), in human infants this skill has been argued to be a fundamental component of socialization, and

infants' early gaze following has been linked repeatedly to their later language development (e.g. Brooks & Meltzoff, 2005; Carpenter et al., 1998; Morales et al., 2000). Gaze following appears to emerge from around five months of age (Gredebäck et al., 2018; Senju & Csibra, 2008; Szufnarowska et al., 2014) including in non-WEIRD (Western, educated, industrialized, rich, and democratic) populations (Hernik & Broesch, 2018). Thus, in human infants, gaze following appears to be an early-emerging, universal ability, which is intimately linked to our uniquely human social capacities.

While gaze following in infancy has been repeatedly demonstrated, the origins of this ability and the cognitive representations at play remain controversial. Some influential representational theories assume that gaze following relies on infants' innate capacity to be receptive to others' communicative intentions or mental states (e.g. Baron-Cohen, 1995; Csibra & Gergely, 2009). On these accounts, infants are born with the understanding that adults are intentional communicative agents, and based on this understanding, look where adults look in order to obtain information. In support of these accounts are studies that suggest infants may have some rudimentary form of gaze following (gaze cueing) from birth, evidenced by the finding that newborns detect an object on the screen faster if it appears in a location previously cued by another's gaze (Farroni et al., 2004). Others argue that although this is not an ability that is present from birth, at a certain time point the ability to read others' intentions and hence follow gaze 'switches on', supported by the finding that younger infants need the presence of ostensive cues (direct gaze and infant directed speech) in order to follow gaze (Senju & Csibra, 2008; but see Gredebäck et al. 2018; Szufnarowska et al. 2014).

In contrast to these representational accounts, learning-based accounts assume that infants learn to follow gaze via reinforcement learning. Specifically, during

interaction, a caregiver may avert their gaze to look at something interesting entering the environment. If the infant by chance looks in the same direction, they are rewarded by the interesting information, and will be more likely to look in the same direction as the caregiver in the future (e.g. Triesch et al., 2006). In this scenario, social scaffolding by the caregiver, for example holding objects directly in the infant's field of view and rapid alternation of head turns from infant to object, may increase opportunities for learning (Moore & Corkum, 1994). Alternatively, infants may learn to follow gaze by initially learning that the caregiver's hands are a reliable cue to the focus of the caregiver's attention (Deák et al., 2014; Yu & Smith, 2013). Although the mechanisms proposed in learning-based theories differ, critically all share the assumption that, at least early in development, associative learning and reinforcement are sufficient to support the emergence of gaze following given a suitably structured learning environment – without needing to invoke a rich ability to understand others' mental states.

Recently, researchers have begun to address this debate by implementing the assumptions of these theories in computational and developmental robotic models. Notably, these models demonstrate that low-level processes can support the development of gaze following without the need for any understanding of intention. Rather, these simulations offer mechanisms by which infants may learn this skill through social reinforcement (Moore et al., 1997; Triesch et al., 2006). In the current paper we focus on the predictions of a developmental robotic simulation of the emergence of gaze following based on a combination of low-level learning mechanisms and social reinforcement. Nagai et al. (2006) successfully taught a developmental robot with a simple neural network cognitive architecture to gaze follow through supervised associative learning. The robot was equipped with a camera that fed images of a human

experimenter to a visual system consisting of a connectionist map, which encoded these images, and a retinal smoothing layer, simulating the development of infants' visual acuity. Training consisted of five steps. First, the robot's camera was focused on the experimenter's face. Next, the experimenter held up an object and shifted her gaze towards it. Visual input from the camera was then processed by the neural network after smoothing, generating a representation of the experimenter's face. Based on this representation, the robot generated a motor command, adjusting the joint angles in its head and neck, resulting in a head turn and a change in its visual field. The robot was then given feedback based on the output error between the location of the object in the visual field and its gaze direction: if the object was centred in the visual field, gaze following was considered successful and no adjustments to the neural network were made. When the object was outside the visual field or off-centre, random noise was added to the connection weights in the robot's neural network. Across training, therefore, head movements resulting in incorrect gaze following were less likely to be produced, increasing the relative strength of connections which produced correct gaze following.¹

Following training, testing with previously untrained images demonstrated that the robot learned to follow the experimenter's gaze. Thus, this work raises the possibility that low-level perceptual and proprioceptive information coupled with social reinforcement are sufficient to support the emergence of this important ability. However, testing during training revealed that in the robot, the ability to gaze follow developed in stages: early in learning, the robot initially learned to follow gaze in the horizontal plane, and only later in the vertical plane. Thus, this work makes the

¹ A formal description of the model is outside the scope of this chapter; for details see Nagai et al. (2006)

empirically testable prediction that infants too should initially follow gaze shifts more successfully in the horizontal than the vertical direction. Importantly, Nagai and colleagues' model implements a learning-based account of gaze following, and the same prediction does not follow clearly from representational theories that assume an ability to understand others' communicative intent. An important means to verify the validity of a computational model of developmental processes is to test if its predictions are borne out in studies with children (Asada et al., 2009; Morse & Cangelosi, 2017; Schlesinger & McMurray, 2012; Westermann & Mareschal, 2012). In the present study we therefore provide this critical test. Specifically, we ask whether, like the robot, infants learn to gaze follow in stages, acquiring horizontal gaze following before vertical gaze following. If this is the case, this would confirm a central prediction of the robot model. A pilot study showed that 12-month-old infants were able to follow the gaze direction in our stimuli and were not at ceiling or floor. This age group are therefore at an intermediate developmental stage in which gaze following is not consistently accurate, raising the possibility that at this age, we may observe the differences in horizontal/vertical tracking predicted by the robot. Therefore, in Experiment 1 participants were 12-month-old infants. In order to investigate the developmental trajectory of gaze following, we then tested 6-month-olds (Experiment 2) and 18-month-olds (Experiment 3). All preregistered hypotheses, materials, code, and data can be found on the Open Science Framework (OSF; <https://osf.io/fqp8z/>).

Experiment 1 (12-month-olds)

Method

Participants were recruited from a database of families at an infancy lab of a UK university, and were given a book as a gift for participation and £10 travel reimbursement. Parents gave informed, written consent before participation, and were free to withdraw their consent. All data were kept confidential. All experiments were approved by the university ethics committee and adhered to the British Psychological Society guidelines.

Participants

Sample size was determined using Bayesian sequential testing (Schönbrodt et al., 2017). In Bayesian sequential testing, Bayes factors are calculated after every participant (pending a pre-defined minimum number of participants) until a preregistered Bayes factor threshold is reached for all analyses run (or a pre-defined maximum number of participants is reached). We determined a minimum *a priori* sample size of 20 infants (set due to a recommended by Schönbrodt et al., 2017), with a maximum of 40 infants (set due to resource constraints), and set a threshold of a BF10 (i.e., evidence for the research hypothesis) and BF01 (i.e., evidence for the null) of 10 or above for all preregistered analyses. We did not reach this threshold for all tests when we reached our maximum number of infants. Therefore, our final sample consisted of 40 typically developing 12-month-old infants (mean age: 364 days; range: 349 days to 383 days; 27 female; all Caucasian; 34 monolingual English). Four additional infants were excluded due to fussiness, defined as providing less than 50% of usable trials where they were not crying. All infants were reported to have no developmental delays and no visual impairments.

Stimuli and design

Example stimuli are depicted in Figure 1. Each trial consisted of a three-second long video. Trials were split into control trials and experimental trials. Control trials consisted of a central fixation cross (1000 ms) followed by a novel object appearing in one of four locations 200 pixels left, right, up, or down from the centre of the screen (2000 ms), and were designed to test whether infants found gaze shifts (without gaze following) *a priori* easier in the horizontal than the vertical plane. Experimental trials consisted of a human face looking directly at the infant, surrounded by four exemplars of the same object in each of the locations (1000 ms) followed by averted gaze to one of the four exemplars (2000 ms). We used five photographs of one female face looking left, right, up, down, and directly at the camera. The eyes from the left, right, up and down photographs were superimposed onto the face looking directly at the camera, ensuring the face was identical apart from gaze direction. We selected objects from the NOUN database (Horst & Hout, 2016) that had no obvious top/bottom in order to avoid biasing infants' attention. In total, 64 videos were made, consisting of the eight objects in the four locations for both the control and experimental conditions. These videos were pseudorandomized into four orders such that infants never saw the same object, location or trial type (control, experimental) on more than two successive trials. Between trials there was a variable animated attention-getter. All videos are available on OSF.

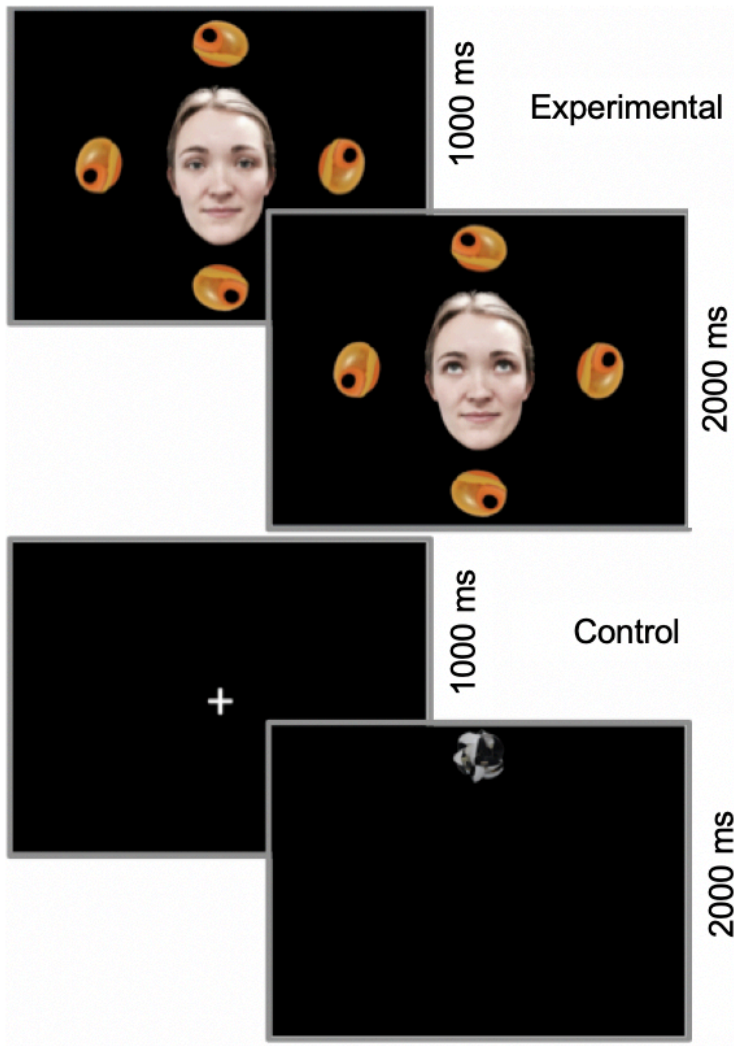


Figure 1. Experimental and control conditions. Central cue (direct gaze or fixation cross) for 1000 milliseconds, followed by a directional cue (averted gaze or object) for 2000 milliseconds.

Procedure

Infants sat on their caregiver's lap during the experiment in front of a 23-inch screen (seated approximately 0.6 meters away). An eye-tracker (Tobii X120) captured infant looking times and gaze locations on screen. We used Tobii Studio 3.3.1 to present stimuli and gather eye-tracking data. We performed a five-point calibration for all infants before beginning the experiment. After this calibration, we instructed caregivers

not to talk to or interact with their infant, and that they could stop the experiment at any time if the infant became too fussy, and the experiment began. Infants saw up to a maximum of 64 trials in one of the four pseudorandomized orders (range: 28-64, $M = 57$; $SD = 9$).

