Learning to reason about desires: An infant training study

Citation for published 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 Doan1,2 (tdoan@uwaterloo.ca), Stephanie Denison1 (stephanie.denison@uwaterloo.ca), Christopher G. Lucas3 (c.lucas@ed.ac.uk), & Alison Gopnik1 (agopnik@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 between 14 and 18 months of age, allowing 18- to 14-month-olds to understand that different people can have different preferences (Lucas et al., 2014; Ma & Xu 2011; Rapachol & Gopnik, 1997). The present research investigates the kind of evidence that is relevant for inducing this shift and whether younger infants can be trained to learn about the diversity of preferences. In Experiment 1, infants younger than 15 months of age were shown demonstrations in which two experimenters either liked the same objects as each other (in one training condition) or different objects (in another training condition). Following training, all infants were asked to share one of two foods with one of the experimenters – they could either share a food that the experimenter showed disgust towards (and the infants themselves liked) or a food that the experimenter showed happiness towards (and the infants themselves did not like). We found that infants who observed two different experimenters liking different objects during training later provided the experimenter with the food she liked, even if it was something they disliked themselves. However, while both experimenters liked the same objects, they later incorrectly shared the food that they themselves preferred. Experiment 2 controlled for an alternative interpretation of these findings. Our results suggest that training allows infant to overturn an initial theory in the domain of Theory of Mind for a more appropriate theory.

Keywords: Theory of mind; Desire-based reasoning; Infant 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 actions 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 emerges 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 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 study that asked whether infants understand that preferences can serve as an underlying cause of people’s behaviors (Rapacholi & 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, the 14-month-olds gave the experimenter the food that they themselves preferred, regardless of her demonstrated preferences. This difference in performance has been 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 view of preferences in which they initially assume that preferences are universal, rather than varying between people. In contrast, older infants seem to recognize that desires are private.

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 encounter. For example, 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 different 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 inconsistent 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 suitable account of data from a wide range of experiments on children’s understanding of preferences. It accounts for preschoolers’ ability to consider 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 (Fernández et al., 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 providing a suitable evidence of diverse desires through lab-based training, then they might be able to transition to the more complex model of preference attribution. We test 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 Rapacholi & Gopnik’s Goldfish/Broccoli task. All infants were tested in the critical unmatched trial type, wherein the experimenter’s preference conflicted with the infant’s preference (henceforth, ODIT) 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 the percentage of infants completing training did not already know that preferences are diverse. Twenty infants per condition were tested in the full experiment with 15 months of age (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 19 infants were excluded from the final data analysis due to failing to complete the study because of fussiness (2) or refusing to share on the pre-test and all test trials (17). 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 Cherries, 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 were toys were 4 cars 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
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 to pass it back by placing it in the experimenters’ hands.

Pre-test: We based on Repacholi & Gopnik (1997). Experimenter 1 slid a plate of food consisting of a few pieces of vegetables and snacks (e.g., raw broccoli and Goldfish crackers) towards the infant and encouraged the infant to try some. The experimenter gave the infant a 45 second time frame to taste the foods and the experimenter determined which of the two foods the infant preferred. We used the same coding as in Repacholi & Gopnik (1997) to determine food preferences on all trials (pre- and post-tests). Inter-coder agreement for preferences was 91%. When the infant’s preference was determined, the experimenter took out a container consisting of the same foods the infant had tried. The experimenter then demonstrated that she liked the food that the infant did not show a preference for and was disgusted by the food that the infant preferred. The experimenter showed her preferences by saying, e.g., “Eww! Crackers! I tasted the crackers! Eww!”, and “Mmm! Broccoli! I tasted the broccoli! Mmm!”. The experimenter showed a liking and disliking towards each food three times and she did this using facial expressions based on the descriptions of Ekmän & Friesen (1975). Next, the experimenter placed broccoli on one side of a tray and Goldfish crackers on the other, placed her hand with her palm face up towards the infant, said, “can you give me some?” and slid the tray towards the infant. The infant was given 45s to pass the food to the experimenter. If the infant gave the experimenter the food that the experimenter showed a preference towards, then the infant passed the pre-test. If the infant gave the experimenter the food that she disliked, or did not provide the experimenter with any food, then the infant failed the pre-test.

Training Trials. Infants who failed the pre-test were introduced to either the DDT condition or the N-DDT condition. Infants in the DDT condition saw two experimenters liking and disliking different toys and infants in the N-DDT condition saw two experimenters disliking the same toys.

Training proceeded as follows: Training trial 1 occurred right after the pre-test. During training trial 1, Experimenter 1 put a toy (e.g., books or toys) onto the table and subsequently pulled out three toys of one type (e.g., dogs) and showed these to the child. Then, the experimenter pulled out three toys of the other type (e.g., trucks) and expressed dislike towards them. The dialogue and facial expressions used were similar to that used during the pre-test. The experimenter expressed her preferences by saying, “Yay! A dog! I got a dog! Yay!”, and “Eww! A truck! I picked up a truck! Eww!”. Once Experimenter 1 expressed her emotions for each type of toy three times, Experimenter 2 took over. Experimenter 2 showed liking and disliking towards the same toys as Experimenter 1 if the infant was in the N-DDT condition (e.g., liked dogs and disliked trucks) and she showed liking and disliking towards the opposite toys as Experimenter 1 if the infant was in the DDT condition (e.g., liked trucks and disliked dogs).

