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 Denison2 (stephanie.denison@uwaterloo.ca), Christopher G. Lucas1 (c.lucas@ed.ac.uk), & Alison Gopnik3 (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 between 14 and 18 months of age, allowing 18- to not 14-month-olds to understand that different people can have different preferences. 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 would 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 they liked, even if it was something they disliked themselves. However, when infants observed two experimenters liking the same objects, they later incorrectly shared the food that they themselves liked. Overall, Experiment 2 controlled for an alternative interpretation of these findings. Our results suggest that training allows infants to overturn an initial theory in the domain of Theory of Mind for a more advanced one.

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 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 undergoes 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 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 & Woolley, 1990).

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 (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 – 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, only 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 notion of desires and preferences (Republichi & Gopnik, 1997). 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 that 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 useful account of data from a wide range of experiments on children’s understanding of preferences. It accounts for preschoolers’ ability to correctly attribute preferences from the statistical properties of a collection of objects and an agent’s choices (Kushnir, Xiu, & 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 you wanted to train children to consider the 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 Repacholi & Gopnik’s Goldfish/Broccoli task. All infants were tested in the critical unmatched trial type, wherein the experimenter’s preference conflicted with the infants’ preference (Repacholi & Gopnik, 1997). 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). We predicted that infants in the DDT condition should show improved performance in attributing preferences during 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 likelihood that infants completing training did not already know that preferences are diverse. Twenty infants per condition were tested in the full procedure (DDT: mean age = 15.7 months; 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 same area, but 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 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 divided into two groups of toys: goldfish and broccoli (D1), four airplanes and four monkeys (D2). The infants were randomly assigned to each set.

Procedure, Design and Predictions

All infants participated in a training session. They sat in a chair in front of a table and their parent sat in a chair beside them. Before the training began, two experimenters played a passing game with the infant. This allowed the infant to warm up to the experimenters and to

While this is an important distinction we will not discuss it further, because both processes result in identical behavior in our task.
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 (Experiment 1) 
Infants entered a same objects condition (e.g., liked dogs and disliked trucks) and a different objects condition. We ran a second experiment to tease apart these explanations.

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 liking and 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 jar of toys (e.g., dogs and trucks) onto the table and then the experimenter pulled out three toys of the same type (e.g., dogs) and expressed liking for them. Then, the experimenter pulled out three toys of the other type (e.g., trucks) and expressed dislike towards them. The dialogue and facial expressions were the same during the pre-test. The experimenter expressed her preferences by saying, “Yay! A dog! I got a dog! Yay!” and “Ewww! A truck! I picked up a truck! Ewww!” 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 toy set (e.g., a set of food, e.g., raw broccoli and Goldfish crackers). The experimenter then demonstrated which of the two foods the infant preferred. We used the same coding as in Rapcholi & Gopnik (1997) to determine food preferences on all trials (pre- and post-tests). Inter-coder agreement for preferences was 91.5%.

When the infant's preference was determined, the experimenter took 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., “Ewww! Crackers! I tasted the crackers!” Ewww!” 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 Ekman & 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 up towards the infant, said, “can you give me some?” and slid the tray broccoli. The infant was given 45s to pass 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 trial 2 occurred three times, Experimenter 1 demonstrated that she liked the food and 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., “Ewww! Crackers! I tasted the crackers!” Ewww!” 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 Ekman & 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 up towards the infant, said, “can you give me some?” and slid the tray broccoli. The infant was given 45s to pass 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 trial 1 was identical to training trial 1, but with a different set of toys (e.g., monkeys and planes). Experimenter 1 expressed liking to one type of toy and dislike towards the opposite type of toy. Experimenter 2 had a turn expressing her emotions towards each of the toys. She expressed happiness and dislike towards the same toys as Experimenter 1 if the infant was in the N-DDT condition. After Experimenter 2 finished her demonstrations, infants completed training task 1. Experimenter 2 put one of each type of toy on both sides of a tray (e.g., a monkey on right, a plane on left), placed her palms face up towards the infant, pushed the tray towards the infant and asked the infant to share one with her. The infants were given 45s to share a toy with the experimenter. Once the infant shared a toy with Experimenter 2, Experimenter 1 had a chance to ask the infant to share with her the toy she liked. Training trial 4 was a repetition of training trial 3 and included a training task that was identical to the one completed after training trial 3.