Analysis

We performed all analyses in R 3.5.2 (R Core Team, 2017). We exported raw data from Tobii Studio 3.3.1 and analysed them using a combination of the `eyetrackingR` package (Dink & Ferguson, 2015) and our own code (all code is available on the OSF). Areas Of Interest (AOIs) were defined for centre (400 x 400 pixels), and left, right, up, and down object locations (each 340 x 400 pixels). All preregistered analyses are reported either in the main paper or in supplementary results on OSF. Our outcome variables included two measures of accuracy: proportion looking (length of looking at target AOI divided by total looking at all four object AOIs); dichotomous target looking (whether or not the infant looked at the target AOI for any frames after the gaze shift); and a measure of latency (time taken for the infant to look at the target AOI). Corresponding Bayesian analyses were carried out for all frequentist analyses (Dienes & McLatchie, 2018). All Bayesian analyses used a default Bayes factor with a wide Cauchy distribution (scale of effect = 0.707) and were calculated using the `BayesFactor` R package (Morey & Rouder, 2015). For all Frequentist analyses, we used a significance threshold of $p < .05$. BF_{10} or BF_{01} of more than 3 are considered moderate evidence, using the system outlined by Jeffreys (1961). Data are visualized by plane; visualizations by position are available on OSF (Figures S1-S4).

Results

Accuracy

We submitted proportion looking to a 2 x 2 repeated measures ANOVA with main effects of condition (gaze vs. control) and plane (horizontal vs. vertical) (Figure 2). Accuracy was higher for control than gaze trials [main effect of condition: $F(1,39) = 1222, p < .001$; $BF_{10} = 2.93e^{63}$], and higher for the horizontal than the vertical plane [main effect of plane: $F(1,39) = 9.02, p < .005$; $BF_{10} = 3.08$]. The ANOVA revealed a significant interaction between condition and plane [$F(1,39) = 23.72, p < .001$; $BF_{10} = 1.65$]. Planned, two-tailed, paired t-tests revealed that in the gaze condition, accuracy was higher for the horizontal [$M = 0.34, SD = 0.15$] than the vertical plane [$M = 0.18, SD = 0.14$; $t(39) = 4.14, p < .001$; $BF_{10} = 2774$]. In contrast, in the control condition, there was no evidence for a difference in accuracy across the two planes [Horizontal: $M = 0.90, SD = 0.11$; Vertical: $M = 0.92, SD = 0.08$; $t(39) = -1.77, p = .08$; $BF_{01} = 2.35$]. Finally, paired t-tests against chance (0.5) revealed that in the gaze condition, accuracy was above chance for the horizontal plane [$t(39) = 3.72, p < .001$; $BF_{10} = 70$], and below chance for the vertical plane [$t(39) = -3.14, p < .01$; $BF_{10} = 15$]. Dichotomous target looking results mirrored the proportion looking results and can be found on OSF (Figure S5).

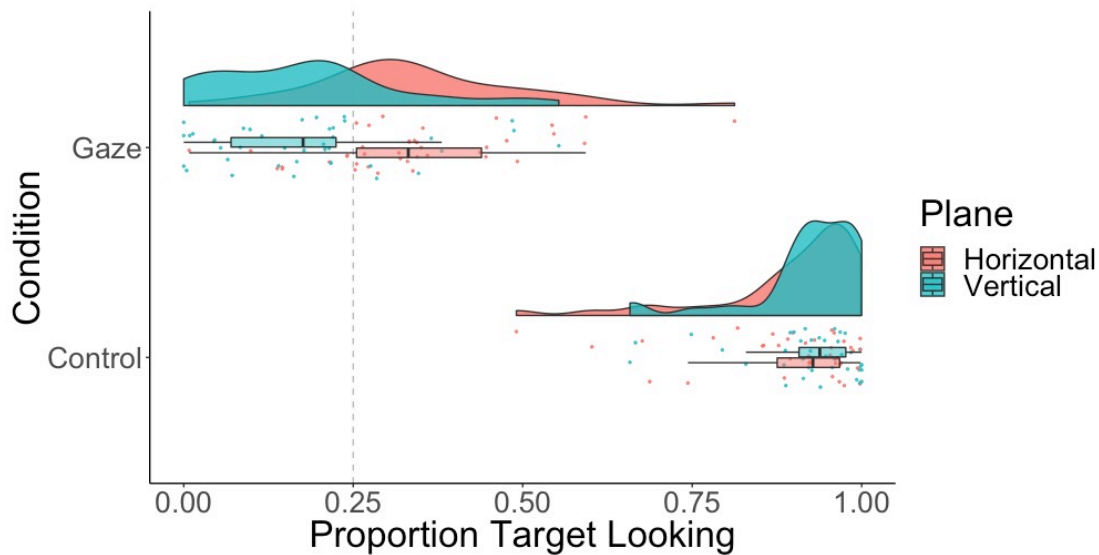


Figure 2. Raincloud plot showing proportion target looking across conditions (gaze vs. control) and planes (horizontal vs. vertical) for 12-month-olds. Dashed line indicates chance performance.

Latency

We submitted latency to a 2 x 2 repeated measures ANOVA with main effects of condition (gaze vs. control) and plane (horizontal vs. vertical) (Figure S6). The ANOVA revealed moderate evidence for no interaction between condition and plane for latency [$F(1,39) = .02, p = .90; BF_{01} = 5.66$]. Latency was shorter for control than gaze trials [main effect of condition: $F(1,39) = 19.45, p < .001; BF_{10} = 283$], and there was moderate evidence for no difference in latency in the horizontal and vertical planes [main effect of plane: $F(1,39) = .01, p = .93; BF_{01} = 5.71$].

Discussion

In this experiment we tested the predictions of Nagai et al.'s (2006) robotic implementation of the development of gaze following, specifically that infants in the early stages of gaze following should first learn in the horizontal plane, and then the

vertical plane. Our results capture this prediction: they suggest that 12-month-old infants have only learned horizontal gaze following and are not yet able to gaze follow vertically. We find no such difference in our control condition, where objects appear in the same locations but are not cued with gaze. We find that latency is faster in the control than the gaze condition, which could be due to it being easier to disengage from the fixation cross than the face. We find no difference in latency across planes, which suggests that infants are not slower to make gaze shifts in either plane. Importantly, the results of our control condition indicate that the accuracy difference found in the experimental condition is not because infants' eye movements are *a priori* more accurate in the horizontal than the vertical plane. Our results are compatible with learning-based theories, specifically Nagai and colleagues' robotic implementation, in which a combination of associative learning mechanisms and social reinforcement were sufficient to support the emergence of gaze following. In contrast, representational accounts predict no difference between horizontal and vertical gaze following. Importantly, however, to strengthen the learning-based account of gaze following, we should be able to demonstrate developmental changes in horizontal and vertical gaze following. Thus, in Experiment 2, we tested 6-month-old infants in the same paradigm.

Although in Experiment 1, when using Bayesian sequential testing we reached our maximum number of infants, this was due to not meeting our threshold for *all* analyses. This is due to the fact it takes a larger amount of evidence to reach any given Bayes Factor in support of the null hypothesis than against it (Johnson & Rossell, 2010). However, frequentist results were the same at 20 participants as at 40 participants, i.e. all results significant at the $p < 0.05$ threshold at 40 participants also were significant at 20 participants. This is likely due to the high number of trials (64) compared to many

infancy studies. Hence, in each further experiment reported, we collected a final sample of 20 infants.

Experiment 2 (6-month-olds)

Method

All stimuli and procedures were identical to Experiment 1. Infants viewed up to 64 trials (range: 28-64, $M = 57$; $SD = 10$). As we did not observe latency differences in our original sample, we did not analyse latency for the other age groups.

Participants

20 typically developing 6-month-old infants took part (mean age: 182 days; range: 166 days to 196 days; 7 female; all Caucasian; 18 monolingual English).

Results

We submitted proportion looking to a 2 x 2 repeated measures ANOVA with main effects of condition (gaze vs. control) and plane (horizontal vs. vertical) (Figure 3). Accuracy was significantly higher for control than gaze trials [main effect of condition: $F(1,19) = 175, p < 0.001$; $BF_{10} = 5.09e^{21}$], and we found moderate evidence for no difference between horizontal and vertical accuracy [main effect of plane: $F(1,19) = 0.08, p = .78$; $BF_{01} = 4.28$]. The ANOVA revealed moderate evidence for no interaction between condition and plane for proportion looking [$F(1,19) = 1.34, p = .26$; $BF_{01} = 3.81$]. Finally, planned, two-tailed paired t-tests found no evidence for accuracy at levels greater than expected by chance for either plane [horizontal: $t(19) = 0.36, p = .73$; $BF_{01} = 3.12$; vertical: $t(19) = -1.13, p = .27$; $BF_{01} = 1.96$]. Dichotomous

target looking results mirrored the proportion looking results and can be found on OSF (Figure S7).

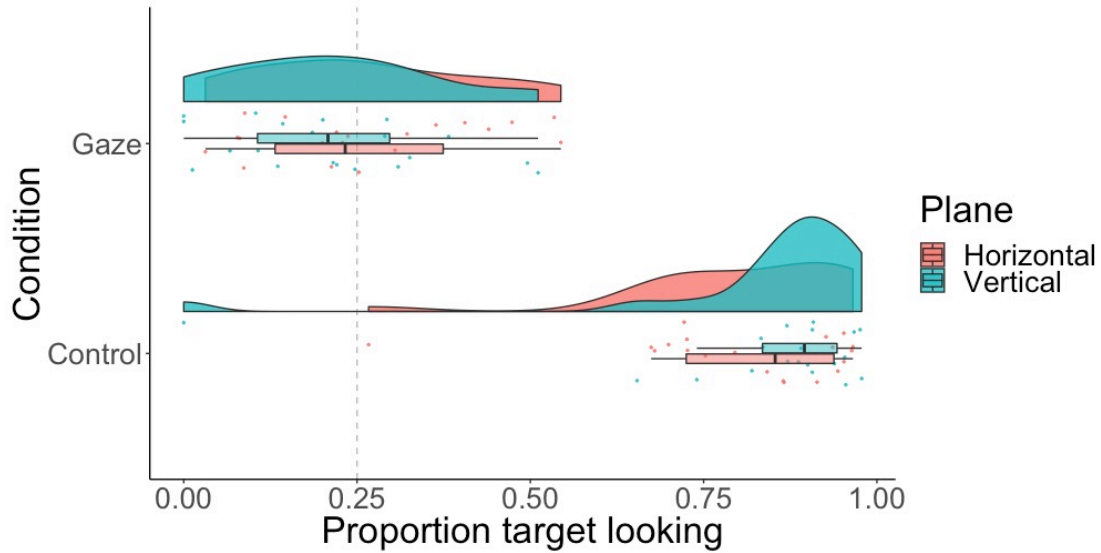


Figure 3. Raincloud plot showing proportion target looking across conditions (gaze vs. control) and planes (horizontal vs. vertical) for 6-month-olds. Dashed line indicates chance performance.

Comparison with Experiment 1

In order to assess whether 6-month-olds were indeed performing differently to 12-month-olds, we submitted proportion looking in gaze trials to a 2 x 2 repeated measures ANOVA with main effects of age (12 months vs. 6 months) and plane (horizontal vs. vertical). Although according to our frequentist analysis the ANOVA did not reveal a significant interaction between age and plane for proportion looking, Bayes factor analysis revealed strong evidence for an interaction, [$F(1,58) = 2.73, p = .10; BF_{10} = 14.56$].

Discussion

In Experiment 2 we tested whether 6-month-old infants could follow gaze in either the horizontal or vertical plane. In contrast to 12-month-olds, this younger group showed no difference in their horizontal and vertical gaze following and did not show above-chance accuracy in either plane. As in Experiment 1, these results capture the predictions of the robotic model by Nagai et al (2006), suggesting that younger infants have not yet learned to follow gaze in either plane. Thus, the difference in gaze following ability across the horizontal and vertical planes that we find in 12-month-olds is not an artefact of our stimuli (as the stimuli used were identical), strengthening our interpretation that infants may learn to follow gaze in stages. In Experiment 3, we investigated whether, as we predicted according to the robot model, older infants can follow gaze equally successfully in the horizontal and vertical plane.

Experiment 3 (18-month-olds)

Method

All stimuli, procedures, and analyses were identical to Experiments 1 & 2. Infants viewed up to 64 trials (range: 42-64, $M = 60$; $SD = 7$).

Participants

20 typically developing 18-month-old infants took part (mean age: 546 days; range: 530 days to 567 days; 10 females; all Caucasian; all monolingual English). Two additional infants were excluded due to fussiness.

