Learning to reason about desires: An infant training study

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Learning to reason about desires: An infant training study

Tiffany Doan1 (tidoan@uwaterloo.ca), Stephanie Denison1 (stephanie.denison@uwaterloo.ca), Christopher G. Lucas2,3 (c.lucas@ed.ac.uk), & Alison Gopnik1 (gopnik@berkeley.edu)

1University of Waterloo, Department of Psychology
2University of Edinburgh, School of Informatics
3University of California, Berkeley, Department of Psychology

Abstract

A key aspect of theory of mind is the ability to reason about other people’s desires. As adults, we know that desires and preferences are subjective and specific to the individual. However, research in cognitive development suggests that a significant conceptual shift occurs in desire-based reasoning during infancy and early childhood, as acts, and to predict their future behavior. These abilities number of important functions. It allows us to please or mental states, such as their desires and beliefs, serves a what other people are thinking. The ability to infer others’ Keywords: learning; Social cognition; Preferences.

Introduction

As social creatures, we are constantly trying to figure out what other people are thinking. The ability to infer others’ mental states, such as their desires and beliefs, serves a number of important functions. It allows us to please or irritate others, to understand why they engage in particular acts, and to predict their future behavior. These abilities hinge on our having a well-developed theory of mind – the understanding that people have mental states (e.g., desires, beliefs, intentions) and that these mental states can differ from person to person (Gopnik & Wellman 1994).

Explicit theory of mind underlies significant development during infancy and early childhood, as children first reason based on knowledge about others’ desires and then later incorporate knowledge about others’ beliefs. How do children arrive at these more sophisticated beliefs about the minds of other people? This paper focuses on the development of desire-based reasoning, or the ability to consider a person’s wants, likes, and dislikes when reflecting on their behavior. For example, children as young as two years of age understand that people’s actions and emotions are influenced by their desires; they know that a person will attend to objects that they want to obtain and will be sad if their desires go unfulfilled (Wellman and Woolley, 1999).

The present experiments examine a shift that occurs in infants’ desire-based reasoning, specifically in their reasoning about preferences. The paradigm is based on a task that asked whether infants understand that preferences can serve as an underlying cause of people’s behaviors (Repacholi & Gopnik, 1997). Fourteen- and eighteen-month-old infants were presented with two different types of food: Goldfish crackers and broccoli. The experimenter determined which food the infants liked (the majority preferred Goldfish crackers). She then demonstrated, using emotional expressions and simple language, that she preferred either that same food (Goldfish crackers in a “matched” trial) or the opposite food (broccoli in an “unmatched” trial), depending on the experimental condition. When infants were asked to share some food with the experimenter, the two age groups differed in their responses. The 18-month-olds were able to correctly determine the experimenter’s preferences based on her previous behaviors, and thus correctly gave her the food that she liked, whether the infant themselves preferred this food or not. However, 14-month-olds gave the experimenter the food that they themselves preferred, regardless of her demonstrated preferences. This difference in performance is interpreted to suggest that around 18 months of age, infants’ desire-based reasoning undergoes a significant conceptual change, moving from a simple to a more complex model of preferences. That is, infants younger than 18 months may have a very simple sense of preferences in which they initially assume that preferences are universal, rather than varying between people1. In contrast, older infants seem to recognize that desires are

What occurs between the ages of 14 and 18 months to promote such a significant advance in Theory of Mind? In a recent paper, Lucas et al. (2014) suggested that infants might first favor the simpler or “universal” model of preferences because it gives a parsimonious explanation for most of the data they encountered. For example, if it is often the case that preferences converge – most people like the taste of pizza but they aren’t as enthusiastic about lima beans. However, as children observe more choices, they have increasingly robust evidence that people have divergent desires. The hypothesis is that as children grow older they accumulate evidence pushing them away from the simple but incorrect initial model toward a more complex and flexible model, which allows them to consider the consequences of distinct preferences. The suggestion is that during this transition, children must observe or participate in many desire-based interactions where people make choices or produce other signals to suggest that their preferences are incongruent with one another or with the infants themselves.

