COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2017) 1–31

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12480

Learning the Structure of Social Influence

Samuel J. Gershman,[a] Hillard Thomas Pouncy,[a] Hyowon Gweon[b]

[a]*Department of Psychology and Center for Brain Science, Harvard University*
[b]*Department of Psychology, Stanford University*

Received 23 December 2015; received in revised form 27 November 2016; accepted 13 December 2016

Abstract

We routinely observe others' choices and use them to guide our own. Whose choices influence us more, and why? Prior work has focused on the effect of perceived similarity between two individuals (self and others), such as the degree of overlap in past choices or explicitly recognizable group affiliations. In the real world, however, any dyadic relationship is part of a more complex social structure involving multiple social groups that are not directly observable. Here we suggest that human learners go beyond dyadic similarities in choice behaviors or explicit group memberships; they infer the structure of social influence by grouping individuals (including themselves) based on choices, and they use these groups to decide whose choices to follow. We propose a computational model that formalizes this idea, and we test the model predictions in a series of behavioral experiments. In Experiment 1, we reproduce a well-established finding that people's choices are more likely to be influenced by someone whose past choices are more similar to their own past choices, as predicted by our model as well as dyadic similarity models. In Experiments 2–5, we test a set of unique predictions of our model by looking at cases where the degree of choice overlap between individuals is equated, but their choices indicate a latent group structure. We then apply our model to prior empirical results on infants' understanding of others' preferences, presenting an alternative account of developmental changes. Finally, we discuss how our model relates to classical findings in the social influence literature and the theoretical implications of our model. Taken together, our findings demonstrate that structure learning is a powerful framework for explaining the influence of social information on decision making in a variety of contexts.

Keywords: Bayesian inference; Social cognition; Structure learning

1. Introduction

Imagine you are trying to buy a bar of soap, and you don't see the brand that you normally use. Confronted with a dozen different kinds of soaps you haven't used before, you

Correspondence should be send to Samuel Gershman, Department of Psychology, Harvard University, 52 Oxford St., Room 295.05, Cambridge, MA 02138. E-mail: gershman@fas.harvard.edu

might be at a loss. However, if you see another customer choosing a particular brand of soap, you might be tempted to make the same choice. Moreover, if her basket is full of other items that you like (e.g., the same brand of toothpaste and shampoo you use), you might be even more convinced that you will also like her chosen soap.

From these mundane, everyday purchases to important life decisions, our choices are strongly influenced by the choices of others. Indeed, humans are not the only species that learn from others; extensive prior literature documents various social influences on behaviors of non-human animals. For instance, rats learn to avoid poison from observing other rats (Steiniger, 1950), and lab-raised rhesus monkeys develop fear responses to snakes after seeing wild-raised monkeys' responses to snakes (Cook & Mineka, 1989). Observations of others' behaviors also influence many aspects of foraging decisions in non-human animals, such as what, where, and when to eat (Galef & Giraldeau, 2001).

Social learning in humans goes far beyond simple imitation. Recent empirical and theoretical work suggests even infants have sophisticated inferential mechanisms that allows them to go beyond simple associations (Schulz, 2012; Tenenbaum, Kemp, & Goodman, 2011), as well as the ability to understand mental states such as goals, preferences, and beliefs (Repacholi & Gopnik, 1997; Wellman & Liu, 2004; Woodward, 1998). From observations of others' actions, infants draw rich causal inferences about the goals and intentions behind these actions and use them to guide their own behavior (Gergely, Bekkering, & Király, 2002; Gweon & Schulz, 2011; Meltzoff, 1995). From just a few instances of an another person's choice behaviors (e.g., selections of toys), infants infer whether other toys share similar features (Gweon, Tenenbaum, & Schulz, 2010) and how much the person likes the toys (Hu, Lucas, Griffiths, & Xu, 2015; Kushnir, Xu, & Wellman, 2010). Infants also use the perceived overlap between their own and others' choices to guide their future choices (Fawcett & Markson, 2010).

The flexibility of children's inferences suggests that they may be supported by an abstract understanding of preferences (i.e., utility functions that underlie choices) rather than mere associations (Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016). Around 18 months of age, children already understand that another agent's food preference may differ from their own (Repacholi & Gopnik, 1997) and provide food items that she likes even when it conflicts with their own preferences. Recent computational work has formalized preference understanding as a rational comparison of utility functions between individuals (Bergen, Evans, & Tenenbaum, 2010; Jern & Kemp, 2015) and showed that such models predict inferences about preferences in young children (Lucas et al., 2014). However, these models rely on choice similarity between pairs of individuals (i.e., dyadic choice similarity), and thus they may not fully capture our ability to flexibly learn from multiple agents, even in the absence of strong differences in perceived overlap, such as in situations where group affiliation is informative about preferences.

Prior work on social learning from multiple agents focused on how individual choices are influenced by shared affiliations with social groups. An important assumption that underlies many studies and theories of social influence is that individuals within a social group share a common set of utility functions (i.e., shared preferences), such that copying behaviors of in-group members is likely to lead to more rewarding outcomes (e.g., Bikhchandani, Hirshleifer, & Welch, 1992; Rendell et al., 2010; Weitzman, 1965). Consistent with this general idea, previous developmental research suggests that young children, and even infants, are sensitive to explicit cues that distinguish in-group members from out-group members (e.g., gender, age, language) in making food choices (Frazier, Gelman, Kaciroti, Russell, & Lumeng, 2012; Shutts, Kinzler, McKee, & Spelke, 2009; Shutts, Banaji, & Spelke, 2010; Whittler & DiMeo, 1991).

However, there are reasons to question the effectiveness of such cues in making real-world choices. First, individuals hold richly heterogeneous preferences across many domains; while some of them may be rooted in culturally shared utility functions, others may be highly idiosyncratic and vary widely even among members within a group (e.g., music or movie choices). Furthermore, even when affiliation with a certain group is potentially helpful, observable features for various groups are often simultaneously present and confounded within individuals and across domains (e.g., a male, French-speaking lawyer in his 50s vs. a female, Korean-speaking engineer in her 30s), making it difficult to choose which ones are actually informative. Thus, even though these directly observable features of group affiliations can signal underlying similarities in utility functions, the causal connection between these features and observed choices are often unclear; by contrast, overlaps in choice behaviors provide strong support for similar utility functions within a given domain. This suggests that the most useful groups for guiding one's future choices might be latent and must be learned from observing others' choices. In order to account for the richness and the sophistication of the human ability to learn from others' choices even when explicit cues to group affiliation are lacking or unhelpful, we must go beyond the simple assumption that choice similarities between individuals arise from explicitly recognizable groups. How do human learners face the complex challenge of learning from multiple agents whose heterogeneous choice behaviors emerge from unobservable, latent social groups?

Here we present a formal account of how humans use others' choices to generate representations of latent social structures, and we present a series of experiments that demonstrate how these representations influence people's choice behaviors. We revise the idea that the strength of social influence depends directly on dyadic choice similarity (Brock, 1965; Simons, Berkowitz, & Moyer, 1970). We argue that social influence involves inferences about the latent structure of shared utility functions among multiple individuals. For example, the group "people who enjoy outdoor sports and historical documentary films" is not directly observable; it is an abstraction that enables us to capture certain regularities in observable choice behavior. Such abstractions are not only useful for learning about about others but are also useful for guiding one's own choices; if you infer that someone belongs to the same group as you (e.g., both you and your friend enjoy outdoor sports and historical documentary films), then her choices of other hobbies or films might be a more reliable guide to what you will like compared to the choices of someone who presumably belongs to a different latent group (e.g., people who enjoy knitting and romantic comedies).