Results

We submitted proportion looking to a 2 x 2 repeated measures ANOVA with main effects of condition (gaze vs. control) and plane (horizontal vs. vertical) (Figure 4). Accuracy was greater for control than gaze trials [main effect of condition: $F(1,19) = 356, p < 0.001$; $BF_{10} = 7.38e^{26}$], and greater for the horizontal than the vertical plane [main effect of plane: $F(1,19) = 9.37, p < .005$; $BF_{10} = 2.40$]. The ANOVA revealed a significant interaction between condition and plane for proportion looking [$F(1,19) = 9.92, p < .01$; $BF_{10} = 2.83$]. Planned, two-tailed, paired t-tests revealed that in the gaze condition, accuracy was higher for the horizontal [$M = 0.34, SD = 0.13$] than the vertical plane [$M = 0.17, SD = 0.13$; $t(19) = 3.52, p < .005$; $BF_{10} = 72.68$], with moderate evidence for no difference in the control condition [Horizontal: $M = 0.87, SD = 0.14$; Vertical: $M = 0.86, SD = 0.14$; $t(19) = .53, p = .60$; $BF_{01} = 3.13$]. Finally, paired t-tests against chance (0.5) revealed that in the gaze condition, accuracy was significantly above chance for the horizontal plane [$t(19) = 2.93, p < .01$; $BF_{10} = 7.66$], and significantly below chance for the vertical plane [$t(19) = -2.61, p < 0.05$; $BF_{10} = 4.07$]. Dichotomous target looking results mirrored the proportion looking results and can be found on OSF (Figure S8).

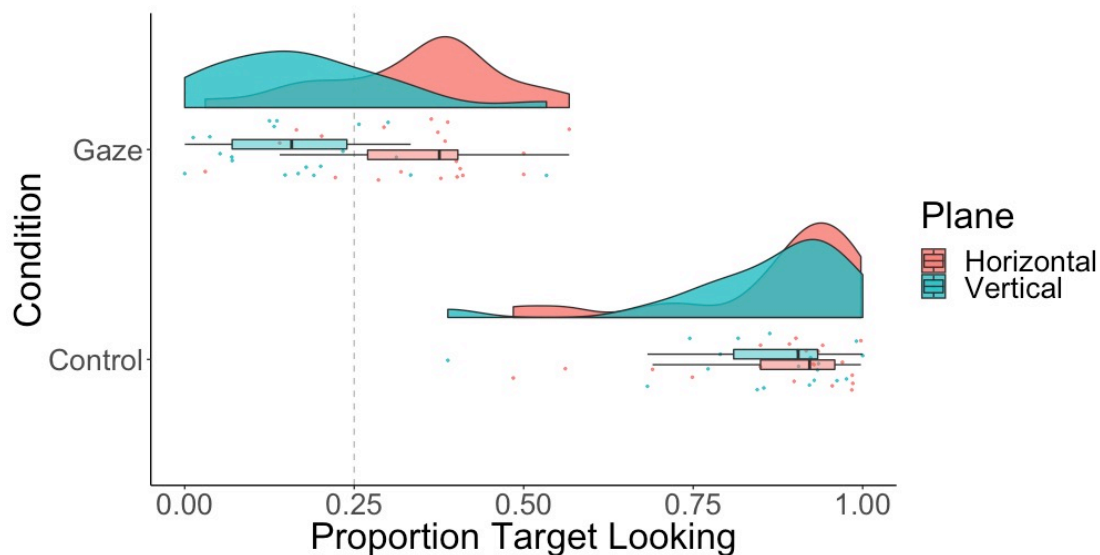


Figure 4. Raincloud plot showing proportion target looking across conditions (gaze vs. control) and planes (horizontal vs. vertical) for 18-month-olds. Dashed line indicates chance performance.

Discussion

In this experiment, we tested whether 18-month-infants were yet able to follow gaze in both the horizontal and vertical plane. We found that they showed the same pattern as the 12-month-olds and were only able to follow gaze in the horizontal plane. Thus, in Experiment 3 we did not identify the age at which infants have acquired both horizontal and vertical gaze following abilities. Critically, however, we replicated the 12-month-old results with 18-month-olds; the results of Experiment 3 are therefore consistent with an account of gaze following in which infants learn in stages.

General discussion

In the current study we tested the predictions of Nagai et al.'s (2006) robotic implementation of the development of gaze following. In this work, the robot learned

to gaze follow based on associative learning and visuomotor input and did so in stages: first in the horizontal plane, followed by the vertical plane. Our results capture this prediction: while 6-month-old infants are unable to gaze follow in either plane, by 12 months they have learned to follow gaze in the horizontal plane, but critically, not the vertical plane. This pattern is replicated in 18-month-olds. We find no such difference in our control conditions, where objects are not cued with gaze; thus, the difference we observe in the experimental conditions is specific to gaze following and is not driven by motoric differences in the ease of horizontal and vertical eye movements. Our results therefore support the robotic prediction and learning-based theories of gaze following, suggesting that a combination of associative learning mechanisms and social reinforcement may be sufficient for the emergence of gaze following. Importantly, Nagai and colleagues' model offers an explicit mechanism by which the horizontal-first gaze following behaviour we observed may emerge. In the robotic model, visual input in the horizontal plane was more variable than visual input in the vertical plane. The model therefore learned to detect differences in the horizontal plane first. Thus, taken together with our replication of the model's behavioural prediction, this work makes the further, empirically testable prediction that input to infants, in particular gaze direction information, should be richer in the horizontal than the vertical plane.

While the current data offer support for learning-based accounts of gaze following, we cannot rule out alternative accounts based on innate cognitive processes. In particular, the shape of the human eye is such that horizontal eye movements are easier to see, due to the amount of visible white sclera being larger. For this reason, one could argue for an innate perceptual system tuned to spot horizontal eye movements (and for evidence of gaze cueing in newborns, see Farroni et al., 2004). However, if this were the case, in the current study we might expect to find better gaze following

for upwards vertical gaze (where there is more visible sclera) than the downwards vertical gaze (where no sclera is visible), which is not the case. In fact, we find worse accuracy for upwards looks than all three other positions (Figures S1-S4). Intuitively, this is consistent with learning theories, as infants are typically situated below adults' line of gaze, which would elicit more frequent downwards than upwards looks from the adult. Furthermore, if information in the horizontal direction is generally richer, we might expect differences also when objects appear in horizontal and vertical locations without being cued by gaze (as in our control condition), which we do not. Again, empirical evidence from naturalistic input to infants can assess this possibility. Equally, it is possible that while elements of the learning environment (such as increased variability in the horizontal plane) may shape infants' behavioural responses, they may still possess an ability to read others' mental states. We also cannot rule out the possibility that some other innate mechanism biases gaze following to horizontal planes. However, we are not aware of any such theory making this prediction.

It is interesting to note that infants in our study are “bad” at gaze following, i.e. even the 18-month-olds were only above chance for horizontal gaze following. This is to be expected, as it has been shown repeatedly that infants respond best to head and eye movements in combination (Moore et al., 1997). The movement of just the eyes is a very subtle cue and so is understandably more difficult for infants to follow. We chose to manipulate only eyes in this experiment in order to create stimuli that were as controlled as possible, but future research should explore whether our findings can be replicated for combined head movement and eye cues, since, if gaze following is learned, these additional cues should also be integrated by the learning mechanism.

Conclusion

Taken together, our results and those of Nagai et al. provide evidence that a low-level learning mechanism provided with structured input may be sufficient to support the development of gaze following in human infants. Whether infants do additionally possess an innate ability to read others' intentions remains an open question. Importantly, however, current prominent theories in developmental psychology (e.g. Baron-Cohen, 1995; Csibra & Gergely, 2009) do not predict that there would be differences in accuracy of gaze following across the horizontal and vertical planes, as an understanding of shared attention or communicative intentions (the basis for these theories) is not dependent on direction of gaze. As such, the current results point to new opportunities for the development of mechanistic theories that can account for these behavioural data.

General discussion

1. Summary of findings

This thesis set out to investigate the claims made by Natural Pedagogy (NP; Csibra & Gergely, 2009; 2011) about how infants and young children interpret and use ostensive-referential cues. NP has three main claims: (i) infants are sensitive to ostensive cues; (ii) ostensive cues create a referential expectation; (iii) infants are biased to encode kind-generalisable knowledge in an ostensive-referential context, and to interpret this information as culturally relevant.

In Study 1, we attempted two replications of a seminal study by Yoon et al. (2008). This study is one of the key pieces of evidence for the most specific claim of NP: that an ostensive-referential context biases infants to encode kind-generalisable information. Importantly, the findings of this study showed that the information infants retained was qualitatively different in ostensive and non-ostensive contexts. This was theoretically important, as it could not be explained by some low-level mechanism whereby an ostensive context is merely more attractive to infants, and therefore causes them to encode or retain information from this context more easily. We were unable to replicate the study in two versions: one before and one after input from the senior author of the original study. This outcome has great theoretical importance. With little other evidence supporting the claim that an ostensive-referential context biases infants to encode kind-generalisable information, we are left with a body of literature with a lot of support for infant sensitivity to ostensive cues in terms of enhancing their attention, but little beyond that suggesting that these signals are any different from any other attention-enhancing cues.

In Study 2, we went beyond the explicit claims of NP, and tried to make a new prediction based on an assumption stemming from the same claim investigated in Study 1. We reasoned that if an ostensive context should bias infants to encode kind-generalisable, culturally relevant information, then this may also transfer to the types of information that they should be biased to transmit to others. Although infants have been investigated heavily as the receivers of information, any cues relevant to learning might also be relevant to their teaching. For this study, we attempted to extend a study by Vredenburg et al. (2015), who found evidence for toddlers' preferential transmission of actions learnt in an ostensive context. We extended this by manipulating the complexity of the actions learnt by the toddler, reasoning that if they truly perceived the action learnt in the ostensive context to be culturally relevant, they should preferentially transmit this action, even if it is more complex, over a simpler action learnt in a non-ostensive context. Instead, we found that infants preferentially transmitted the simple action. To follow up on the mechanism for this, we conducted a replication of the original study, and were unable to replicate the original finding of preferential transmission of actions learnt in an ostensive context. So again, we were unable to find any effect of ostensive cues in this study. Although this study was not as clear a test of NP as Study 1, we believe it highlights the importance of making theories more specified rather than keeping them relatively vague. If theories are hypothesised to only explain the behaviour of certain ages, or only extend to certain contexts, then this should be tested, and findings embedded within an updated version of the theory. Keeping theories vague means that a wide variety of outcomes will be claimed to be consistent with the theory, which stops the cycle of the updating of theory based on new evidence.

In Study 3, we looked beyond NP, to test an alternative hypothesis for how infants may learn to gaze-follow. One of the claims of NP is that ostensive cues create a “referential expectation”, due to the communicative intention conveyed by ostensive cues, i.e. that observing ostensive cues tells the infant they should look out for some oncoming information that is going to be conveyed to them through reference. Instead, we looked to the developmental robotics literature to theorise about other ways that infants may learn to gaze follow. A study by Nagai et al. (2006) showed that a developmental robot could be taught to follow gaze purely through low-level reinforcement learning, without any high-level understanding of communicative intention. Interestingly, the robot learnt in stages, first learning horizontal, and then vertical gaze following (due to the higher perceptual variability in the horizontal plane). We tested this prediction in infants, and found that they too learnt in stages, learning horizontal gaze-following first. As the infant results mirrored the robot results, this suggested that infants might learn to gaze-follow in the same way as the robot: through low-level reinforcement learning.

Overall, in the studies conducted in this thesis we found no support for NP, and some preliminary support for a low-level theory of how infants respond to ostensive-referential cues. We did not set out to find evidence against NP; we merely were unable to find any evidence that supported NP. As we know that at least adults do understand the communicative and referential intentions of others, this thesis cannot speak to the developmental trajectory of this ability, but our findings suggest that this is a process that is learnt rather than innate. The main way in which this thesis contributes to the literature is to question the assumption that findings from published studies can be assumed to be true, and in its emphasis on the importance of using open and reproducible methods in developmental psychology. Beyond the

specific evidence found in the three studies, this thesis represents an attempt to show how to implement open science practices across a doctoral thesis in developmental psychology, and the benefits of doing so (see Kathawalla, Silverstein, & Syed, 2020 for further discussion of these benefits).

