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, Stephanie Denison2, Christopher G. Lucas3, Alix Goebel4, and Alison Gopnik5

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 cross-cultural development suggests that a significant conceptual shift occurs in desire-based reasoning between 14 and 18 months of age, allowing 18- to 24-month-olds to understand that different people can have different preferences (Lucas et al., 2014; Ma & Xu 2011; Repacholi & 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, we also found that younger infants sometimes liked the same objects, they later incorrectly share the food that they themselves liked with the experimenter, as they would see it as the object she liked, even if it was something they disliked themselves. While this is an important distinction we will not discuss it further, our results suggest that training allows infant to overturn an initial theory in the domain of Theory of Mind for a more sophisticated understanding of preferences.

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 or 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 in infancy and early childhood, although 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 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 another person who preferred the opposite food, they selected one of the objects that the experimenter had demonstrated a preference for. In 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 sophisticated understanding of preferences.

Keywords: Theory of mind; Desire-based reasoning; Infant learning; Social cognition; Preferences.

Procedure, Design and Predictions

Toxins

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 of toys were small stuffed animals; and 4 and 8 dogs, and small and large monkeys. The toys within each type were not identical; they varied in color and shape.

Procedure, Design and Predictions

While this is an important distinction we will not discuss it further, because both processes result in identical behavior in our task.

contrasting, 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, Doan 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 towards 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 from 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 obtain experiences 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 (Fargas et al., 2009; 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 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 (1997) Goldfish/broccoli task. All infants were tested in the critical unmatched trial type, wherein the experimenter’s preference conflicted with the infant’s prior training (Lucas et al., 2014). They were then asked if they observed that the experimenter was “thinking” (henceforth, DTT) 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-DTT), 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 DTT 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. Infants completing training did not already know that preferences are diverse. Twenty infants per condition were tested in the full training protocol (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 first condition due to failing to complete the study because of fussiness (2) or refusal to share on the pre-test and test trials (13).
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.

**Procedure**

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 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 of the toys 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 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 towards the infant, pushed the tray towards the infant and asked the infant to share one with her. The infants were 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 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 toy (e.g., dogs and trucks) onto the table and subsequently pulled out three toy of one type (e.g., dogs) and expressed dislike towards 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 between Experimenter 1 and the infant 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 toy (e.g., left side of the tray and a right side of the tray and fruits) and each food item was part of the pre-test. Experimenter 2 gave the infant a plate of food and determined which food the infant preferred within 45s. In the DDT condition, the experimenter then demonstrated that she preferred the food that the infant disliked and disliked the food that the infant preferred. In the N-DDT condition, the experimenter demonstrated that she liked and disliked the same foods as the infant. The infant was not asked to share any food with the experimenter, as this was a training trial and not a test.

Training trial 3 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 other 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 and expressed happiness and dislike towards the opposite toys as Experimenter 1 if the infant was in the DDT condition. After Experimenter 2 finished her demonstrations, infants completed training task 1. Experimenter 1 put two 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 that he/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 remember which toy and in fact we found that infants did not reliably remember the experimenters’ preferences in either condition of Experiment 1 or in Experiment 2 (all $p’s > .25$ for ANOVA’s examining infants’ passing behavior on the experimenters’ preferences). Post-training test 1 immediately followed training. It was identical to the pre-test, except with different food (e.g., cucumbers and Cheerios). Once the infant shared a food on post-test 1, the first day of the study was complete.

Infants returned on Day 2 to complete post-training test 2. Infants again warmed up with Experimenter 1 by playing the warm-up game from Day 1. This was followed by post-training test 2, 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-test 2 was identical to the food on the 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 in determining about whose 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 Theory of Mind.

But it did 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, “In this game I’m playing, people always get opposite things—they 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 simply 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 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 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 tended to share the correct food with the experimenter on post-training test 2, then they should show the correct 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 this 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 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 and randomization were identical to Experiment 1.

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 toddler 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. We also thank Elizabeth Attisano, Emily Mcintosh, Meghan McGrath, Julia Heunis, Christen Caubada, and Gina Mandracchia for assistance with data collection. Thanks also to the parents and infants for their participation.

References