To formalize these ideas, we appeal to computational theories of structure learning (Gershman & Niv, 2010), which explain how learners can discover a set of latent variables (e.g., groups) that could plausibly give rise to observable data. In the next section, we introduce a computational model of latent group inference from choice patterns. We then test the model predictions across five experiments using real-world movie posters. In Experiment 1, we reproduce the well-established finding that people's choices are more influenced by agents whose past choices are similar to their own (Brock, 1965; Simons et al., 1970), confirming that our experimental paradigm can induce such social influence on choice behaviors, and that our model can adequately account for this basic effect. In Experiments 2–5, we test the distinctive predictions of our model: If people represent latent group structure rather than dyadic similarities, then even when two agents' choices are equally similar to their own past choices, their decisions should be influenced by the presence of a third agent that alters the group structure. Importantly, prior models that rely on dyadic choice similarity would not predict differential social influence in these cases (since choice overlap is equated for the two agents in question). Our model, in contrast, predicts that individuals will be differentially influenced by the agent who belongs to the same latent group as themselves. Experiments 2 and 3 test this prediction by measuring choice behavior. Experiment 4 provides more direct evidence by asking people to cluster agents into separate groups. Experiment 5 shows that the influence of latent groups extends beyond choices between two novel items; it is powerful enough to change people's ratings for real-world items.

2. A computational model of social structure learning

Our model embodies a "rational analysis" of social influence, taking a probabilistic approach to understanding human cognition and decision making (Anderson, 1990; Chater & Oaksford, 1999; Tenenbaum et al., 2011). We start with a description of the *generative model*: a set of assumptions about how latent groups are formed and how they can be inferred from observed patterns of choices. We then describe how an individual should reason rationally under the assumptions of the generative model.

Before specifying the model formally, we provide some intuitions about how the model works (see Fig. 1 for a schematic). Each individual is assumed to belong to a latent group that parameterizes the distribution over choices. Due to the stochastic nature of choice, different individuals within the same group may not make exactly the same choices. The assignment of individuals to groups is unknown and must be inferred from observed data. The number of groups is also unknown; we assume a distribution over group assignments that accommodates an unbounded number of groups, with a bias for smaller numbers. This simplicity bias prevents the model from overfitting the choice data (e.g., by assigning each individual to a unique group), which would hinder generalization to new choice problems. Similar distributional assumptions have been used to model other learning problems in which the number of groups or "clusters" is unknown (see Austerweil, Gershman, Tenenbaum, & Griffiths, 2015, for a review).

Within this framework, reasoning rationally about preferences means combining prior beliefs about group structure with information from choice patterns to draw inferences

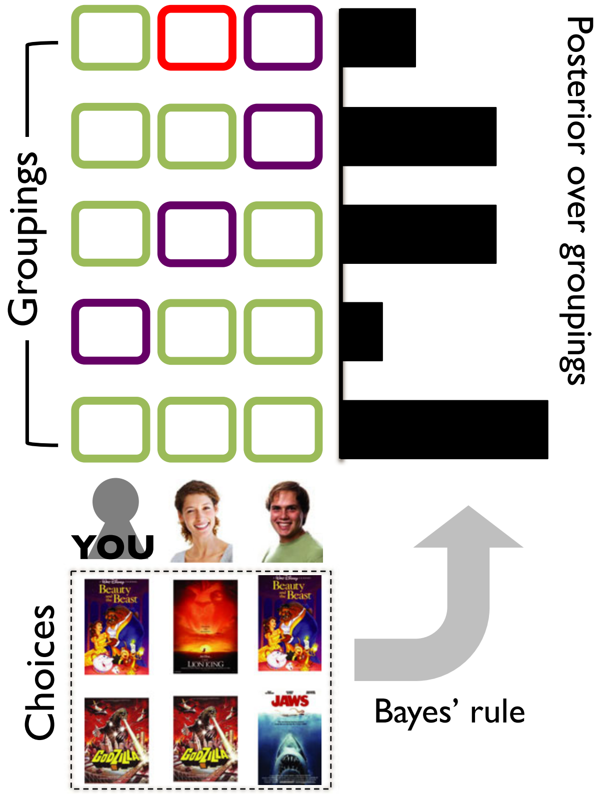

Fig. 1. Model schematic. Bayes' rule is used to compute a posterior distribution over groupings given the choices of several individuals (including the participant, labeled "You"). Each row of colored boxes represents possible groupings (with each color representing a group), and the black bars indicate the posterior probability over these groupings.

about the likely group assignments that generated the choice patterns. As we detail in our model description below, Bayes' rule specifies how prior beliefs and choice patterns are combined to generate a posterior distribution that quantifies the probability over possible groupings of individuals. The posterior can then be used to compute predictions about choice behavior for new sets of objects. The same computation can be used to generate predictions for one's own choices, as well as for those of other individuals.

In our usage, reasoning rationally does *not* mean reasoning "optimally," and predicting others' choices does not mean inferring their "true preferences." In fact, because actual experiences are inherently subjective, it is difficult to even define what counts as optimal; certainly, an agent whose behavior is consistent with our model cannot be said to have made "better" choices (i.e., choices that are actually more enjoyable) than another agent who does not. Rather, our claim is that people systematically combine their prior knowledge with new social observations to make choices in ways that are well described by probabilistic reasoning. In other words, we are not claiming that people rationally infer what they actually like (their *true* preferences); rather, we propose that people can make rational choices about what they *think* they might like (i.e., they form rational expectations of expected utility observing others' choices).

Thus, rather than examining how people infer the structure of others' abstract, stable preferences, we focus on how people infer the latent structure of social groups that give rise to choice behaviors, and how such structure informs people's subsequent choices. As such, we do not attempt to empirically validate whether participants' preference predictions match their actual hedonic experience; while important for other reasons, such a match is orthogonal to our present goals.

2.1. Generative model

In our model, M individuals are presented with a collection of N choice problems, $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_N\}$, where $\mathcal{X}_n$ is a set of objects. Let $c_{mn} \in \mathcal{X}_n$ denote the object chosen by individual m on problem n. The probability of this choice is given by:

$$P(c_{mn} = c | z_m = k, \theta) = \theta^c_{kn}, \quad (1)$$

where z_m denotes the group assignment of individual m, and θ is a set of multinomial parameters (although for generality we present our model for choice sets with arbitrary numbers of items, the experiments we describe later only involve binary choices). We assume that θ_{kn} was drawn from a symmetric Dirichlet distribution with parameter γ. For all of our simulations, we set $\gamma = 1$, which corresponds to a uniform distribution. While we did not directly fit this parameter to our data, we found that changing its value to 0.1 or 2 had a negligible effect on the model predictions (squared error changed by <0.001).

Finally, we assume that group assignments are drawn from a distribution known as the *Chinese restaurant process* (CRP; Aldous, 1985). This distribution has "infinite capacity" in the sense that it can generate an infinite number of groups but will tend to produce a parsimonious grouping (see Gershman & Blei, 2012, for a tutorial introduction). Mathematically, the probability of grouping $\mathbf{z} = [z_1, \ldots, z_M]$ under the CRP is given by:

$$P(\mathbf{z}|\alpha) = \frac{\alpha^K \Gamma(\alpha) \prod_k \Gamma(T_k)}{\Gamma(M + \alpha)}, \quad (2)$$

where $\alpha \geq 0$ is the *concentration parameter*, T_k is the number of individuals assigned to group k, and $\Gamma(\cdot)$ is the gamma function. A random draw from the CRP will on average generate $\alpha \ln M$ groups. Thus, the number of groups will generally grow as more individuals are observed. The concentration parameter controls the tendency to create new groups: When $\alpha = 0$, only a single group will be created, whereas in the limit $\alpha \to \infty$, every individual will be assigned to a unique group.