2. Replication

A key component to this thesis has been replication. Replication features in some form in all three studies. Study 1 can be considered a pure replication, in that it doesn't aim to extend upon the study it is attempting to replicate (save the additional eye-tracking measure, which has no effect on the paradigm itself). Study 2 uses a replication to clarify findings from an extension. Study 3 includes replication within itself – the results with 18-month-olds replicate those found with the 12-month-olds in the same paradigm. Replication is essential for the progress of science. In psychology, for too long we have taken the published literature at face value, perhaps questioning the interpretation of results or whether or not authors may be hiding some additional results, but nonetheless taking at face value that if $p < 0.05$, then this is a statistically significant result that is meaningful. This is an error. Even if we assume that all research that has ever been done has been published and this research has been conducted without any QRPs, then we still would expect to find 5% false positive results being published. However, we are aware that both of those conditions are completely false (see the first chapter of this thesis), and therefore that the false positive rate is far higher.

This is not all to say that the studies that we were unable to replicate in this thesis can be known to be wrong, or that ours can be known to be right. Indeed, there is also the possibility that both are right, and that “hidden moderators” explain the

difference in findings (Brown et al., 2014; Gilbert et al., 2016). However, it has been noted that this explanation is more likely when original authors did not approve the design of the replication, which we do in Study 1, and inadvertently in Study 2, as we had one of the original authors as a signed reviewer for the manuscript. Although this is still a possibility, one of the key findings coming from large-scale replication projects is that although they see that effect sizes are influenced by small differences between labs, paradigms, and participants, it has been found that if an effect replicates, it does so across these contexts (Inbar, 2016; Klein et al., 2018; Open Science Collaboration, 2015).

However, we must keep in mind that these studies have very large sample sizes in order to investigate these differences, and so it is still possible that a non-replication is the product of a reduced effect size due to small differences between studies, meaning that with the sample sized used we were underpowered to detect the effect. This is a critical issue with all studies in this thesis, as well as infancy research in general (Davis-Kean & Ellis, 2019).

Many suggestions have been made for how sample sizes for replication studies should be calculated, including conducting a power analysis for 50% of the original effect size, to allow for inflated effect size estimates in original studies (Camerer et al., 2018). These methods were not used in the studies in this thesis, with all replication experiments matching the sample sizes of original studies. In future, it would be preferable to use one of these methods in order to maximise any confidence that not finding an effect is because none exists. However, due to the extremely high attrition in Study 1 and the general difficulty recruiting very large samples of infants, this was not done. One possible solution for this issue is discussed in detail in the next section.

It must be noted that the way we conceptualised whether or not we were able to replicate effects is not the only way in which this can be done. Our way of formalising whether or not we were able to replicate was whether the original analyses conducted on our new data produced the same pattern of results (using Frequentist statistics), and for when a null result was obtained, how much support this showed for the null hypothesis (using the threshold of a BF01 of 3 to be able to have moderate support for the null hypothesis; Lee & Wagenmakers, 2013). Other options include conducting a meta-analysis including both the replication and all previous published and unpublished versions of the experiment, conducting prediction intervals (predicting the range of effects that researchers should expect in a replication study, Patil et al., 2016), and the ‘small telescopes’ approach (assessing the extent to which the replication result is consistent with an effect size big enough to have been detectable in the original study, Simonsohn, 2015). A more comprehensive approach may be to combine multiple of these approaches (e.g. Camerer et al., 2018).

2.1 Large-Scale Collaboration

One solution to the problem of infancy studies having low power comes from the exciting ManyBabies initiative (Frank et al., 2017). Taking inspiration from the ManyLabs studies, ManyBabies aims to conduct large-scale infancy research across labs around the world, with the intention of both a high-powered replication study of a proposed effect, as well as the power to investigate individual differences between participants and labs and how these may influence the effect (or not). Importantly, rather than doing an exact replication of a particular paradigm, researchers work together to come up with a “best-practice” version of the study which takes into account any proposed methodological flaws in previous versions from both sides of

the argument. This approach ensures that what is truly being assessed is the theorised underlying effect itself, rather than any one particular paradigm. Going forward, I'm excited to take inspiration from this approach when designing future studies.

Interestingly, and highly relevant to this thesis, the first ManyBabies project investigated the Infant Directed Speech (IDS) preference in infants across ages and labs (The ManyBabies Consortium, 2019). They found a robust preference for IDS, but much smaller than that found in previous studies, which they attributed to differences in procedure and stimulus. The effect was moderated by experimental method, age, and the child's native language (whether it matched the North American (NA) English stimuli). Importantly, beyond these moderators, there was little evidence for heterogeneity. Rather than being the result of subtle procedural variations, differences between labs appeared to be due to sampling error. This initiative showed that replication at this scale leads to more precise measurements and insight into which variables matter, making hidden mediators explicit and offering the chance to integrate them into theory. The ManyBabies model is now being used for several follow-up projects, both extending the IDS preference work (one of which I have collected data for, investigating the test-retest reliability of this preference: Schreiner et al., in prep.), and some on other topics such as implicit theory of mind.

Going forward, it seems like it will be a good model for infancy research to use these large-scale collaborations to see which effects in infancy research can be relied on, and the intricacies of what moderates these effects. But for the single researcher or single lab model, there are still important lessons to be learnt:

1. Design a best-practice paradigm, rather than relying on previous research.

2. Ensure your study is adequately powered not only to detect the effect size predicted from the literature, but for a substantially smaller effect size likely to be closer to the “true” effect size.
3. Use open, transparent, reproducible methods, to ensure your study could be replicated by others, and that your data can be integrated into meta-analyses.

3. Future directions

3.1 Study extensions

3.1.1 Study 1

It is important for future research to discern why we were unable to replicate the original study. The first step for this would be for the original lab to conduct a direct replication with the original stimuli. If this did not replicate, we would be fairly certain that the original result was a Type 1 error. If the original result were to replicate, we could then attempt our own replication with the original stimuli. If this did not replicate, we would know that it is likely that the original result is due to some lab effect rather than the intended manipulation. If it did replicate, it is likely that our original non-replication was due to stimuli differences. It would be important to then follow up on what the differences between the stimuli were that made a difference to whether or not the effect is observed. This may lead to the understanding that only a very stringent manipulation of ostensive context leads to observing the effect (in which case NP would need to be modified to include these constraints), or perhaps that the original effect is due to spurious factors that were not intended to be manipulated, and is not a reflection of the theoretical prediction. It is also important that all existing replications of this study are made public, as I am aware through

word of mouth of some data being in existence but not being published. Preprints and data sharing can be a useful avenue for studies like this where perhaps there isn't enough data for a full publication, but the dataset can be added to potential meta-analyses.

3.1.2 Study 2

Future research would need to take similar steps to those following Study 1 in order to discern why we were unable to replicate the original finding. If, through this process, we find that the original effect does exist, but we did not manipulate ostensive context effectively, then further research can be done to answer our original question of how the effect of ostensive context interacts with other factors such as complexity. Additionally, if it is differences between our version and the original study that were responsible for the difference in results, then this can be investigated, and theory can be adjusted to include these specifications. For example, from a linguistic perspective it may be very interesting to look into how different types of verbal prompts are interpreted by infants and lead to different outcomes. If we find that the original effect cannot be replicated, we may assume that ostensive context does not have an effect on information transmission in two-year-olds, and further research can be done into how other factors that do affect information transmission interact.

Additionally, it is unclear whether 'teaching' is really what is being measured in the paradigm used in this study and by Vredenburg and colleagues. Future studies on this topic could aim to create better paradigms that would more readily be able to distinguish teaching from merely playing with the toy in the presence of another interlocutor. Some ways of doing this could be to measure whether toddlers generalise

the information they have learned to a new object or seeing if they independently correct someone who can't work out how to play with the toy without being asked to do so.

3.1.3 Study 3

Although from Study 3 we know that infants do learn horizontal gaze following first, we were not able to discern the age at which they can follow both horizontal and vertical gaze. Testing the same paradigm in older infants and adults will help complete this story. There are many follow up studies using the original paradigm that could be designed to disentangle this effect further. Moving stimuli could be created that include both a gaze shift and a head turn. These stimuli would be more naturalistic and more similar to the visual input given to the robot. We could also investigate other referential cues, for example pointing, in order to see whether the same stage learning is found.

Using a directional control condition that is a better match to the gaze shift could shed light on whether this is not something specific to gaze but is to do with any directional cue. Simply keeping the control condition the same but with the face remaining in the middle (with direct gaze) could help control for the possibility that it may be easier to disengage from a face with direct gaze in the horizontal than in the vertical plane.

Additionally, if gaze following is learned through reinforcement, it may be possible to train infants to show a preference for vertical gaze following for a short time. A training study where caregivers are asked to repeatedly show interesting stimuli in the vertical plane could provide converging evidence for this theory.

Going forward, it would be important to acquire naturalistic infant head-camera data at different ages in order to answer important questions about the input the infant is getting. Is it true that there is more perceptual variability in the horizontal plane in the input for gaze? Do infants have less experience turning their head up or down to look away from caregivers' faces? Are interesting stimuli less likely at the top or bottom of caregivers' faces? These are all promising avenues of research and would get us closer to understanding how gaze following develops in infants, as we have only scratched the surface.

3.2 Theory development

It has been said that following the replication crisis comes the theory crisis (Oberauer & Lewandowsky, 2019). Developmental psychology has an abundance of theories attempting to explain behaviour and learning. Some theories attempt to explain a small subset of behaviours, some attempt to explain a wide variety. There are many issues with the way we theorise in developmental psychology, and in psychology in general. Firstly, as previously discussed, theory isn't regularly updated based on incoming evidence. This is worsened by the fact that we have access to a biased set of evidence, as discussed at length. Another issue is that theories are not turned into explicit models as often as they should be, making it difficult to evaluate evidence for or against a theory (Guest & Martin, 2020). Theories are rarely integrated together to include both low-level and high-level mechanisms, more often pitting extremes against each other in the age-old "nature or nurture" debate that remains across psychology. Lastly, current theories are often imprecise in determining which level they explain behaviour at. Instead, it needs to be made explicit that

theories should make predictions at a specific level: neural, cognitive, individual behavioural, or group behavioural.

3.2.3 Conclusion

In this thesis, we were able to question previous findings in the literature supporting NP through two non-replications. We were also able to show support for a learning theory for how infants may begin to follow ostensive cues (in this case, gaze following). As well as this, through the use of open science practices this thesis can contribute further than the studies presented here. Open data, code, and materials mean that the reproducibility and replicability of our results can be verified, and that our work can be easily extended and adapted. While we may have raised more questions than we have answered (but all good research should do this), we have also illuminated promising avenues for future research. While much of the published literature cannot be taken at face value (due to few existing replications of work in this field, and up until recently very little in the way of open and transparent science), the work presented in this thesis (by being open, transparent, and reproducible) can meaningfully contribute to a cumulative science of developmental psychology. For now, when evaluating the existing research, we can try to do what we can with what exists: encourage the publication of file-drawered research on these topics, conduct meta-analyses on existing data, encourage the sharing of data with all published work, and encourage new hypothesis-testing research to be preregistered. In this way, we will feel like we know less, as the field is exposed as messy and confusing, but it is within this mess and confusion that we will begin to get closer to some semblance of truth and insight on this (or any) topic.

References:

- Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*(7021), 68–72.
<https://doi.org/10.1038/nature03086>
- Anderson, B. A. (2013). A value-driven mechanism of attentional selection. *Journal of Vision*, *13*(3), 7–7. <https://doi.org/10.1167/13.3.7>
- Anderson, D. R., & Pempek, T. A. (2005). Television and Very Young Children. *American Behavioral Scientist*, *48*(5), 505–522.
<https://doi.org/10.1177/0002764204271506>
- Baldwin, I. T., & Schultz, J. C. (1983). Rapid Changes in Tree Leaf Chemistry Induced by Damage: Evidence for Communication Between Plants. *Science*, *221*(4607), 277–279.
<https://doi.org/10.1126/science.221.4607.277>
- Bannard, C., Rosner, M., & Matthews, D. (2017). What's Worth Talking About? Information Theory Reveals How Children Balance Informativeness and Ease of Production. *Psychological Science*, *28*(7), 954–966.
<https://doi.org/10.1177/0956797617699848>
- Baron-Cohen, S. (1995a). *Mindblindness An Essay on Autism and Theory of Mind*.
- Baron-Cohen, S. (1995b). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In *Joint attention: Its origins and role in development* (pp. 41–59). Lawrence Erlbaum Associates.

- Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., & Williams, S. C. R. (1999). Social intelligence in the normal and autistic brain: An fMRI study. *European Journal of Neuroscience*, 9.
- Barry, R. A., Graf Estes, K., & Rivera, S. M. (2015). Domain general learning: Infants use social and non-social cues when learning object statistics. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00551>
- Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development*, 23(2), 223–229. [https://doi.org/10.1016/S0163-6383\(01\)00037-6](https://doi.org/10.1016/S0163-6383(01)00037-6)
- Bear, A., & Bloom, P. (2016). A Simple Task Uncovers a Postdictive Illusion of Choice. *Psychological Science*, 27(6), 914–922. <https://doi.org/10.1177/0956797616641943>
- Beier, J. S., & Spelke, E. S. (2012). Infants' Developing Understanding of Social Gaze: Infants' Developing Understanding of Social Gaze. *Child Development*, no-no. <https://doi.org/10.1111/j.1467-8624.2011.01702.x>
- Benasich, A. A., & Bejar, I. I. (1992). The Fagan test of infant intelligence: A critical review. *Journal of Applied Developmental Psychology*, 13(2), 153–171. [https://doi.org/10.1016/0193-3973\(92\)90027-F](https://doi.org/10.1016/0193-3973(92)90027-F)
- Bertenthal, B. I., Boyer, T. W., & Harding, S. (2014). When do infants begin to follow a point? *Developmental Psychology*, 50(8), 2036–2048. <https://doi.org/10.1037/a0037152>
- Blake, J., O'Rourke, P., & Borzellino, G. (1994). Form and function in the development of pointing and reaching gestures. *Infant Behavior and*

Development, 17(2), 195–203. [https://doi.org/10.1016/0163-6383\(94\)90055-8](https://doi.org/10.1016/0163-6383(94)90055-8)

Blaser, E., & Kaldy, Z. (2010). Infants Get Five Stars on Iconic Memory Tests: A Partial-Report Test of 6-Month-Old Infants' Iconic Memory Capacity. *Psychological Science*, 21(11), 1643–1645. <https://doi.org/10.1177/0956797610385358>

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, 64(4), 215–234. <https://doi.org/10.1016/j.cogpsych.2011.12.002>

Brembs, B. (2018). Prestigious Science Journals Struggle to Reach Even Average Reliability. *Frontiers in Human Neuroscience*, 12. <https://doi.org/10.3389/fnhum.2018.00037>

Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2019). Young children consider the expected utility of others' learning to decide what to teach. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-019-0748-6>

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, 8(6), 535–543. <https://doi.org/10.1111/j.1467-7687.2005.00445.x>

Brown, S. D., Furrow, D., Hill, D. F., Gable, J. C., Porter, L. P., & Jacobs, W. J. (2014). A Duty to Describe: Better the Devil You Know Than the Devil You Don't. *Perspectives on Psychological Science*, 9(6), 626–640. <https://doi.org/10.1177/1745691614551749>

Brugger, A., Lariviere, L. A., Mumme, D. L., & Bushnell, E. W. (2007). Doing the Right Thing: Infants' Selection of Actions to Imitate From Observed Event

- Sequences. *Child Development*, 78(3), 806–824.
<https://doi.org/10.1111/j.1467-8624.2007.01034.x>
- Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, 120(3), 331–340.
<https://doi.org/10.1016/j.cognition.2010.12.001>
- Bugnyar, T., Stöwe, M., & Heinrich, B. (2004). Ravens, *Corvus corax*, follow gaze direction of humans around obstacles. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1546), 1331–1336.
<https://doi.org/10.1098/rspb.2004.2738>
- Burdett, E. R. R., Dean, L. G., & Ronfard, S. (2018). A Diverse and Flexible Teaching Toolkit Facilitates the Human Capacity for Cumulative Culture. *Review of Philosophy and Psychology*, 9(4), 807–818.
<https://doi.org/10.1007/s13164-017-0345-4>
- Butler, L. P., & Markman, E. M. (2012). Preschoolers Use Intentional and Pedagogical Cues to Guide Inductive Inferences and Exploration: Pedagogical Cues and Exploration. *Child Development*, 83(4), 1416–1428.
<https://doi.org/10.1111/j.1467-8624.2012.01775.x>
- Butler, L. P., & Markman, E. M. (2014). Preschoolers use pedagogical cues to guide radical reorganization of category knowledge. *Cognition*, 130(1), 116–127. <https://doi.org/10.1016/j.cognition.2013.10.002>
- Butler, L. P., & Markman, E. M. (2016). Navigating pedagogy: Children's developing capacities for learning from pedagogical interactions. *Cognitive Development*, 38, 27–35.
<https://doi.org/10.1016/j.cogdev.2016.01.001>

- Butler, L. P., & Tomasello, M. (2016). Two- and 3-year-olds integrate linguistic and pedagogical cues in guiding inductive generalization and exploration. *Journal of Experimental Child Psychology, 145*, 64–78.
<https://doi.org/10.1016/j.jecp.2015.12.001>
- Butterworth, G. (1991). The ontogeny and phylogeny of joint visual attention. In *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 223–232). Basil Blackwell.
- Butterworth, G. (2003). Pointing is the royal road to language for babies. In *Pointing: Where language, culture, and cognition meet* (pp. 9–33). Lawrence Erlbaum Associates Publishers.
- Butterworth, G., & Grover, L. (1990). Joint visual attention, manual pointing, and preverbal communication in human infancy. In *Attention and performance 13: Motor representation and control* (pp. 605–624). Lawrence Erlbaum Associates, Inc.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. <https://doi.org/10.1038/nrn3475>
- Caldwell, C. A., Renner, E., & Atkinson, M. (2018). Human Teaching and Cumulative Cultural Evolution. *Review of Philosophy and Psychology, 9*(4), 751–770. <https://doi.org/10.1007/s13164-017-0346-3>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., ... Wu, H. (2018). Evaluating the replicability of social science

- experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644. <https://doi.org/10.1038/s41562-018-0399-z>
- Carey, S., & Xu, F. (2001). *Infants' knowledge of objects: Beyond object @les and object tracking*. 35.
- Caron, A. J., Kiel, E. J., Dayton, M., & Butler, S. C. (2002). Comprehension of the Referential Intent of Looking and Pointing Between 12 and 15 Months. *Journal of Cognition and Development*, 3(4), 445–464. <https://doi.org/10.1080/15248372.2002.9669677>
- Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, 63(4), i. <https://doi.org/10.2307/1166214>
- Casler, K., Terziyan, T., & Greene, K. (2009). Toddlers view artifact function normatively. *Cognitive Development*, 24(3), 240–247. <https://doi.org/10.1016/j.cogdev.2009.03.005>
- Churcher, J., & Scaife, M. (1982). How infants see the point. In *Social Cognition* (pp. 110–136). Harvester Press.
- Cooper, R. P., & Aslin, R. N. (1990). *Preference for Infant-Directed Speech in the First Month after Birth*. 61(5), 1584–1595.
- Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology*, 34(1), 28–38. <https://doi.org/10.1037/0012-1649.34.1.28>
- Corriveau, K. H., DiYanni, C. J., Clegg, J. M., Min, G., Chin, J., & Nasrini, J. (2017). Cultural differences in the imitation and transmission of inefficient

- actions. *Journal of Experimental Child Psychology*, 161, 1–18.
<https://doi.org/10.1016/j.jecp.2017.03.002>
- Cristia, A., Seidl, A., Singh, L., & Houston, D. (2016). Test-Retest Reliability in Infant Speech Perception Tasks. *Infancy*, 21(5), 648–667.
<https://doi.org/10.1111/infa.12127>
- Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., Orben, A., Parsons, S., & Schulte-Mecklenbeck, M. (2019). Seven Easy Steps to Open Science: An Annotated Reading List. *Zeitschrift Für Psychologie*, 227(4), 237–248. <https://doi.org/10.1027/2151-2604/a000387>
- Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567), 1149–1157. <https://doi.org/10.1098/rstb.2010.0319>
- Csibra, Gergely. (2007). Teachers in the wild. *Trends in Cognitive Sciences*, 11(3), 95–96. <https://doi.org/10.1016/j.tics.2006.12.001>
- Csibra, Gergely. (2010). Recognizing Communicative Intentions in Infancy. *Mind & Language*, 25(2), 141–168. <https://doi.org/10.1111/j.1468-0017.2009.01384.x>
- Csibra, Gergely, & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of Change in Brain and Cognitive Development. Attention and Performance*, XXI (pp. 249–274). Oxford University Press.
- Csibra, Gergely, & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13(4), 148–153. <https://doi.org/10.1016/j.tics.2009.01.005>
- Csibra, Gergely, Hernik, M., Shamsudheen, R., Tatone, D., & Senju, A. (2015). *For the Record: A Commentary by Csibra, Senju, et al. on Gaze Following*.

<http://cognitionandculture.net/blogs/icci-blog/for-the-record-a-commentary-by-csibra-senju-et-al-on-gaze-following/>

Csibra, Gergely., & Volein, Ágnes. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, 26(1), 1–11. <https://doi.org/10.1348/026151007X185987>

Daum, M. M., Ulber, J., & Gredebäck, G. (2013). The development of pointing perception in infancy: Effects of communicative signals on covert shifts of attention. *Developmental Psychology*, 49(10), 1898–1908. <https://doi.org/10.1037/a0031111>

Davis-Kean, P. E., & Ellis, A. (2019). An overview of issues in infant and developmental research for the creation of robust and replicable science. *Infant Behavior and Development*, 57, 101339. <https://doi.org/10.1016/j.infbeh.2019.101339>

Deak, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of Gesture and Target on 12- and 18-Month-Olds' Joint Visual Attention to Objects in Front of or Behind Them. *Developmental Psychology*, 36(4), 511–523.

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science*, 17(2), 270–281. <https://doi.org/10.1111/desc.12122>

Del Bianco, T., Falck-Ytter, T., Thorup, E., & Gredebäck, G. (2019). The Developmental Origins of Gaze-Following in Human Infants. *Infancy*, 24(3), 433–454. <https://doi.org/10.1111/infa.12276>

- D'Entremont, B., Hains, S. M., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development*, *20*(4), 569–572.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. <https://doi.org/10.3389/fpsyg.2014.00781>
- Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89.
<https://doi.org/10.1016/j.jmp.2015.10.003>
- Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, *25*(1), 207–218.
<https://doi.org/10.3758/s13423-017-1266-z>
- Dink, J. W., & Ferguson, B. (2015). *EyetrackingR: An R Library for Eye-tracking Data Analysis*.
- Dunst, C. J., Gorman, E., & Hamby, D. W. (2012). *Preference for Infant-Directed Speech in Preverbal Young Children*. *5*(1).
- Eaves Jr., B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant directed speech is consistent with teaching. *ArXiv:1606.01175 [Cs]*.
<http://arxiv.org/abs/1606.01175>
- Elan* (5.0). (2019). [Computer software]. Max Planck Institute for Psycholinguistics.
- Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*(6), 581–604.
[https://doi.org/10.1016/S0149-7634\(00\)00025-7](https://doi.org/10.1016/S0149-7634(00)00025-7)

- Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*(7), 878–879.
<https://doi.org/10.1038/nn1729>
- Fanelli, D. (2010). Do Pressures to Publish Increase Scientists' Bias? An Empirical Support from US States Data. *PLoS ONE*, *5*(4), e10271.
<https://doi.org/10.1371/journal.pone.0010271>
- Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605. <https://doi.org/10.1073/pnas.152159999>
- Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences*, *102*(47), 17245–17250. <https://doi.org/10.1073/pnas.0502205102>
- Farroni, Teresa, Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze Following in Newborns. *Infancy*, *5*(1), 39–60.
- Farroni, Teresa, Menon, E., & Johnson, M. H. (2006). Factors influencing newborns' preference for faces with eye contact. *Journal of Experimental Child Psychology*, *95*(4), 298–308.
<https://doi.org/10.1016/j.jecp.2006.08.001>
- Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Development*, *60*(6), 1497–1510.
- Fernald, A. (1991). Prosody in speech to children: Prelinguistic and linguistic functions. *R Annals of Child Development*, *8*, 43–80.