The idea that infants might shift from a simple to a more complex model was formalized as part of a broader look into whether children learn preferences in a way that is rational or optimal under certain assumptions (Lucas et al., 2014). Lucas et al. explored the idea that children have tacit hypotheses about others’ behaviors or underlying mental states, and evaluate those hypotheses against incoming data in a manner consistent with Bayes Theorem. If children expect others to have consistent preferences for options or features (like goldfish crackers, or saltiness) and choose the most attractive option based on the combined desirability of its features – including some features that might be hidden to the child – their preference attributions should be consistent with the predictions of a widely-used economic model, the Mixed Multinomial Logit (MML). The MML is generally used to predict consumer behavior, but it also succeeded in providing a survival account of data from a wide range of experiments on children’s understanding of preferences. It accounts for preschoolers’ ability to accumulate evidence from the statistical properties of a collection of objects and an agent’s choices (Kushnir, Xu, & Wellman, 2010) and for children’s ability to use shared preferences, as well as their knowledge of category membership, as a means for making generalizations (Fawcett & Markson, 2010; see Lucas et al. for details).

This modeling work also yielded an important empirical prediction about the development of desire-based reasoning: if younger children were provided evidence of diverse desires through lab-based training, then they might be able to transition to the more complex model of preference attribution. We tested this hypothesis here using a training study with 14- to 17-month-old infants in two experiments. In Experiment 1 we began by assessing infants understanding of preferences by testing them in a modified version of Repacholi & Gopnik’s Goldfish/Broccoli task. All infants were tested in the critical unmatched trial, wherein the experimenter’s preference conflicted with the infants’ desires. Only infants who failed to give the experimenter the food that she liked continued to a training condition. The critical manipulation is that half of the infants completed a “Diverse Desires Training” condition (henceforth, DDT) where they observed multiple training trials with two experimenters demonstrating different preferences from one another. The other half completed a “Non-Diverse Desires” Training condition (henceforth, N-DDT), where they observed multiple training trials with two experimenters demonstrating the same preferences. Following training, infants were tested again on two unmatched test trials, one directly after training and the other approximately 24 hours later. The second test trial occurred 1 day later to examine how enduring the effects of training might be – would the effect still be evident following a delay? We predict that only infants in the DDT condition should show improved performance in attributing preferences on the test trials.

Experiment 1: Methods

Participants

Infants in both experiments were recruited by phone and email from the California East Bay Area and Southwestern Ontario. In Experiment 1, 55 infants were tested. We used the strict criterion that only infants who did not share the correct item on an initial pre-test (described below) continued to training, increasing confidence that infants completing training did not already know that preferences are diverse. Twenty infants per condition were tested in the full pre-test (10 in DDT, 10 in N-DDT; Range = 14.1 months to 17.5 months; N-DDT, mean age = 15.6 months; Range = 14.4 months to 17.2 months). An additional 15 infants were tested from the pre-test condition due to failing to complete the study because of fussiness (2) or refusing to share on the pre-test and all test trials (13).

Materials

Food: Four sets of food pairs were used in the experiment. The pairs were broccoli and Goldfish crackers, celery and rice puffs, cucumbers and Cherrios, and green peppers and wheel-shaped infant crackers.

Toys: Two sets of toys were used during the training sessions; each set consisted of one type of animal and one type of vehicle in a transparent container. The sets of toys were 4 trucks and 4 dogs, and 4 planes and 4 monkeys. The toys within each type were not identical; they varied in color and shape.

Procedure, Design and Predictions

All infants were tested individually in a quiet lab setting. They sat in a high chair in front of a table and their parent sat in a chair beside them. Before the study began, two experimenters played a passing game with the infant. This allowed the infant to warm up to the experimenters and to

Footnote

1For simplicity, we will characterize younger infants as assuming that preferences are universal. An alternative possibility is that infants are instead sensitive to the relative desirability of objects, reasoning that some things are inherently more desirable than others. This reasoning would also result in infants always sharing the item that they personally like with the experimenter, as they would see it as the objectively better item.
ensure that they could share with the experimenters. The warm up consisted of each experimenter passing a toy (e.g., a ball or toy keys) to the infant and asking her/him to pass it back by placing it in the experimenters’ hands.