To allow for uncertainty about α, we placed a prior over α and numerically marginalized over all possible values:

$$P(\mathbf{z}) = \int_{\alpha} P(\mathbf{z}|\alpha)P(\alpha)d\alpha. \tag{3}$$

In the analyses reported below, we used a prior that was uniform between 0.00001 and 10. This range was chosen to cover the entire high-density region of the parameter space. The non-zero lower bound was chosen because the Gamma function is infinite at $\alpha = 0$.

2.2. *Rational reasoning about latent social groups*

Let us suppose that individual m observes choices $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_M]$, including her own, where $\mathbf{c}_m = [c_{m1}, \ldots, c_{mN}]$. This set of choices also includes those from a final test problem $\mathcal{X}_*$, where individual m first observes the choices of other individuals. The task facing individual m is to determine the probability that she chooses option c on the test problem:

$$P(c_{m*} = c|\mathbf{C}) = \sum_{\mathbf{z}} P(c_{m*} = c|\mathbf{z}, \mathbf{C})P(\mathbf{z}|\mathbf{C}), \tag{4}$$

where

$$P(c_{m*} = c|\mathbf{z}, \mathbf{C}) = \frac{\gamma + L^{c}_{z_m *}}{|\mathcal{X}_*|\gamma + \sum_{c' \in \mathcal{X}_*} L^{c'}_{z_m *}}. \tag{5}$$

Here $|\mathcal{X}_n|$ is the number of options on problem n and L^c_{kn} is the number of individuals assigned to group k who chose object c on problem n. This equation says that the probability of choosing c monotonically increases with the number of other individuals assigned to the same group who also chose c, averaging over uncertainty about the group assignments. Individuals with similar choices have higher probability of being assigned to the same group and hence will exert a stronger influence on individual m's subsequent choices.

In practice, we add some modeling flexibility by assuming that choice predictions are made using the Luce choice rule (Luce, 1959):

$$\pi_{m*}(c) \propto [P(c_{m*} = c|\mathbf{C})]^{\beta} \tag{6}$$

where $\pi_{m*}(c)$ denotes the choice prediction and β is an "inverse temperature" parameter that governs response stochasticity. When β is close to 0, responses are equiprobable; in the limit $\beta \rightarrow \infty$, the maximum probability response is deterministically chosen. Note that the inverse temperature does not alter the relative ordering of choice predictions, only their magnitude, so the qualitative behavior of the model is the same regardless of the value of β. In modeling our experimental data, we performed a coarse grid search over β (our only free parameter) to find the value that minimized the squared error between model predictions and data from Experiment 1. All model predictions are reported with the best-fit parameter value.

The posterior over groups for individual m is given by Bayes' rule:

$$P(\mathbf{z}|\mathbf{C}) = \frac{P(\mathbf{C}|\mathbf{z})P(\mathbf{z})}{\sum_{\mathbf{z}} P(\mathbf{C}|\mathbf{z})P(\mathbf{z})}. \tag{7}$$

The prior $P(\mathbf{z})$ is given by Eq. 2 above. The likelihood is obtained by analytically marginalizing the latent parameters under the Dirichlet-Multinomial model described in the previous section:

$$\begin{aligned} P(\mathbf{C}|\mathbf{z}) &= \int_{\theta} P(\mathbf{C}|\theta, \mathbf{z})P(\theta)d\theta \\ &= \prod_n \prod_k \frac{\Gamma(|\mathcal{X}_n|\gamma)}{\Gamma(T_k + |\mathcal{X}_n|\gamma)} \prod_c \frac{\Gamma(L^c_{kn} + \gamma)}{\Gamma(\gamma)}. \end{aligned} \tag{8}$$

The likelihood favors groupings for which choices patterns are similar between individuals assigned to the same group.

In summary, by using choice data from individuals, the model infers a posterior distribution over latent group assignments. This allows us to test a number of predictions about social influence on choice behavior. In what follows, we present a series of experimental tests of our model.

3. Experiment 1

Experiment 1 was designed to test the prediction that similar individuals will exert a stronger social influence on choice behavior than dissimilar individuals, where similarity is defined in terms of overlap in choice patterns between two agents. Note that this prediction need not appeal to latent group structure; existing models that rely on the frequency of choice overlap also make the same prediction. The purpose of this experiment was to verify that our current experimental paradigm and stimuli can induce this most basic effect, and that our model shows comparable performance in explaining the results. In this experiment, participants first made choices between pairs of movies and then observed the choices of two other individuals (Fig. 2). We varied the proportion of trials on which these individuals agreed with the participant, so that one of the individuals tended to agree with the participant more than the other. In the critical test trial, we presented a "mystery choice" in which the individuals made opposite choices between a pair of unknown movies (the poster was replaced with a question mark). We then asked participants to choose the movie from the pair that they would prefer to watch. We predicted that, in the absence of other information about the choice set, participants would be more likely to follow the individual with whom they had greater choice overlap.

Fig. 2. Schematic of task in Experiment 1. (A) Participants were presented with choices between pairs of movies. (B) Participants then predicted (i.e., guessed) the choice of another individual and received feedback about his/her actual choice. (C) On the Mystery Choice trials, participants first saw the two other individuals make opposite choices between two unknown movies and then indicated which one they would like to watch.

3.1. Methods

3.1.1. Participants

One hundred and seventy-two adults were recruited via the Amazon Mechanical Turk web service (N = 61, 51, and 60 for 25–75, 50–75, and 75–75 conditions, respectively. We used TurkGate to ensure that participants did not repeatedly participate within and across experiments (Goldin & Darlow, 2013). The current and subsequent experiments (Experiments 1–5) were approved by the Stanford Internal Review Board; all participants in these experiments received informed consent and were paid for their participation.

3.1.2. Materials

Stimuli consisted of movie poster images, drawn from a set of 48 movies. The movies were organized into 24 pairs that were matched by genre (e.g., comedy, horror, action, romance, science fiction). Each participant viewed a random set of 16 movie pairs.

3.1.3. Procedure

On each trial, participants first saw a pair of movie posters and chose which of the two they would rather watch. After making a choice, participants observed two other individuals' choices between the same pair; they first saw a picture of an individual along with his or her name (henceforth A) and were asked to predict which movie this individual would rather watch. The individual's actual choice was then revealed with an arrow pointing to one of the movies on the screen. Participants repeated this with a new individual B, predicting this individual's choice and then observing the actual choice. There were four trials in a block, each with different pairs of movies but all with same individuals. To reduce the memory demand for past choices, participants were presented with a table showing the history of choices for all players, including the participant's own. Each block involved a different set of individuals and movies.

At the end of the block, participants were given a "mystery choice" trial. Participants saw two colored boxes with question marks, representing unknown movies. They first observed A and B make choices, indicated by an arrow pointing to each individual's choice (we made explicit to the participants that A and B knew the movies they were choosing). Critically, A and B always made different choices. Finally, participants were asked which one of the two unknown movies they would rather watch. Each participant completed a total of four blocks. Identities of Individuals A and B changed every block, and all movie pairs were presented only once.

Across different groups of participants, we varied A's choice overlap with the participant (25%, 50%, and 75%) while holding constant B's choice overlap with the participant at 75%. For instance, in the 25–75 condition, A agreed with the participant on 25% of choices while B agreed with the participant on 75% of the trials. There were three conditions in total: 25–75, 50–75, and 75–75. All elements of the blocks (individuals, movies, choices) were randomized except the choice overlap.