- Fernald, A. (1993). Approval and Disapproval: Infant Responsiveness to Vocal Affect in Familiar and Unfamiliar Languages. *Child Development*, 64, 657–674.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychological Science*, 12(6), 499–504. <https://doi.org/10.1111/1467-9280.00392>
- Flom, R., Deák, G. O., Phill, C. G., & Pick, A. D. (2004). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior and Development*, 27(2), 181–194. <https://doi.org/10.1016/j.infbeh.2003.09.007>
- Flynn, E., Turner, C., & Giraldeau, L.-A. (2016). Selectivity in social and asocial learning: Investigating the prevalence, effect and development of young children's learning preferences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1690), 20150189. <https://doi.org/10.1098/rstb.2015.0189>
- Flynn, E., & Whiten, A. (2012). Experimental “Microcultures” in Young Children: Identifying Biographic, Cognitive, and Social Predictors of Information Transmission: Identifying Predictors of Information Transmission. *Child Development*, 83(3), 911–925. <https://doi.org/10.1111/j.1467-8624.2012.01747.x>
- Forstmeier, W., Wagenmakers, E.-J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings - a practical guide: Avoiding false-positive findings. *Biological Reviews*, 92(4), 1941–1968. <https://doi.org/10.1111/brv.12315>

- Franco, F. (2005). Infant pointing: Harlequin, servant of two masters. In *Joint attention: Communication and other minds. Problems in philosophy and psychology* (pp. 129–164). Oxford University Press.
- Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S., & Yurovsky, D. (2017). A Collaborative Approach to Infant Research: Promoting Reproducibility, Best Practices, and Theory-Building. *Infancy*, 22(4), 421–435. <https://doi.org/10.1111/infa.12182>
- Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133(4), 694–724. <https://doi.org/10.1037/0033-2909.133.4.694>
- Gale, A., Spratt, G., Chapman, A. J., & Smallbone, A. (1975). EEG correlates of eye contact and interpersonal distance. *Biological Psychology*, 3(4), 237–245. [https://doi.org/10.1016/0301-0511\(75\)90023-X](https://doi.org/10.1016/0301-0511(75)90023-X)
- Gelman, S. A., Ware, E. A., Manczak, E. M., & Graham, S. A. (2013). Children's sensitivity to the knowledge expressed in pedagogical and nonpedagogical contexts. *Developmental Psychology*, 49(3), 491–504. <https://doi.org/10.1037/a0027901>
- Gergely, G., Egyed, K., & Király, I. (2007). On pedagogy. *Developmental Science*, 10(1), 139–146. <https://doi.org/10.1111/j.1467-7687.2007.00576.x>
- Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on 'Estimating the reproducibility of psychological science'. *Science*, 351(6277), 1037–1037. <https://doi.org/10.1126/science.aad7243>

- Gliga, T., & Csibra, G. (2009). One-Year-Old Infants Appreciate the Referential Nature of Deictic Gestures and Words. *Psychological Science*, *20*(3), 347–353. <https://doi.org/10.1111/j.1467-9280.2009.02295.x>
- Gopnik, A., O’Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., Aboody, R., Fung, H., & Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences*, *114*(30), 7892–7899. <https://doi.org/10.1073/pnas.1700811114>
- Gredebäck, G., Astor, K., & Fawcett, C. (2018). Gaze Following Is Not Dependent on Ostensive Cues: A Critical Test of Natural Pedagogy. *Child Development*, *89*(6), 2091–2098. <https://doi.org/10.1111/cdev.13026>
- Gredebäck, G., Theuring, C., Hauf, P., & Kenward, B. (2008). The Microstructure of Infants’ Gaze as They View Adult Shifts in Overt Attention. *Infancy*, *13*(5), 533–543. <https://doi.org/10.1080/15250000802329529>
- Grossmann, T. (2017). The Eyes as Windows Into Other Minds: An Integrative Perspective. *Perspectives on Psychological Science*, *12*(1), 107–121. <https://doi.org/10.1177/1745691616654457>
- Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., & Csibra, G. (2008). Early cortical specialization for face-to-face communication in human infants. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1653), 2803–2811. <https://doi.org/10.1098/rspb.2008.0986>
- Grossmann, T., Parise, E., & Friederici, A. D. (2010). The Detection of Communicative Signals Directed at the Self in Infant Prefrontal Cortex.

Frontiers in Human Neuroscience, 4.

<https://doi.org/10.3389/fnhum.2010.00201>

Guest, O., & Martin, A. (2020). How computational modeling can force theory building in psychological science. *PsyArXiv*.

<https://doi.org/10.31234/osf.io/rybh9>

Gweon, H., & Schulz, L. (2019). From Exploration to Instruction: Children Learn From Exploration and Tailor Their Demonstrations to Observers' Goals and Competence. *Child Development*, 90(1), e148–e164.

<https://doi.org/10.1111/cdev.13059>

Háden, G. P., Mády, K., Török, M., & Winkler, I. (2020). Newborn infants differently process adult directed and infant directed speech.

International Journal of Psychophysiology, 147, 107–112.

<https://doi.org/10.1016/j.ijpsycho.2019.10.011>

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development*, 21(2), 167–179.

[https://doi.org/10.1016/S0163-6383\(98\)90001-7](https://doi.org/10.1016/S0163-6383(98)90001-7)

Hamilton, A. F. de C. (2016). Gazing at me: The importance of social meaning in understanding direct-gaze cues. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1686), 20150080.

<https://doi.org/10.1098/rstb.2015.0080>

Harris, P. L. (2012). *Trusting what you're told*. Cambridge: Belknap/Harvard.

Haun, D. B. M., Call, J., Janzen, G., & Levinson, S. C. (2006). Evolutionary Psychology of Spatial Representations in the Hominidae. *Current Biology*,

16(17), 1736–1740. <https://doi.org/10.1016/j.cub.2006.07.049>

- Helminen, T. M., Kaasinen, S. M., & Hietanen, J. K. (2011). Eye contact and arousal: The effects of stimulus duration. *Biological Psychology*, *88*(1), 124–130.
<https://doi.org/10.1016/j.biopsycho.2011.07.002>
- Hernik, M., & Broesch, T. (2018). Infant gaze following depends on communicative signals: An eye-tracking study of 5- to 7-month-olds in Vanuatu. *Developmental Science*, e12779.
<https://doi.org/10.1111/desc.12779>
- Heyes, C. (2016). Born Pupils? Natural Pedagogy and Cultural Pedagogy. *Perspectives on Psychological Science*, *11*(2), 280–295.
<https://doi.org/10.1177/1745691615621276>
- Heyes, C. (2017). When does social learning become cultural learning? *Developmental Science*, *20*(2), e12350.
<https://doi.org/10.1111/desc.12350>
- Hirai, M., & Kanakogi, Y. (2018). Communicative Hand-Waving Gestures Facilitate Object Learning in Preverbal Infants. *Developmental Science*, e12787. <https://doi.org/10.1111/desc.12787>
- Hirsh-Pasek, K., Golinkoff, R. M., & Hollich, G. (2000). An emergentist coalition model for word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 136–164). Oxford University Press.
- Hoehl, S., Michel, C., Reid, V. M., Parise, E., & Striano, T. (2014). Eye contact during live social interaction modulates infants' oscillatory brain activity. *Social Neuroscience*, *9*(3), 300–308.
<https://doi.org/10.1080/17470919.2014.884982>

- Hood, B. M., Macrae, C. N., Cole-Davies, V., & Dias, M. (2003). Eye remember you: The effects of gaze direction on face recognition in children and adults. *Developmental Science*, 6(1), 67–71. <https://doi.org/10.1111/1467-7687.00256>
- Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, 48(4), 1393–1409. <https://doi.org/10.3758/s13428-015-0647-3>
- Inbar, Y. (2016). Association between contextual dependence and replicability in psychology may be spurious. *Proceedings of the National Academy of Sciences*, 113(34), E4933–E4934. <https://doi.org/10.1073/pnas.1608676113>
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), 6.
- Ioannidis, J. P. A. (2012). Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science*, 7(6), 645–654. <https://doi.org/10.1177/1745691612464056>
- Itier, R. J., Van Roon, P., & Alain, C. (2011). Species sensitivity of early face and eye processing. *NeuroImage*, 54(1), 705–713. <https://doi.org/10.1016/j.neuroimage.2010.07.031>
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press, Oxford, UK.
- Jessen, S., & Grossmann, T. (2014). Unconscious discrimination of social cues from eye whites in infants. *Proceedings of the National Academy of Sciences*, 111(12), 4373–4378. <https://doi.org/10.1073/pnas.1312500111>

Sciences, 111(45), 16208–16213.

<https://doi.org/10.1073/pnas.1411333111>

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling.

Psychological Science, 23(5), 524–532.

<https://doi.org/10.1177/0956797611430953>

Johnson, M. H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience & Biobehavioral Reviews*, 50, 169–179.

<https://doi.org/10.1016/j.neubiorev.2014.10.009>

Johnson, V. E., & Rossell, D. (2010). On the use of non-local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(2), 143–170.

<https://doi.org/10.1111/j.1467-9868.2009.00730.x>

Kaiser, D., & Losick, R. (1993). How and why bacteria talk to each other. *Cell*, 73(5), 873–885. [https://doi.org/10.1016/0092-8674\(93\)90268-U](https://doi.org/10.1016/0092-8674(93)90268-U)

Kampe, K. K. W., Frith, C. D., & Frith, U. (2003). “Hey John”: Signals Conveying Communicative Intention toward the Self Activate Brain Regions Associated with “Mentalizing,” Regardless of Modality. *The Journal of Neuroscience*, 23(12), 5258–5263.

<https://doi.org/10.1523/JNEUROSCI.23-12-05258.2003>

Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature*

Communications, 2(1). <https://doi.org/10.1038/ncomms1342>

- Kathawalla, U., Silverstein, P., & Syed, M. (2020). Easing Into Open Science: A Guide for Graduate Students and Their Advisors. *PsyArXiv*.
<https://doi.org/10.31234/osf.io/vzjdp>
- Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Ito, K., Fukuda, H., Kojima, S., & Nakamura, K. (1999). The human amygdala plays an important role in gaze monitoring. *Brain*, *122*(4), 779–783.
<https://doi.org/10.1093/brain/122.4.779>
- Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217.
- Kibbe, M. M., & Leslie, A. M. (2011). What Do Infants Remember When They Forget? Location and Identity in 6-Month-Olds' Memory for Objects. *Psychological Science*, *22*(12), 1500–1505.
<https://doi.org/10.1177/0956797611420165>
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLoS ONE*, *7*(5), e36399.
<https://doi.org/10.1371/journal.pone.0036399>
- Kim, S., Kalish, C. W., Weisman, K., Johnson, M. V., & Shutts, K. (2016). Young Children Choose to Inform Previously Knowledgeable Others. *Journal of Cognition and Development*, *17*(2), 320–340.
<https://doi.org/10.1080/15248372.2014.952731>
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. [https://doi.org/10.1016/S0010-0277\(02\)00004-5](https://doi.org/10.1016/S0010-0277(02)00004-5)

- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., ... Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443–490. <https://doi.org/10.1177/2515245918810225>
- Kleinke, C. L., & Pohlen, P. D. (1971). Affective and emotional responses as a function of other person's gaze and cooperativeness in a two-person game. *Journal of Personality and Social Psychology*, 17(3), 308–313. <https://doi.org/10.1037/h0030600>
- Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain Sciences*, 38. <https://doi.org/10.1017/S0140525X14000090>
- Krøjgaard, P. (2009). The human ability to single out and track specific objects through space and time: Origin and application. In H. Høgh-Olesen, P. Bertelsen, & J. Tønnesvang, *Human characteristics: Evolutionary perspectives on human mind and kind* (pp. 89–116). Cambridge Scholars Publishing.
- Kuhn, P. K. (1962). *The structure of scientific revolutions*. University of Chicago Press.
- Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., & Dienes, Z. (2018). Improving Inferences About Null Effects With Bayes Factors and Equivalence Tests. *Journals of Gerontology: Psychological Sciences*, 1–13.

- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
<http://dx.doi.org/10.1017/CBO9781139087759>
- Lempers, J. D. (1979). Young children's production and comprehension of nonverbal deictic behaviors. *Journal of Genetic Psychology*, *135*(1), 93–102.
- Leung, E. H. L., & Rheingold, H. L. (1981). Development of Pointing as a Social Gesture. *Infant Behavior & Development*, *30*(4), 562–567.
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*(3), 732–739.
<https://doi.org/10.1016/j.cognition.2008.06.013>
- Lloyd-Fox, S., Széplaki-Köllöd, B., Yin, J., & Csibra, G. (2015). Are you talking to me? Neural activations in 6-month-old infants in response to being addressed during natural interactions. *Cortex*, *70*, 35–48.
<https://doi.org/10.1016/j.cortex.2015.02.005>
- Lyons, D. E., Damrosch, D. H., Lin, J. K., Macris, D. M., & Keil, F. C. (2011). The scope and limits of overimitation in the transmission of artefact culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1158–1167. <https://doi.org/10.1098/rstb.2010.0335>
- Mares, I., Smith, M. L., Johnson, M. H., & Senju, A. (2016). Direct gaze facilitates rapid orienting to faces: Evidence from express saccades and saccadic potentials. *Biological Psychology*, *121*, 84–90.
<https://doi.org/10.1016/j.biopsycho.2016.10.003>

- Mareschal, D., & Johnson, M. H. (2003). The “what” and “where” of object representations in infancy. *Cognition*, *88*(3), 259–276.
[https://doi.org/10.1016/S0010-0277\(03\)00039-8](https://doi.org/10.1016/S0010-0277(03)00039-8)
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.
<https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Marno, H., & Csibra, G. (2015). Toddlers Favor Communicatively Presented Information over Statistical Reliability in Learning about Artifacts. *PLOS ONE*, *10*(3), e0122129. <https://doi.org/10.1371/journal.pone.0122129>
- Marno, H., Davelaar, E. J., & Csibra, G. (2014). Nonverbal communicative signals modulate attention to object properties. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 752–762.
<https://doi.org/10.1037/a0035113>
- Marno, H., Davelaar, E. J., & Csibra, G. (2016). An object memory bias induced by communicative reference. *Acta Psychologica*, *163*, 88–96.
<https://doi.org/10.1016/j.actpsy.2015.11.005>
- Mather, E., & Plunkett, K. (2012). The Role of Novelty in Early Word Learning. *Cognitive Science*, *36*(7), 1157–1177. <https://doi.org/10.1111/j.1551-6709.2012.01239.x>
- Meltzoff, A. N. (1988a). Imitation of Televised Models by Infants. *Child Development*, *59*(5), 1221–1229.
- Meltzoff, A. N. (1988b). Infant Imitation and Memory: Nine-Month-Olds in Immediate and Deferred Tests. *Child Development*, *59*(1), 217–225.

- Meltzoff, A. N. (2008). The case for developmental cognitive science: Theories of people and things. In *Theories of infant development* (pp. 143–173). Blackwell Publishing Ltd.
- Meltzoff, A. N., & Moore, M. K. (1989). Imitation in Newborn Infants: Exploring the Range of Gestures Imitated and the Underlying Mechanisms. *Developmental Psychology*, *25*(6), 954–962.
- Mendelson, M. J., Haith, M. M., & Goldman-Rakic, P. S. (1982). Face Scanning and Responsiveness to Social Cues in Infant Rhesus Monkeys. *Developmental Psychology*, *18*(2), 222–228.
- Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science*, *7*(1), F1–F9.
<https://doi.org/10.1111/j.1467-7687.2004.00315.x>
- Moore, C., Angelopoulos, M., & Bennett, P. (1997). The role of movement in the development of joint visual attention. *Infant Behavior and Development*, *20*(1), 83–92. [https://doi.org/10.1016/S0163-6383\(97\)90063-1](https://doi.org/10.1016/S0163-6383(97)90063-1)
- Moore, C., & Corkum, V. (1994). Social understanding at the end of the first year. *Developmental Review*, *14*, 349–372.
- Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Neal, R., & Schwartz, H. K. (2000). Gaze following, temperament, and language development in 6-month-olds: A replication and extension. *Infant Behavior and Development*, *23*(2), 231–236. [https://doi.org/10.1016/S0163-6383\(01\)00038-8](https://doi.org/10.1016/S0163-6383(01)00038-8)
- Morey, R. D., & Rouder, J. N. (2015). *Bayes factor: Computation of Bayes factors for common designs. R package version 0.9.12-2*. <https://CRAN.R-project.org/package=BayesFactor>

- Morgan, T. J. H., Laland, K. N., & Harris, P. L. (2015). The development of adaptive conformity in young children: Effects of uncertainty and consensus. *Developmental Science, 18*(4), 511–524. <https://doi.org/10.1111/desc.12231>
- Morissette, P., Richard, M., & Décarie, T. G. (1995). Joint visual attention and pointing in infancy: A longitudinal study of comprehension. *British Journal of Developmental Psychology, 13*(2), 163–175.
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1). <https://doi.org/10.1038/s41562-016-0021>
- Murphy, C. M., & Messer, D. J. (1977). Mothers, infants and pointing: A study of a gesture. In *Studies in mother-infant interaction*. Academic Press.
- Myowa-Yamakoshi, M., & Tomonaga, M. (2001). Development of face recognition in an infant gibbon (*Hylobates agilis*). *Infant Behavior and Development, 24*(2), 215–227. [https://doi.org/10.1016/S0163-6383\(01\)00076-5](https://doi.org/10.1016/S0163-6383(01)00076-5)
- Nagai, Y., Asada, M., & Hosoda, K. (2006). Learning for joint attention helped by functional development. *Advanced Robotics, 20*(10), 1165–1181. <https://doi.org/10.1163/156855306778522497>
- Nakao, H., & Andrews, K. (2014). Ready to Teach or Ready to Learn: A Critique of the Natural Pedagogy Theory. *Review of Philosophy and Psychology, 5*(4), 465–483. <https://doi.org/10.1007/s13164-014-0187-2>
- Naoui, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., Nakamura, K., Yamamoto, J., & Kojima, S. (2012). Cerebral responses to infant-directed

- speech and the effect of talker familiarity. *NeuroImage*, 59(2), 1735–1744.
<https://doi.org/10.1016/j.neuroimage.2011.07.093>
- Newcombe, N. S. (2002). The Nativist-Empiricist Controversy in the Context of Recent Research on Spatial and Quantitative Development. *Psychological Science*, 13(5), 395–401. <https://doi.org/10.1111/1467-9280.00471>
- Newman, R. S., & Hussain, I. (2006). Changes in Preference for Infant-Directed Speech in Low and Moderate Noise by 4.5- to 13-Month-Olds. *Infancy*, 10(1), 61–76. https://doi.org/10.1207/s15327078in1001_4
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48(2), 127–162.
[https://doi.org/10.1016/S0010-0285\(03\)00128-2](https://doi.org/10.1016/S0010-0285(03)00128-2)
- Nichols, K. A., & Champness, B. G. (1971). Eye gaze and the GSR. *Journal of Experimental Social Psychology*, 7(6), 623–626.
[https://doi.org/10.1016/0022-1031\(71\)90024-2](https://doi.org/10.1016/0022-1031(71)90024-2)
- Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, 42(3), 555–565.
<https://doi.org/10.1037/0012-1649.42.3.555>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606.
<https://doi.org/10.1073/pnas.1708274114>
- Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, 26(5), 1596–1618.
<https://doi.org/10.3758/s13423-019-01645-2>

- Okamoto-Barth, S., Call, J., & Tomasello, M. (2007). Great Apes' Understanding of Other Individuals' Line of Sight. *Psychological Science*, *18*(5), 462–468.
<https://doi.org/10.1111/j.1467-9280.2007.01922.x>
- Okumura, Y., Kobayashi, T., & Itakura, S. (2016). Eye Contact Affects Object Representation in 9-Month-Old Infants. *PLOS ONE*, *11*(10), e0165145.
<https://doi.org/10.1371/journal.pone.0165145>
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, *308*(5719), 255–258.
<https://doi.org/10.1126/science.1107621>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716.
<https://doi.org/10.1126/science.aac4716>
- Over, H., & Carpenter, M. (2012). Putting the social into social learning: Explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology*, *126*(2), 182–192.
<https://doi.org/10.1037/a0024555>
- Papoušek, M., Papoušek, H., & Haekel, M. (1987). Didactic adjustments in fathers' and mothers' speech to their 3-month-old infants. *Journal of Psycholinguistic Research*, *16*(5), 491–516.
<https://doi.org/10.1007/BF01073274>
- Parise, E., & Csibra, G. (2013). Neural Responses to Multimodal Ostensive Signals in 5-Month-Old Infants. *PLoS ONE*, *8*(8), e72360.
<https://doi.org/10.1371/journal.pone.0072360>
- Patil, P., Peng, R. D., & Leek, J. T. (2016). What Should Researchers Expect When They Replicate Studies? A Statistical View of Replicability in Psychological

- Science. *Perspectives on Psychological Science*, 11(4), 539–544.
<https://doi.org/10.1177/1745691616646366>
- Paulus, M., Kim, S., & Sodian, B. (2015). Clarifying the range of social-cognitive processes subserving human teaching. *Behavioral and Brain Sciences*, 38.
- Paulus, Markus, Hunnius, S., Vissers, M., & Bekkering, H. (2011). Imitation in Infancy: Rational or Motor Resonance?: Imitation in Infancy. *Child Development*, 82(4), 1047–1057. <https://doi.org/10.1111/j.1467-8624.2011.01610.x>
- Perrett, D. I., & Emery, N. J. (1994). Understanding the intentions of others from visual signals: Neuropsychological evidence. *CPC*, 13(5), 683–694.
- Peter, V., Kalashnikova, M., Santos, A., & Burnham, D. (2016). Mature neural responses to Infant-Directed Speech but not Adult-Directed Speech in Pre-Verbal Infants. *Scientific Reports*, 6(1).
<https://doi.org/10.1038/srep34273>
- Pinkham, A. M., & Jaswal, V. K. (2011). Watch and Learn? Infants Privilege Efficiency Over Pedagogy During Imitative Learning: EFFICIENCY AND PEDAGOGY IN IMITATIVE LEARNING. *Infancy*, 16(5), 535–544.
<https://doi.org/10.1111/j.1532-7078.2010.00059.x>
- Popper, K. (1963). *Conjectures and refutations: The growth of scientific knowledge*. Routledge.
- Povinelli, D. J., Eddy, T. J., Hobson, R. P., & Tomasello, M. (1996). What Young Chimpanzees Know about Seeing. *Monographs of the Society for Research in Child Development*, 61(3), i. <https://doi.org/10.2307/1166159>
- Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The Birth of Words: Ten-Month-Olds Learn Words Through Perceptual Salience.