Presentation of the pre-training test and post-training test. Infants were assigned to one of the two conditions—infants’ preferences (N-DDT condition) or the control condition (DDT condition). The purpose of the training tasks was to demonstrate that infants would remember which experimenter preferred which objects. We were able to demonstrate that infants used this knowledge about other’s minds, allowing infants to pass the test on Day 2 but not on Day 1. We will address these possibilities more fully in the General Discussion.

Before we can speculate as to why children appeared to learn something new about preferences in the DDT condition, we must first investigate an alternative interpretation of the Experiment 1 data. It is possible that the infants in the DDT condition did not learn that preferences are diverse, but instead learned something less cognitively powerful like, “In this game I’m playing, people always get the opposite thing, so I should give the other person the thing that I didn’t take.” If this is the case, then the participants did not learn that preferences are specific to the individual; they play a game of opposites. We ran a second experiment to tease apart these explanations.

Experiment 2

Experiment 2 explored the alternative interpretation that infants in the DDT condition of Experiment 1 only learned to give the experimenter what they did not like. Infants completed the same training as in the DDT condition of Experiment 1 but with a “matched” trial on post-training test 2. In a matched trial type, the experimenter demonstrates the same preference as the infant, instead of demonstrating opposite preferences. In this case, if infants in Experiment 1 DDT condition learned that preferences are specific to the individual, and that is why they tend to share the correct food with the experimenter on post-training test 2, then they should give the experimenter the food she likes even though this is also the food that the infant herself likes. Conversely, if infants in the DDT condition of Experiment 1 learned through the course of the session that people should simply always be given opposite things to their partner, then they will give the experimenter the food that they themselves do not like on post-training.
Experiment 2: Methods

Participants
Participants were 29 infants and, as in Experiment 1, only children who failed to give the correct food on the initial post-test continued to training with 20 infants tested in the full training procedure (mean age = 15.5 months; Range = 14.4–17.0 months). An additional 10 infants were tested but not included in data analyses due to failing to complete the study because of fussiness (1), parental interference (1) or refusing to share anything with the experimenters on all test trials (8).

Materials
Food. The food was the same as in Experiment 1 except that the wheel-shaped crackers were replaced with Animal Crackers. This was done because we could no longer find the wheel-shaped crackers.
Toys. The sets of toys were 4 hippos and 4 trucks, and 4 cars and 4 planes. Again, all of the toys within an individual type were slightly different in shape and/or color.

Procedure and Design
The experimental procedure, counterbalancing and randomization were identical to that in Experiment 1.

Predictions
We predicted that infants would perform at chance on post-training test 1, as they did in Experiment 1. If infants give the experimenter the correct food on post-test 2 (the food that both the teacher and the infant prefer), then we will suggest that infants in Experiment 1 did not simply learn to play a game of opposites but instead learned that preferences are diverse.

Experiment 2: Results

Again we replicated the findings from Repacholi & Gopnik (1997). Again, 9/30 infants passed the pre-test (p = 0.06, binomial, marginally significantly fewer than chance), 18 infants shared the incorrect food and 2 infants shared nothing.

Six out of 20 infants were correct on post-training test 1 and 13 out of 20 were correct on post-training test 2, both not significantly different from chance (p = .12 and p = .26, respectively).

The critical comparison is between infants’ performance on post-training test 2 in the Experiment 1 DDT condition and in Experiment 2. This comparison addresses whether infants in Experiment 1 simply learned to play a game of opposites and would have shared the opposite food type to their own preference regardless of what the experimenter demonstrated on post-test 2. For this analysis, we coded infants’ performance in terms of whether they gave the experimenter the opposite food to what the infant preferred (which is correct in Exp 1 DDT but incorrect in Exp 2). We gave infants a score of 1 for sharing the opposite food and a score of zero for sharing the same (non-opposite) food. This test resulted in a score of 7/20 for Experiment 1 and 15/20 on post-training test 2 in the DDT condition of Experiment 1. Using a Fisher’s Exact test, we found that this performance on these trials were significantly different from one another, \( X^2 (1, N = 40) = 6.46, p < .01, \) suggesting that infants in Experiment 1 were more likely to share the opposite food than infants in Experiment 2, where they would have been incorrect in doing so.