3.2. Results and discussion

Choice probabilities for the mystery choice trials, along with model fits, are shown in Fig. 3. The only free parameter fit to the data was the choice inverse temperature β, with maximum likelihood estimate of $\beta = 11.05$. In the 25–75 condition, the model tends to group individual B and the participant together because of the relatively higher overlap between B and the participant. As the agreement between A and the participant increases, the probability of grouping both individuals with the participant increases, leading to an indifference between the two options in the 75–75 condition. Thus, our model captures the way in which people's choices are socially influenced by others, and how this influence is modulated by the degree of similarity in past choice behaviors.

Participants' choices on the mystery choice trials were consistent with these predictions (Fig. 3). As expected, the probability of making the same choice as Individual B declined as the choice overlap between the participant and Individual A increased. A one-way ANOVA showed that there is a significant effect of condition ($F(171) = 7.33$, $p < .001$), and t tests showed that probabilities were significantly greater than chance (0.5) for both the 25–75 condition ($t(60) = 5.03$, $p < .0001$) and the 50–75 condition ($t(50) = 3.03$, $p < .005$). When both individuals equally agreed with the participant (75–75 condition), the probability was not significantly greater than chance ($p = .77$). Participants in the

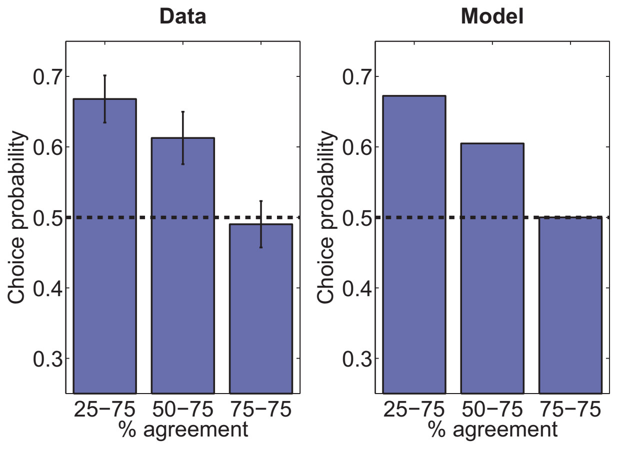

Fig. 3. Experiment 1 results. (Left) Probability of participants making the same choice as B on the mystery choice trial. The degree of agreement between the participant and the other individuals was varied between conditions. For example, "25–75" means that A agreed with the participant on 25% of trials while B agreed with the participant on 75% of trials. The dashed line indicates indifference between the two choices. Error bars represent standard error of the mean. (Right) Model behavior.

25–75 condition chose to follow Individual B significantly more often than participants in the 75–75 condition ($t(119) = 3.79$, $p < .001$). Thus, these results show how the degree of prior overlaps in choices of our own and others influence our future decisions, and it confirms the validity of our model in capturing this effect.

4. Experiment 2

In Experiment 1, we replicated the well-established effect of inter-individual similarity on choice behavior in a simple but naturalistic decision-making scenario (Brock, 1965; Simons et al., 1970) and showed that our model captures the way in which people's choices are more strongly influenced by those with greater choice overlap. However, structure learning via grouping is not necessary to predict and explain these results; a model that represents dyadic similarity in terms of choice overlap (e.g., Akerlof, 1997) could also capture these findings, without invoking a grouping process.

In Experiment 2, we test a novel prediction of our model that cannot be captured by dyadic models. We examine whether differential social influence can occur even when dyadic similarity is equated. A conceptual schematic of our design is shown in Fig. 4. Individuals A and B have equal choice overlap with the participant (P), and hence their dyadic similarity is equated. By adding a third individual (C) who tends to agree with both P and B (75% overlap with B and P, respectively), we can cause P, B, and C to be grouped together. This enhances the influence of B (relative to the influence of A) on P's choice. As C's overlap with P decreases (25% overlap with P but 75% overlap with B), the relative strength of B's influence on P would also diminish.

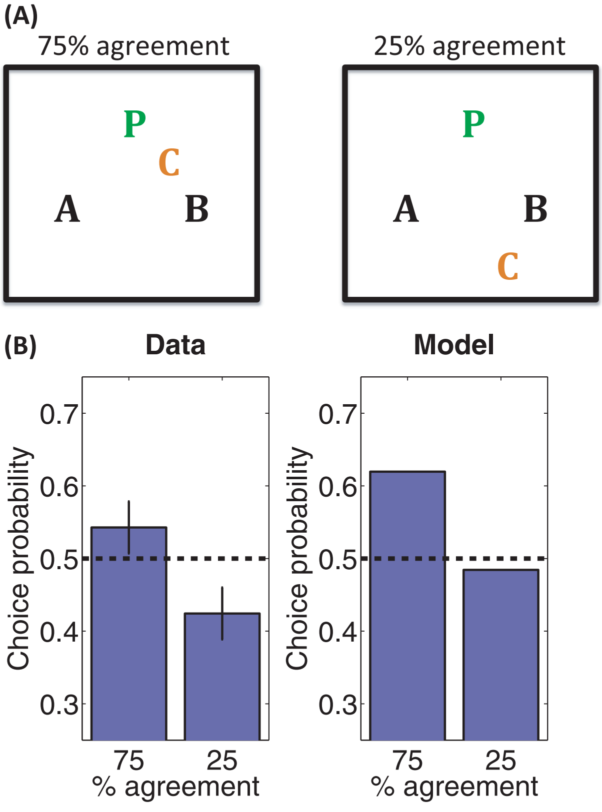

Fig. 4. Experiment 2: schematic and results. (A) Individuals are represented as letters in an abstract space, where the distance between letters indicates the proportion of the same choices made between pairs of individuals (i.e., choice overlap). In the 75% agreement condition, individual C made the same choices as individual P (the participant) and B on 3 out of 4 trials. In the 25% agreement condition, individual C made the same choices as B on 3 out of 4 trials, but the same choices as P on only 1 out of 4 trials. (B) Probability of making the same choice as individual B for the 75% agreement and 25% agreement conditions. Error bars represent standard error of the mean.

Note that we are not arguing against the role of dyadic similarity per se; rather, we are claiming that it is not by itself a sufficient basis for social influence on choice. In order to capture people's choice behaviors, the appropriate measure of similarity must take into account the latent clustering structure. This claim echoes other work which argues that unsupervised learning processes underlie similarity judgments (e.g., Love, Medin, & Gureckis, 2004; Shepard & Arabie, 1979; Tenenbaum, 1996).

4.1. Methods

4.1.1. Participants

One hundred and fifty-two adults were recruited via the Amazon Mechanical Turk web service.

4.1.2. Materials

Stimuli were the same as those used in Experiment 1.

4.1.3. Procedure

The procedure in Experiment 2 was largely the same as in Experiment 1. The main difference was that participants, after choosing between a pair of movies, observed the choices of three individuals (A, B, and C), rather than two. A and B always agreed with the participant on 50% of the trials, while C always agreed with A on 25% of the trials and agreed with B on 75% of the trials. Critically, we varied the degree of overlap between C and the participant (P). In the 75% condition, C agreed with the participant on 75% of choices in a given block; in the 25% condition, C agreed with the participant on 25% of the trials. These conditions were varied within-participant, such that each participant was presented with two blocks of each condition. The final Mystery choice trial was identical to Experiment 1; the participant saw a pair of unknown movies, one chosen by A and the other by B. Participants were asked to choose which one they themselves would rather watch. C did not appear in the Mystery choice trials. As in Experiment 1, identities of A, B, and C remained the same within a block but changed on every block.

4.2. Results and discussion

By design, the choices of the participant were equally similar to the choices of individuals A and B. Thus, if people simply keep track of dyadic choice overlap, they should be indifferent between following A and B's mystery choices. However, our model predicts a bias in people's choices caused by the presence of C. In the 75% condition, B, C, and the participant should be clustered into a single group, leading to a bias to follow the choice of B rather than A. However, this grouping would be weaker in the 25% condition, leading to a relatively weaker tendency to follow B's mystery choice, and possibly even a reversal.