Training trial 2 involved Experimenter 2 and the infant. It was similar to the pre-test, except that it involved a different type of toy (e.g., a set of food, e.g., broccoli and chicken). After each food item appeared on the tray during the sharing part of the pre-test and post- training tests were randomized. For training trials, if Experimenter 1 liked and disliked things, Experimenter 2 would be the first to play with the toy, she would continue to like animals and dislike vehicles in subsequent training trials. For half the participants, Experimenter 1 liked animals and for the other half, Experimenter 1 liked vehicles. This was crossed with half of the infants seeing dislikes expressed first and half seeing likes expressed first. None of these counterbalancing factors led to any systematic differences in the data (when entered into ANOVAs, all p’s > .05 for these factors).

Predictions. We predicted that infants in the DDT condition would be more likely to offer the experimenter the correct food on the post- than the pre-test if infants in the N-DDT condition even though both conditions provided infants with practice in considering other people’s preferences and desires. In the N-DDT condition, infants saw two experimenters liking the same objects – this does not provide the infants with any information that allows them to learn that different people can have different mental states. In the DDT condition, infants saw two experimenters display different preferences from each other, which would provide a great deal of evidence to suggest that different people can have different preferences.

Experiment 1: Results

Of the initial 55 infants who participated in the experiment, 15 passed the pre-test by giving the correct food (p < .01, binomial, significantly fewer than chance), 34 infants shared the incorrect food, and 6 infants shared nothing, replicating that infants this age perform below chance on this task (Repacholi & Gopnik, 1997). This confirms that, in general, infants below 18 months are inclined to share the item that they themselves prefers, not the item for which another person has shown a preference. For the 40 infants that failed the pre-test and continued to training, performance on post-training test 1 was significantly different to that of infants in the General Discussion.

Experiment 2. Experiment 2 explored the alternative interpretation that infants in the DDT condition of Experiment 1 only learned to give the experimenter from which they liked. 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, they should play a game of opposites and run a second test to tease apart these possibilities.

Experiment 2

For the first 10 infants in both training conditions, the food on post-training test 2 was identical to the food on training trial 2 (which the infant used with Experimenter 2 on Day 1 but did not share). We switched this to a new food type to ensure that any improvement in infants’ performance on Day 2 in DDT could not be explained by already being familiar with these foods.
test 2, even though the experimenter demonstrates that she likes the food that the infant also prefers. We maintained the exact same procedure as in the DDT condition of Experiment 1, including using an “unmatched” trial type for post-training test 1, as the effect was observed only in post-training test 2 and so every aspect of the experimental session must remain the same until that point.

**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 pre-test continued to training with 20 infants tested in the full training procedure (mean age = 15.5 months; Range = 14.4 months to 17.0 months). An additional 10 infants were tested but not included in 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 cats 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 Experiment 1 DDT. **Predictions** We predicted that infants would perform at chance on post-training test 1, as they did in Experiment 1. If infants gave the experimenter the correct food on pre-test 2 (the food that both the experimenter and the infant like), then this will suggest that infants in Experiment 1 did not simply learn to play a game of opposites but instead learned that preferences are diverse. 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 although infants younger than 18 months can learn about the subjectivity of preferences when provided with appropriate training trials (i.e., infants were exposed to any training, they provided an adult with the food that they personally liked and not the one the experimenter liked), marginally significantly fewer than chance, 18 infants shared the incorrect food and 2 infants shared nothing at all. 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 significantly different from chance (p = .06, binomial, marginally significantly fewer than chance), 18 infants shared the incorrect food and 2 infants shared nothing at all.

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 resulted in a score of 72/20 for post-training test 2 in Experiment 2 and 15/20 on post-training test 2 in the DDT condition of Experiment 1. Using a Fisher’s Exact test, we found that preferences on these trials were significantly different from one another, X²(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 she 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).

One concern regarding these data is the relatively low statistical power that results from our experimental design and the small sample size for each experiment. Although the results in the Experiment 1 DDT condition were significant, it will be prudent to replicate these findings. This replication experiment is currently underway in our lab.

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 we now 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, to include 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 – in particular, that involves first-person experience such as training trial 2 and post-training test 1, or training trials that involve observing two actors display divergent 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 is the two nights of sleep. Research has shown that sleep is important for the consolidation of memories, and improvements in children’s and infants’ learning is enhanced with longer and more intense sleep (Wilhelm, Prehn-Kristensen & Born, 2012). For example, Hupbach, Gomez, Bustin, and Nadel (2009) found that when 5-5 month old infants napped after they 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, all 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 Attisano, Emily McNichol, Meghan McGrath, Julia Heunsin, Christien Caluza, and Gina Mandrachi for assistance with data collection. Thanks also to the parents and infants for their participation.

**References**