Experiment 2: Discussion

Overall, most infants gave the experimenter the food that they preferred (and that the infant also preferred) on post-training test 2 (this was not significantly different from chance using a binomial test). Though we would have expected infants to share the correct food at higher than chance levels in this “matched” trial, we suspect that the non-significant result is due to a lack of statistical power caused by having relatively few participants for binomial statistics. In general, the percentage of infants offering the correct, “matched” food on this trial is very similar to the percentage of younger infants who did so in Repacholi & Gopnik (1997) (65% vs. 72%, respectively). The purpose of Experiment 2 was to eliminate the possible explanation that participants in the Experiment 1 DDT condition only learned to give the experimenter the opposite food of what they themselves wanted. Comparison of Experiments 1 and 2 suggest that this was not the case, as infants shared the food that they preferred in Experiment 2 and did not reflexively give the experimenter the opposite food following training.

General Discussion

Together, these findings show that infants younger than 18 months can learn about the subjectivity of preferences when provided with divergent preferences during training, infants were more likely to remember the general grammatical pattern of that language 24 hours following training test 1, or training trials that involve oppositions of what they themselves desire. This comparison address the opposite of what they themselves desire. This was not the case, as they did in Experiment 1.

An interesting finding in these experiments is that the participants performed identically during the pre-test and post-training test 1, but performed significantly above chance on post-training test 2 in Experiment 1. Both tests occurred after training and we had not predicted this pattern of results, so now we return to the question of why we only saw improvement on post-training test 2.

One possible explanation for this improved performance on post-training test 2 is that post-training test 1 might act as another piece of evidence to train the infants to better understand diverse preferences. That is, post-training test 1 gives infants yet another trial in which the experimenter demonstrates that she likes the opposite food to the infant. It is possible that this extra trial is what allows the infants to learn that preferences are subjective. This possibility can be examined by manipulating the number of training trials, including an additional trial before post-training test 1 on Day 1. Related to this, we can also examine what type of evidence is most informative – evidence that involves first-person experience such as training trial 2 and post-training test 1, or training trials that involve observing others display diverse preferences. By manipulating the number and type of training trials across various conditions in future experiments, we can answer these questions.

Another possible explanation for the improved performance only on Day 2 is the role of memory consolidation in sleep. Post-training test 2 occurs the following day, whereas post-training test 1 occurs on the same day as the training trials. Therefore, a potentially critical difference between the two tests is sleep. Research has shown that sleep is important for the consolidation of memories, and improvements in children’s and infants’ learning are related to longer and more intense sleep (Wilhelm, Prehn-Kristensen & Born, 2012). For example, Hupbach, Gomez, Buzoianu, and Nadel (2009) found that when 5-month-old infants were exposed to an artificial language, they were more likely to remember the general grammatical pattern of that language 24 hours later, compared to infants who did not nap. It is possible that the infants in our experiment performed better on post-training test 2 because they had slept. To address the sleep hypothesis, one could conduct an experiment similar to these here, except with the entire procedure occurring on the same day. After infants complete post-training test 1, half of the infants would take a nap and half would experience a similar delay without taking a nap. Following that infants would complete post-training test 2. If the infants who napped performed better than those who did not, then this would suggest that sleep consolidation is a crucial aspect of their improved performance.

Conclusion

Research on children’s desire-based reasoning has persisted for decades. Here we examined a prediction from a particular model of how children attribute preferences to others, namely that appropriate training regarding the diversity of desires could result in infants undergoing a significant shift in conceptual development (Lucas et al., 2014). We found that following exposure to different people demonstrating divergent desires, infants were able to move from a model of universal preferences to a model that allows for the individualization of preferences. The success of this training procedure more broadly suggests that early advances in Theory of Mind could be due to experience.

Acknowledgments

We thank Justine Hoch and Sophie Bridgers for an immense amount of help with data collection and infant recruitment. We also thank Elizabeth Attianio, Emily McNish, Meghan McGrath, Julia Heinis, Christien Calabeda, and Gina Mandracchia for assistance with data collection. Thanks also to the parents and infants for their participation.

References