The probability of following B 's choice was significantly different between conditions ($t(151) = 3.31$, $p < 0.005$), such that participants were more likely to follow B in the 75% condition than in the 25% condition. These results cannot be explained by a dyadic similarity account, which would predict indifference between the two options. Our structure learning account captures these results qualitatively without any additional parameter tuning; in generating the model predictions, we used the estimate of choice inverse temperature β from Experiment 1 (Fig. 4). Thus, there are no free parameters in our model predictions. Note that while the model does predict a difference between the two conditions, it does not get the scaling right: overall the choice probability is significantly higher in both conditions. This scaling mismatch can be corrected by fitting β and γ, but we do not explore that possibility here.

5. Experiment 3

Our third experiment was designed to replicate and extend the findings of Experiment 2 by exploring a larger range of choice overlap configurations. Using the same procedure, we manipulated the overlap between individual C and the other individuals (including the participant), while holding the overlap between the participant and individuals A and B constant at 50%. Again, if participants simply consider the inter-individual similarity in making decisions on the Mystery choice trials, they would be indifferent between the two options in all of the conditions. In contrast, if participants group themselves (P), A, B, and C based on their choices, we would see systematic differences across conditions.

5.1. Methods

5.1.1. Participants

In Experiment 3, each participant received one block per condition (compared to two blocks in Experiment 2; see Procedure); thus, we doubled our sample size and recruited 305 adults via Amazon Mechanical Turk.

5.1.2. Materials

The same stimuli were used as in Experiments 1 and 2, with additional stimuli so that the stimuli were entirely non-overlapping across blocks.

5.1.3. Procedure

The procedure in Experiment 3 was nearly identical to the procedure in the previous experiments. The only difference was that each participant was presented with seven blocks (in randomized order), each from a different condition. The conditions varied in the degree of overlap between the various individuals; specifically, we explored seven configurations of choice overlap chosen so that the overlap between the participant and individuals A and B was always fixed at 50%. As in Experiments 1 and 2, the participant was shown the choices of individuals A and B (but not C) on the Mystery choice trials.

5.2. Results and discussion

Using the same value of β estimated from the Experiment 1, the results of Experiment 3 (Fig. 5) shows that our model can reliably predict choice data across conditions without any additional parameter tuning. The model shows reasonably strong correlation with the empirical choice probabilities (r = .79, p < .05, permutation test). As in Experiment 2, the model correctly predicts that the strongest effects would be observed when individual C is most similar to both the participant and individual A, and least similar to individual B. In contrast, under a dyadic similarity account, all conditions would show uniform choice probabilities at 0.5, since the overlap with the participant's choices was always the same for individuals A and B.

The results in Experiments 2 and 3 together suggest that our model captures a phenomenon that is not well explained by dyadic similarity models. We predicted that if people can learn the latent groupings underlying others' choice behaviors, their choices would be differentially influenced by others' choices even when the dyadic similarity is

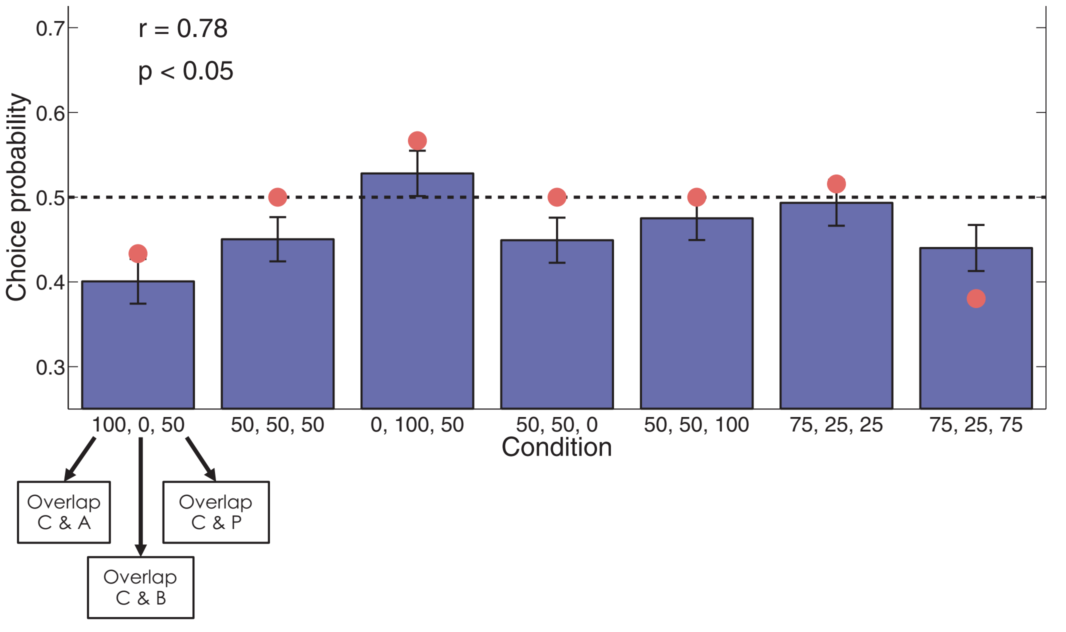

Fig. 5. Experiment 3 results. Average responses (bars) and model predictions (red circles) for each condition. The *x*-axis labels indicate the choice overlap configuration for each condition. Error bars represent standard error of the mean. The first three bars show the results when the overlap between C and P are fixed; the last four bars show the results for different fixed combinations of A–C and B–C overlap.

equated. People's choice behaviors were well predicted by our model, across a broad range of choice overlap configurations.

6. Experiment 4

In previous experiments, we looked at an indirect measure of latent grouping: how people's choices were influenced by those of other individuals based on how they were clustered into latent groups. In Experiment 4, we sought more direct evidence by using the same choice patterns as in Experiment 2 and asking participants to explicitly group the individuals (including themselves) at the end of each block. In essence, we asked participants to do an unsupervised category learning procedure. This procedure allowed us to ask whether participants' ways of grouping individuals were consistent with our model's predictions. We hypothesized that participants would prefer to group themselves with individual B than with individual A, even though A and B had equal overlaps with participants' choices, due to the presence of individual C whose choices were similar to both the participant and individual B than to A.

6.1. Methods

6.1.1. Participants

One hundred and four adults participated via the Amazon Mechanical Turk web service.

6.1.2. Materials

The same stimuli were used as in the previous experiments.

