Learning to reason about desires: An infant training study

Citation for published version:

Published In:
Learning to reason about desires: An infant training study

Tiffany Doan1(1tiffanyuwatervoda.ca), Stephanie Denison2(stephanie.denison@uwatervoda.ca), Christopher G. Lucas4(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 younger than 14-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 18 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 they liked, even if it was something they disliked themselves. However, when both experimenters showed the same object, the infants later incorrectly shared the food that they did not like. 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 complex model.

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 variety of functions, from understanding other people’s motivations 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 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, 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 demonstrated which food the infants liked (the majority preferred Goldfish crackers). The infant 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 notion 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 diverse.

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 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 unified account of preferences because it gives a parsimonious explanation for diverse. Twenty infants per condition were tested in the full training group (DDT), and Range = 14.4 months to 17.2 months). An additional 15 infants were tested from a control group 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 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

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. 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., “Ewww! Crackers! I tasted the crackers! Ewww!” and “Mmmm! Broccoli! I tasted the broccoli! Mmmm!”. 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 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 share a toy with the experimenter. Once the infant shared a toy with Experimenter 2, Experimenter 2 had a chance to ask the infant to share with her the toy that she liked.

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 as Experiment 1 if the infant was in the DDT condition (e.g., liked dogs 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 set of food (e.g., a toy or food item) and each food item was a 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. The experimenter showed a happiness and dislike towards the same toys as Experiment 1 if the infant was in the DDT condition. After Experimenter 2 finished her demonstrations, infants completed training task 1. Experimenter 1 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 toys were given 45s to share a toy with the experimenter. Once the infant shared a toy with Experimenter 2, Experimenter 2 had a chance to ask the infant to share with her the toy that 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 they liked 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 based 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 a 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).

Design. The foods used on each trial were randomized (some infants saw broccoli and Goldfish on the pre-test, some on training trial 2, some on post-training test 1 and some on post-training test 2). In Experimenter 1, some appeared on the sharing part of the pre-test and post-training tests was randomized. For training trials, if Experimentals liked different toys, Experimenter 1 put one toy on both sides of a tray (e.g., a truck on right, a dog 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 toys were given 45s to share a toy with the experimenter. Once the infant shared a toy with Experimenter 1, Experimenter 1 had a chance to ask the infant to share with her the toy that 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 purposes 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 they liked 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 based 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 a 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-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 information 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.

If 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 post-training test. 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. I should give whatever 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 opposite desires. They 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 disliked. 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 choose the food that the infant they liked 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 experimenters 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-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 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 will suggest that infants in Experiment 1 did not simply learn to play a game of opposites but instead learned that preferences are diverse. By manipulating the number of trials, we developed a replication of Experiments 1 and 2 and suggested that this was not the case, as infants shared the food that they preferred in Experiment 2 and did not reflectively 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 diverse experiences during training. For instance, if infants were exposed to any training, they provided an adult with the food that they personally liked and not one the experimenter liked, even though both foods are nutritionally equivalent. Thus, it is possible that infants younger than 18 months may interpret the experimenter’s feedback as more than simply preference. To further test this hypothesis, one could conduct an experiment similar to those 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. Follow that infants who did not complete post-training test 2. If the infants who napped perform better than those who do 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 Bridges for an immense amount of help with data collection and infant recruitment. We also thank Elizabith Attisano, Emily McNichol, Meghan McGrath, Julia Henius, Christen Calabas, and Gina Mandracchia for assistance with data collection. Thanks also to the parents and infants for their participation.

**References**