- Child Development*, 77(2), 266–280. <https://doi.org/10.1111/j.1467-8624.2006.00869.x>
- R Core Team. (2017). *R* (3.5.2) [Computer software].
- Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising? – Linking properties of IDS to statistical learning and attention at the prosodic level. *Cognition*, 178, 193–206. <https://doi.org/10.1016/j.cognition.2018.05.015>
- Rhodes, M., Bonawitz, E., Shafto, P., Chen, A., & Caglar, L. (2015). Controlling the message: Preschoolers' use of information to teach and deceive others. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00867>
- Rohlfing, K. J., Longo, M. R., & Bertenthal, B. I. (2012). Dynamic pointing triggers shifts of visual attention in young infants: Dynamic pointing triggers shifts of visual attention. *Developmental Science*, 15(3), 426–435. <https://doi.org/10.1111/j.1467-7687.2012.01139.x>
- Ronfard, S., & Harris, P. L. (2018). Children's Decision to Transmit Information is Guided by their Evaluation of the Nature of that Information. *Review of Philosophy and Psychology*, 9(4), 849–861. <https://doi.org/10.1007/s13164-017-0344-5>
- Ronfard, S., Was, A. M., & Harris, P. L. (2016). Children teach methods they could not discover for themselves. *Journal of Experimental Child Psychology*, 142, 107–117. <https://doi.org/10.1016/j.jecp.2015.09.032>
- Rosa Salva, O., Regolin, L., & Vallortigara, G. (2012). Inversion of contrast polarity abolishes spontaneous preferences for face-like stimuli in newborn chicks. *Behavioural Brain Research*, 228(1), 133–143. <https://doi.org/10.1016/j.bbr.2011.11.025>

- Rothkirch, M., Madipakkam, A. R., Rehn, E., & Sterzer, P. (2015). Making eye contact without awareness. *Cognition*, *143*, 108–114.
<https://doi.org/10.1016/j.cognition.2015.06.012>
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. <https://doi.org/10.1016/j.jmp.2012.08.001>
- Sage, K. D., & Baldwin, D. (2011). Disentangling the Social and the Pedagogical in Infants' Learning about Tool-Use: Effects of Social Cues on Infant Tool-use. *Social Development*, *20*(4), 825–844. <https://doi.org/10.1111/j.1467-9507.2011.00624.x>
- Saito, Y., Aoyama, S., Kondo, T., Fukumoto, R., Konishi, N., Nakamura, K., Kobayashi, M., & Toshima, T. (2007). Frontal cerebral blood flow change associated with infant-directed speech. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *92*(2), F113–F116.
<https://doi.org/10.1136/adc.2006.097949>
- Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266.
- Schmidt, M. F. H., Butler, L. P., Heinz, J., & Tomasello, M. (2016). Young Children See a Single Action and Infer a Social Norm: Promiscuous Normativity in 3-Year-Olds. *Psychological Science*, *27*(10), 1360–1370.
<https://doi.org/10.1177/0956797616661182>
- Schmitow, C., Kochukhova, O., & Nyström, P. (2016). Social perception: How do 6-month-old infants look at pointing gestures? *Infant Behavior and Development*, *42*, 152–156. <https://doi.org/10.1016/j.infbeh.2015.12.001>

- Schönbrodt, F. D., Zehetleitner, M., Wagenmakers, E.-J., & Perugini, M. (2017). Sequential Hypothesis Testing With Bayes Factors: Efficiently Testing Mean Differences. *Psychological Methods*, *22*(2), 322–339.
- Seehagen, S., Schneider, S., Miebach, K., Frigge, K., & Zmyj, N. (2017). “Should I or shouldn’t I?” Imitation of undesired versus allowed actions from peer and adult models by 18- and 24-month-old toddlers. *Infant Behavior and Development*, *49*, 1–8. <https://doi.org/10.1016/j.infbeh.2017.06.001>
- Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, *18*(9), 668–671. <https://doi.org/10.1016/j.cub.2008.03.059>
- Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants’ preference for object-directed gaze. *Cognition*, *108*(2), 303–319. <https://doi.org/10.1016/j.cognition.2008.02.009>
- Senju, A., & Johnson, M. H. (2009). The eye contact effect: Mechanisms and development. *Trends in Cognitive Sciences*, *13*(3), 127–134. <https://doi.org/10.1016/j.tics.2008.11.009>
- Senju, A., Verneti, A., Ganea, N., Hudry, K., Tucker, L., Charman, T., & Johnson, M. H. (2015). Early Social Experience Affects the Development of Eye Gaze Processing. *Current Biology*, *25*(23), 3086–3091. <https://doi.org/10.1016/j.cub.2015.10.019>
- Silverstein, P., Gliga, T., Westermann, G., & Parise, E. (2019). Probing communication-induced memory biases in preverbal infants: Two replication attempts of Yoon, Johnson and Csibra (2008). *Infant Behavior and Development*, *55*, 77–87. <https://doi.org/10.1016/j.infbeh.2019.03.005>

- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366.
<https://doi.org/10.1177/0956797611417632>
- Simonsohn, U. (2015). Small Telescopes: Detectability and the Evaluation of Replication Results. *Psychological Science*, *26*(5), 559–569.
<https://doi.org/10.1177/0956797614567341>
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*(9), 160384.
<https://doi.org/10.1098/rsos.160384>
- Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually cued attention and early word learning. *Cognitive Science*, *34*(7), 1287–1314.
- Smith, Linda B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343–348. [https://doi.org/10.1016/S1364-6613\(03\)00156-6](https://doi.org/10.1016/S1364-6613(03)00156-6)
- Snow, C. E., & Ferguson, C. A. (1977). *Talking to Children: Language Input and Acquisition*. Cambridge University Press.
<https://www.jstor.org/stable/412603?origin=crossref>
- Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, *120*(4), 779–797. <https://doi.org/10.1037/a0034191>
- Sobel, D. M., & Letourneau, S. M. (2015). Children’s developing understanding of what and how they learn. *Journal of Experimental Child Psychology*, *132*, 221–229. <https://doi.org/10.1016/j.jecp.2015.01.004>

- Spencer, J. P., Dineva, E., & Smith, L. B. (2009). Comment on 'Infants' Perseverative Search Errors Are Induced by Pragmatic Misinterpretation'. *Science*, *325*(5948), 1624–1624.
<https://doi.org/10.1126/science.1172759>
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Blackwell Publishers.
- Stoyanova, R. S., Ewbank, M. P., & Calder, A. J. (2010). "You Talkin' to Me?": Self-Relevant Auditory Signals Influence Perception of Gaze Direction. *Psychological Science*, *21*(12), 1765–1769.
<https://doi.org/10.1177/0956797610388812>
- Striano, T., Chen, X., Cleveland, A., & Bradshaw, S. (2006). Joint attention social cues influence infant learning. *European Journal of Developmental Psychology*, *3*(3), 289–299.
<https://doi.org/10.1080/17405620600879779>
- Sulpizio, S., Doi, H., Bornstein, M. H., Cui, J., Esposito, G., & Shinohara, K. (2018). fNIRS reveals enhanced brain activation to female (versus male) infant directed speech (relative to adult directed speech) in Young Human Infants. *Infant Behavior and Development*, *52*, 89–96.
<https://doi.org/10.1016/j.infbeh.2018.05.009>
- Świątkowski, W., & Dompnier, B. (2017). Replicability Crisis in Social Psychology: Looking at the Past to Find New Pathways for the Future. *International Review of Social Psychology*, *30*(1), 111. <https://doi.org/10.5334/irsp.66>
- Szufnarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2014). Is ostension any more than attention? *Scientific Reports*, *4*(1).
<https://doi.org/10.1038/srep05304>

- Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference. *Science*, 332(6033), 1054–1059.
- The ManyBabies Consortium. (2019). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science (AMPPS)*.
<https://doi.org/10.31234/osf.io/s98ab>
- Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.
- Tomasello, Michael. (2008). *Origins of Human Communication*. MIT Press.
- Tomasello, Michael. (2016). The ontogeny of cultural learning. *Current Opinion in Psychology*, 8, 1–4. <https://doi.org/10.1016/j.copsy.2015.09.008>
- Tomasello, Michael, Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(05).
<https://doi.org/10.1017/S0140525X05000129>
- Tomasello, Michael, Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution*, 52(3), 314–320.
<https://doi.org/10.1016/j.jhevol.2006.10.001>
- Topal, J., Gergely, G., Miklosi, A., Erdohegyi, A., & Csibra, G. (2008). Infants' Perseverative Search Errors Are Induced by Pragmatic Misinterpretation. *Science*, 321(5897), 1831–1834.
<https://doi.org/10.1126/science.1161437>

- Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is Infant-Directed Speech Prosody a Result of the Vocal Expression of Emotion? *Psychological Science*, *11*(3), 188–195. <https://doi.org/10.1111/1467-9280.00240>
- Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335–340. <https://doi.org/10.3758/BF03196290>
- Triesch, J., Teuscher, C., Deak, G. O., & Carlson, E. (2006). Gaze following: Why (not) learn it? *Developmental Science*, *9*(2), 125–147. <https://doi.org/10.1111/j.1467-7687.2006.00470.x>
- Urakawa, S., Takamoto, K., Ishikawa, A., Ono, T., & Nishijo, H. (2015). Selective Medial Prefrontal Cortex Responses During Live Mutual Gaze Interactions in Human Infants: An fNIRS Study. *Brain Topography*, *28*(5), 691–701. <https://doi.org/10.1007/s10548-014-0414-2>
- Vazire, S. (2017). Quality Uncertainty Erodes Trust in Science. *Collabra: Psychology*, *3*(1), 1. <https://doi.org/10.1525/collabra.74>
- Vecera, S. P., & Johnson, M. H. (1995). Gaze detection and the cortical processing of faces: Evidence from infants and adults. *Visual Cognition*, *2*(1), 59–87. <https://doi.org/10.1080/13506289508401722>
- von Hofsten, C., Dahlström, E., & Fredriksson, Y. (2005). 12-Month-Old Infants' Perception of Attention Direction in Static Video Images. *Infancy*, *8*(3), 217–231. https://doi.org/10.1207/s15327078in0803_2
- Vredenburgh, C., Kushnir, T., & Casasola, M. (2015). Pedagogical cues encourage toddlers' transmission of recently demonstrated functions to unfamiliar adults. *Developmental Science*, *18*(4), 645–654. <https://doi.org/10.1111/desc.12233>

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*.
Harvard University.

Waismeyer, A., & Meltzoff, A. N. (2017). Learning to make things happen: Infants' observational learning of social and physical causal events. *Journal of Experimental Child Psychology*, *162*, 58–71.
<https://doi.org/10.1016/j.jecp.2017.04.018>

Whiten, A., & Flynn, E. (2010). The transmission and evolution of experimental microcultures in groups of young children. *Developmental Psychology*, *46*(6), 1694–1709. <https://doi.org/10.1037/a0020786>

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*, *7*.
<https://doi.org/10.3389/fpsyg.2016.01832>

Williams, G. P., & Kleinke, C. L. (1993). *Effects of Mutual Gaze and Touch on Attraction, Mood, and Cardiovascular Reactivity*. *27*, 170–183.

Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. [https://doi.org/10.1016/S0010-0277\(98\)00058-4](https://doi.org/10.1016/S0010-0277(98)00058-4)

Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, *107*(2), 118–136.
<https://doi.org/10.1016/j.jecp.2010.04.014>

- Wu, R., Tummeltshammer, K. S., Gliga, T., & Kirkham, N. Z. (2014). Ostensive signals support learning from novel attention cues during infancy. *Frontiers in Psychology, 5*. <https://doi.org/10.3389/fpsyg.2014.00251>
- Xu, F., & Carey, S. (1996). Infants' Metaphysics: The Case of Numerical Identity. *Cognitive Psychology, 30*(2), 111–153. <https://doi.org/10.1006/cogp.1996.0005>
- Xu, F., & Kushnir, T. (2013). Infants Are Rational Constructivist Learners. *Current Directions in Psychological Science, 22*(1), 28–32. <https://doi.org/10.1177/0963721412469396>
- Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences, 105*(36), 13690–13695. <https://doi.org/10.1073/pnas.0804388105>
- Yu, C., & Smith, L. B. (2013). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLoS ONE, 8*(11), e79659. <https://doi.org/10.1371/journal.pone.0079659>
- Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: Salience and social cues in early word learning. *Developmental Science, 20*(2), e12349. <https://doi.org/10.1111/desc.12349>
- Zangl, R., & Mills, D. L. (2007). Increased Brain Activity to Infant-Directed Speech in 6- and 13-Month-Old Infants. *Infancy, 11*(1), 31–62. https://doi.org/10.1207/s15327078in1101_2

- Zeifman, D., Delaney, S., & Blass, E. M. (1996). Sweet taste, looking, and calm in 2- and 4-week-old infants: The eyes have it. *Developmental Psychology*, 32(6), 1090–1099. <https://doi.org/10.1037/0012-1649.32.6.1090>
- Ziv, M., & Frye, D. (2004). Children's understanding of teaching: The role of knowledge and belief. *Cognitive Development*, 19(4), 457–477. <https://doi.org/10.1016/j.cogdev.2004.09.002>
- Zuberbühler, K. (2001). Predator-specific alarm calls in Campbell's monkeys, *Cercopithecus campbelli*. *Behavioral Ecology and Sociobiology*, 50(5), 414–422. <https://doi.org/10.1007/s002650100383>