6.1.3. Procedure

The procedure was the same as used in Experiment 2 (75% overlap condition), with one difference: the Mystery choice on each block was replaced by a grouping task. After the choice trials, participants saw icons of four individuals (pictures of A, B, C, and an avatar for the participant (P)) and four blank boxes. They were instructed to group these individuals based on their preferences by dragging the icons into the boxes such that "individuals with similar movie preferences would be placed in the same box" (Fig. 6, left). Participants were told that all boxes were the

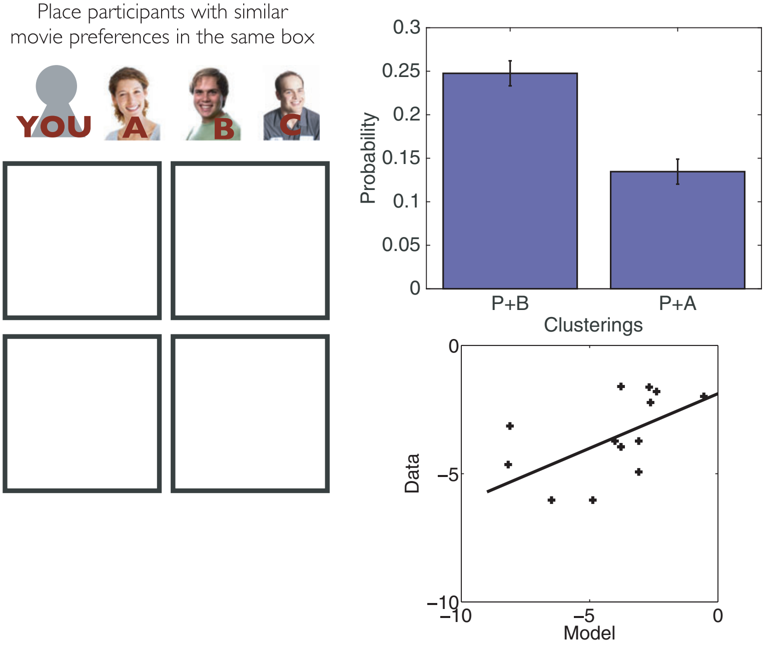

Fig. 6. Experiment 4 task and results. (Left) At the end of each block, participants grouped all the individuals, including themselves, by dragging icons into boxes. The individuals were labeled by names rather than by letters, but we use letters here to maintain consistency with the text and previous figures. (Top right) Groupings in which the participant (P) and individual B (but not A) clustered together, denoted "P + B," were preferred to groupings in which P and A (but not B) were clustered together, denoted "P + A." Error bars represent standard error of the mean. (Bottom right) Scatter plot and least-squares line relating model predictions and empirical cluster probabilities on a log scale (see Table 1).

same (i.e., their locations were meaningless), and that they could choose any assignment of individuals to groups (including assigning each individual to a unique group or putting everyone in the same group) as long as all four individuals were assigned to a group.

6.2. Results and discussion

The complete set of groupings and their choice probabilities is shown in Table 1. We aggregated the grouping responses into two classes: groupings in which the participant (P) and individual B (but not A) were clustered together (denoted "P + B"), and groupings in which P and A (but not B) were clustered together (denoted "P + A"). Fig. 6 (top right) shows the average proportion of blocks in which each grouping class was chosen. The P + B groupings were significantly more likely to be chosen than the P + A groupings ($t(103) = 3.93$, $p < .001$). This is consistent with our model's prediction that P and B will tend to be clustered together because of the presence of C, despite no differential similarity between P and B relative to P and A (see Fig. 4). This result holds when we exclude groupings in which A is clustered with C ($p < .005$).

To compare these results quantitatively to our model predictions, we computed the Pearson correlation between log-transformed model posterior probabilities and empirical probabilities, finding a significant correlation between the two ($r = .57$, $p < .05$; Fig. 6, bottom right). Even though we allowed a large degree of freedom in participants' grouping responses, these results show that our model's predictions about grouping reasonably predict people's responses independently of its predictions about choice. This further

Table 1
Probability for each grouping. Each letter denotes an individual (see text for details). For example (AC, BP) denotes a grouping in which A and C are in one group and B and P are in the other

Grouping	Mean Frequency (± SEM)
(ABCP)	0.002 (0.002)
(AC, BP)	0.024 (0.008)
(AB, CP)	0.202 (0.026)
(AP, BC)	0.108 (0.017)
(CP, AB)	0.137 (0.028)
(A, BCP)	0.009 (0.009)
(B, ACP)	0.024 (0.008)
(ABP, C)	0.002 (0.002)
(P, ABC)	0.002 (0.002)
(AP, C, B)	0.009 (0.009)
(P, C, AB)	0.009 (0.005)
(A, P, BC)	0.176 (0.022)
(P, AC, B)	0.019 (0.023)
(A, P, C, B)	0.019 (0.007)

supports our hypothesis that such representations about latent groups might underlie participants' choices in Experiments 1–3.

7. Experiment 5

In Experiment 4, we asked people to cluster individuals into explicit groups, providing the most direct evidence for the inference of latent groups. Our results from Experiments 1–3 can be construed as a natural consequence of such inference; people's choices were influenced by those of other individuals based on the latent group structure. In these experiments, we looked at the influence of latent groups on people's choices for *unknown items*, thus removing any possible effect of existing preference or prior knowledge. After seeing two individuals choosing each one of two "mystery" items, people were more likely to choose the item that was preferred by the individual who was inferred to belong in the same latent group as themselves. In real-world choices, however, we rarely encounter choices between two novel, unknown items; often we are faced with multiple items, many of which are already colored by our prior knowledge and varying degrees of liking.

Here we provide a more stringent and ecologically valid test of our hypothesis by asking whether the social influence of latent groups can change people's *previous ratings* for familiar items. Specifically, participants first rated a set of movies based on how much they liked the movies. Then they performed the same choice task used in Experiment 2, in which participants indicated their choice between pairs of movies (same movies used in the rating task) and observed two other individuals' choices. After this task, participants were asked to rate the same movies again. We hypothesized that the difference in people's ratings before and after the choice task would reflect the latent group structure inferred from the choice patterns.

During the choice task in this experiment, as in all the previous experiments, we asked participants' choices for movies (i.e., which movie they would rather see), rather than their real preferences for these movies or which movie they actually liked better. In Experiment 5, we asked participants to rate the movies as a coarse measure their existing preference for these items. However, while we made an effort to use only well-known movies, we remain agnostic about whether participants have seen the movies before. Thus, we use "changes in ratings" to refer to changes in their expected (anticipated) utility rather than changes in the actual experienced utility. While these changes might reflect a change in their actual preferences, our data do not necessarily show that changes in ratings indicate changes in the degree to which a participant actually enjoyed a particular movie.

7.1. Methods

7.1.1. Participants

One hundred and nine adults participated via the Amazon Mechanical Turk web service.

7.1.2. *Materials*

For the choice task, we used the same posters used in the previous experiments. For the rating task, 14 additional movie posters were used; all were posters of real movies and were different from those used in the choice task.

7.1.3. *Procedure*

At the beginning of the experiment, participants rated 14 movies on a 1–10 scale based on how much they would like to watch each movie. They were then presented with seven blocks of the choice task (using different movies), with the same structure as in Experiment 2. Instead of the Mystery choice trial at the end of each block, participants were shown a pair of movies randomly selected from the 14 movies they had initially rated, and saw that individual A chose one of them and individual B chose the other (recall that the presence of individual C in the social learning task caused the participant to group themselves with B and C, but not A). Participants were asked to enter new ratings for the two movies. Thus, by the end of seven blocks, participants had given two ratings (before and after the choice task) for all 14 movies.

7.2. *Results and discussion*

Our dependent measure in this study was "change in ratings": the difference in the rating of a movie before and after observing the individuals' choices. We separated the 14 movies based on whether it was endorsed by individual A or B. According to our model, the participant should be more likely to group themselves with B than with A, and therefore B's endorsement should lead to an increase in ratings, whereas A's endorsement should not. This is precisely what we found (Fig. 7); participants exhibited a significantly positive change in ratings in response to B's endorsement ($t(108) = 3.09$, $p < .005$), but not in response to A's endorsement ($t(108) = 1.32$, $p = .19$). There was no significant difference in the magnitude of change between the two groups ($p = .14$). Our experimental

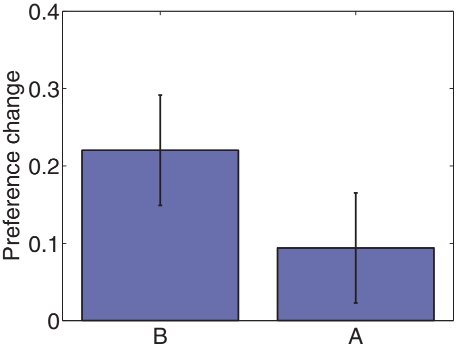

Fig. 7. Experiment 5 results. Participants significantly changed their ratings in response to observing individual B's endorsement, but not in response to observing individual A's endorsement. Error bars represent standard error of the mean.

results suggest that latent grouping can not only influence people's choices about about unknown items, but also lead to graded changes in anticipated utility (i.e., how much they would like to watch the movies in the future) in a more realistic scenario where participants brought in their real-world knowledge (e.g., of movie identity) and previous experiences.