The purpose of the training tasks, where infants were asked to share one of two toys with each experimenter, was simply to ensure that the infants did not get bored and continued to share throughout the study. We did not expect that infants would share the same toy across both training sessions. A second possibility is that a model of how preferences work (i.e., to learn that they appear to the individual). A second possibility is that a model of how preferences work (i.e., to learn that they appear to the individual). A second possibility is that a model of how preferences work (i.e., to learn that they appear to the individual). A second possibility is that a model of how preferences work (i.e., to learn that they appear to the individual).

Infants again warmed up with Experimenter 1 by playing the warm-up game from Day 1. This was followed by post-training test 1, which was identical to the pre-test and post-training test 1, but again with a different set of food (e.g., green peppers and wheel-shaped crackers).

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.

Experiment 1: Discussion
Our results suggest that the type of information provided during training was crucial to infants learning about diverse desires. When infants were provided with a large number of instances indicating that two different people can like different things, they appeared to share the belief that they disliked but the experimenter preferred. However, infants’ performance did not improve when they saw preferences that were not diverse: infants in the N-DDT condition did not share the correct food with the experimenter on any post-training tests. This suggests that training with appropriate evidence can result in significant changes to children’s explicit Theories of Mind.

But did infants in the DDT condition only demonstrate advances in understanding on Day 2 of the experiment, during the second post-training test? We see at least two possible explanations. One possibility is that post-training test 1 served as a final training trial, giving infants the minimum number of examples required to change their model of how preferences work (i.e., to learn that they apply to the individual). A second possibility is that a night of sleep resulted in improved learning of this general 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 conceptually powerful like, “In this game I’m playing, people always get opposite things they should give to 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 and 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 from what they liked. Infants completed the same training as in the DDT condition of Experiment 1 but with a “matched” trial on post-training test 1. 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 tended to share the correct food with the experimenter on post-training test 2, then they should learn the same food 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
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-test 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: 14.4-17.0 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 hippo 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, randomization and training in Experiment 1 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 post-test 2 (the food that both the experimenter and the infant like), then this would 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 (1993). Of 93 infants tested the pre-test (p = .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 significantly different from chance (p = .02 and p = .06, 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 coding resulted in a score of 72/20 for infants 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 infants’ performance on these trials was 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 they preferred (and 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 able to reason correctly about another person’s preferences, providing the experimenter with the food that she liked. In contrast, the infants who only saw congruent expressions of liking and disliking options did not learn to reason correctly about another person’s preferences, and continued to give the experimenter the food that they themselves preferred, regardless of the experimenter’s preference.

Experiment 2 helped to clarify these findings, providing evidence that infants did not simply learn to always give the experimenter the food that the infant liked, but that they learned to give the opposite of what the infant itself did. Post-training test 2 of Experiment 2 was a “matched” trial, meaning that the experimenter showed the infant that she liked the same food as the infant. Because the majority of infants gave the experimenter the food that the infant (and the infant) liked, we can be confident that infants in Experiment 1 learned that preferences are diverse. Taken together, the coding resulted in a score of 72/20 for infants 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 infants’ performance on these trials was 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.

**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 Bridge for an immense amount of help with data collection and infant recruitment. We also thank Elizabeth Attias, Emily McNichols, Meghan McGrath, Julia Heinis, Christen Calabas, and Gina Mandracchia for assistance with data collection. Thanks also to the parents and infants for their participation.

**References**