8. Further applications of the model

In this section, we go beyond applying our model to our experimental tasks to show that the idea of latent social grouping can explain a broader range of phenomena. Specifically, we examine two classes of prior empirical work that are well captured by our model. First, we simulate the effects of informational social influence as studied in the social psychology literature. We then simulate a developmental shift in infants' understanding of others' preferences.

8.1. Modeling informational social influence

Traditionally, research on social influence has focused on the multiple factors giving rise to conformity and compliance (see Cialdini & Goldstein, 2004, for a review). Deutsch and Gerard (1955) made the important distinction between *informational* and *normative* sources of social influence, where the former is based on the desire to correctly interpret observational data, while the latter is based on the desire to obtain social approval. The behavioral phenomena studied in the present paper can be understood as a form of informational social influence. A number of quantitative models have been developed to explain informational social influence, most notably the Social Impact Model of Latané (1981) and the Social Influence Model of Tanford and Penrod (1984). Both models specify a function relating the degree of social influence to the number and strength of influence sources. While these models can account for a large number of social influence phenomena, they are essentially descriptive. These models also cannot explain why individuals with equal choice overlap can exert differential social influence. Here we show that our latent grouping model, which is derived from more fundamental principles of induction and rational choice, can be applied to conformity and compliance effects in choice behavior.

Within the domain of social decision making, a number of studies have explored the effects of parametric variations of group structure on social influence (Burnkrant & Cousineau, 1975; Cohen & Golden, 1972; Pincus & Waters, 1977). These studies, along with others reviewed in Burnkrant and Cousineau (1975), Cohen and Golden (1972), and Pincus and Waters (1977), have shown that social influence increases with the number (i.e., group size) and uniformity of influence sources. As a proof of concept that our model can account for these effects, we have simulated simple choice scenarios in which members of a group choose between two options (Fig. 8). Consistent with the experimental data, the probability of following the majority choice increases with the number of

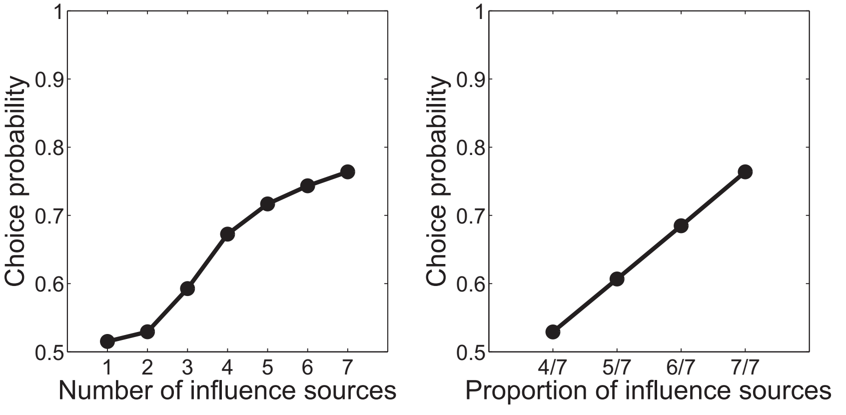

Fig. 8. Modeling classical social influence effects. (Left) The probability of choosing the same option as other group members ("influence sources"), plotted as a function of group size. (Right) Probability of choice as a function of group uniformity. For example, 4/7 indicates that 4 out of 7 group members followed the majority choice (i.e., low uniformity).

group members and the degree to which they agree. These effects stem from the fact that size and uniformity of influence sources both strengthen the inference of a single latent group that includes the experimental subject. Indeed, we do not claim that our model provides a comprehensive account of informational social influence; here we have focused on a particular domain to show that our model can be extended beyond the scope of our experimental paradigm. However, we believe that the same principles can be extended to other domains as well, including young children's tendency to trust majority opinions both in normative and informational contexts (Corriveau, Fusaro, & Harris, 2009; Hu, Whalen, Buchsbaum, Griffiths, & Xu, 2015).

8.2. *Modeling the development of preference understanding*

Motivated by our model's success in predicting empirical data from adults, in this section we ask whether our model can also explain developmental changes in children's understanding of others' preferences. We focus on a study by Repacholi and Gopnik (1997) on infants' ability to represent other people's preferences that are either similar or different from their own. In this study, an adult experimenter expressed a preference that was either consistent or inconsistent with the child's own preference. In the Matched condition, an actor showed positive emotions toward a snack that was also appetizing to the child (e.g., goldfish crackers) and disgust toward the snack that was unappetizing to the child (e.g., raw broccoli). In the Mismatched condition, the actor's expressions were reversed such that her preferences contradicted the child's own (i.e., disgust toward goldfish crackers and pleasure towards raw broccoli). When the experimenter requested the child to give her some snacks, most 14-month-olds subsequently offered the actor the food that was congruent with their own preferences regardless of condition, whereas

18-month-olds offered the actor the food that matched her preferences (i.e., offering broccoli in the Mismatched condition). Repacholi and Gopnik (1997) interpreted these results as evidence that 18-month-olds, but not 14-month-olds, are able to represent others' preferences that are different from their own.

Lucas et al. (2014) recently presented a formal model to explain the data from Repacholi and Gopnik (1997) as a result of "model comparison": 14-month-olds hold a simple model of preferences, such that everyone shares the same preferences, whereas 18-month-olds have a more flexible model in which different individuals have distinct preferences. The developmental change in children's behavior reflects a shift from the simpler to the more flexible model, driven by accumulation of data that cannot be supported by the simpler model (i.e., observing that some people's choices differ from their own, or from other individuals). Such an account predicts a training effect; by providing more data that the simpler model cannot explain, younger infants could be trained to behave more like 18-month-olds.

Consistent with this idea, Doan, Denison, Lucas, and Gopnik (2015) showed that training that involves observations of heterogeneous choices of other agents indeed improves infants' performance on the original Repacholi and Gopnik paradigm. Fourteen- to seventeen-month-old infants saw a series of training trials in which an experimenter who liked different toys and foods either from the other experimenter or the infant herself (the "diverse desires training" condition), or trials in which an experimenter liked the same toys and foods as the other experimenter or the infant (the "non-diverse desires training" condition). Following training, the infants were given a choice between sharing with one of the experimenters a food item that she had previously shown disgust toward (but the infant liked) or a food that she had previously expressed preference toward (but the infant disliked). Infants in the diverse desires training condition chose to share the food that the experimenter liked, whereas infants in the non-diverse desires training condition chose to share the food that they liked. Consistent with the formal model offered by Lucas et al. (2014), these results suggest that infants exposed to a more complex pattern of preferences will acquire a more complex model of individual differences in preferences,

Inspired by prior empirical and formal work, we asked whether the developmental shift and the effect of training can also be explained by our model. A critical difference between our model and the model proposed by Lucas et al. (2014) is that rather than positing a change in the preference models themselves, we suggest that the developmental shift reflects a shift in children's priors over groupings. As noted earlier, the concentration parameter α in our model controls the bias to cluster agents into smaller or larger number of groups. In our modeling of the adult data reported above, the value of α was automatically inferred. In simulating Repacholi and Gopnik's results with our model (using the same value of β estimated from Experiment 1), we assume that young children start off with a small value of α, based on the idea that limited experience inductively supports simpler structures (see Gershman, Blei, & Niv, 2010; Kemp & Tenenbaum, 2008); thus, our model strongly favors a structure in which the actor and the participant are grouped together despite the non-overlapping choices. Older children are assumed to have larger values of α, allowing them to accommodate a more complex social structure with multiple groupings depending on varying preferences of agents. In other words, larger values of α favor structures in which the actor and the participant are assigned to different groups (Fig. 9).

The increase in α is a natural consequence of joint Bayesian inference over $\mathbf{z}$ and α. As more complex data are observed (e.g., conflicting preferences between more agents), the marginal posterior over α will shift toward higher values. This allows us to naturally predict the training effect observed in Doan et al. (2015). Indeed, simulations (shown in Fig. 10) confirm that the model matches the empirical data qualitatively, without additional parameter tuning.

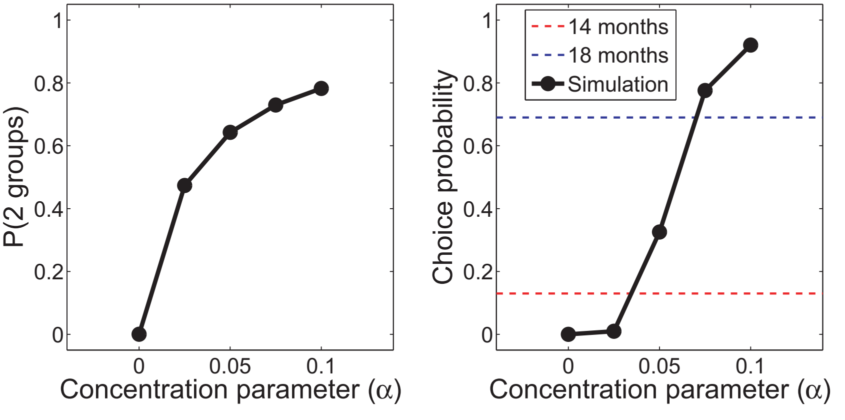

Fig. 9. Simulation of Repacholi and Gopnik (1997), Mismatched condition. (Left) Probability of two-group structure (as opposed to single-group structure) as a function of the concentration parameter α. In the two-group structure, the actor and participant are assigned to separate groups. (Right) Probability of offering the actor the object that matches the actor's preference. The dashed lines show average responses reported by Repacholi and Gopnik (1997).

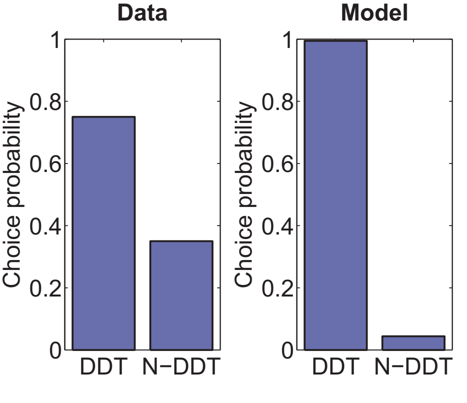

Fig. 10. Effects of training on infant social inferences. (Left) Probability of offering the actor the object that matches the actor's preference, following "diverse desires training" (DDT) or "non-diverse desires training" (N-DDT). Data adapted from Doan et al. (2015). (Right) Simulation results.

The fact that our model requires more data to justify more complex structures has important implications for the development of children's understanding of others' preferences, beliefs, and beyond. Our proposal is consistent with prior empirical and computational work that explains development as a theory formation process in which structural knowledge changes with accumulation of data (Carey, 2009; Gershman et al., 2010; Goodman, Ullman, & Tenenbaum, 2011; Gweon et al., 2010; Kemp & Tenenbaum, 2008). It is important to note that these simulations do not necessarily show that our model performs better than prior models (Lucas et al., 2014) in predicting the empirical data; indeed, both models emphasize the role of data and are consistent with the empirical results. Our goal here is not to compare model performance but to propose an alternative way of explaining developmental change that can be easily extended to include a broader set of results.

Our model naturally predicts young children's developing abilities to learn from others' choices. Prior empirical work shows how young children use information about social groups to guide their own choices (Frazier et al., 2012; Shutts et al., 2009, 2010), and use observations of choice data to make predictions about social affiliations (Liberman, Kinzler, & Woodward, 2014). Our model could be extended to explain these results, further supporting the idea that humans, even early on, readily infer structures of social groups and use these representations in making decisions. Furthermore, our simulation of Repacholi and Gopnik (1997) suggests that people's tendency to assign agents to different groups might develop over time; this would predict a developmental change in children's responses to tasks we used in our experiments. Such results will provide further support for our model and our theoretical proposal, and make empirical contribution to show that even young children infer the structure of their social environment and use such structure to guide their choices even in the absence of explicit cues to group membership. We look forward to extending our work to directly examine this possibility.

9. General discussion

In this paper, we presented a computational model of social structure learning and a series of experiments that tested the predictions of our model. The data suggest that humans naturally cluster individuals into groups on the basis of their choice patterns, and they use these groupings to guide their choices for novel items. With only a single free parameter fit to the data from Experiment 1, our model predictions closely matched human judgments in several experiments that varied in their task format and dependent measures. In Experiment 1, consistent with prior work, we showed that parametrically increasing the degree of choice overlap between pairs of individuals directly increases the strength of social influence between these pairs. In Experiment 2, we tested a novel prediction of our model to provide empirical support for our social structure learning account. We showed that two individuals can have differential influence on our choices even when their previous choices are equally similar with our own, by virtue of a third individual whose choice patterns alter the group structure. These findings were well explained by our Bayesian model, which infers the structure of latent social groups from patterns of choice behavior. Experiment 3 replicated and extended this finding by testing a wider range of choice patterns. Experiment 4 provided more direct support for our model by showing that people's explicit grouping judgments are consistent with our model's predictions of inferred latent groups. Experiment 5 demonstrated that latent grouping can even lead to changes in people's ratings of familiar choices, suggesting the possibility that these latent groups can be powerful enough to influence people's existing preferences. Finally, we showed that our model can be applied to explain a broad range of social phenomena, such as the effect of group size and uniformity on informational social influence and the developmental changes in children's inferences about others' preferences.

Although we used a Bayesian modeling framework to formalize latent grouping, other clustering algorithms (e.g., K-means) might also have been able to capture our findings. The advantage of the Bayesian framework is that it makes explicit the probabilistic assumptions that we impute to the decision-maker and can be flexibly extended to more complex choice settings, a point we return to below. It also provides a natural decision-theoretic framework for modeling choice behavior. Further work is needed to evaluate alternative accounts of our data. The important point is that we have effectively ruled out a large class of models that do not represent latent preference structure (i.e., dyadic similarity models); our claim is that social influence on human choice behaviors cannot be fully understood without explaining how latent groups are inferred from choice data.

The phenomena of interest in our current study are inherently social—the data people observed are choices that reflect others' underlying preferences. Nevertheless, one might wonder if it would be possible to create analogous non-social tasks that produce similar behaviors (e.g., varying the degree to which multiple objects have overlapping and non-overlapping features, instead of presenting agents who exhibit overlapping or non-overlapping choices). We believe that it may be possible, and in fact, such results would be consistent with our idea that common inferential processes might underlie learning the structure of social groups from choice data and learning the structure of object kinds from object features. However, social groups in the real world are presumably more dynamic, complex, and multi-layered compared to object categories, and our reasoning about others' actions and choices often depends on attributions of unobservable yet stable properties of people, such as goals, beliefs, or preferences. Thus, learning the structure of social groups from observing others' choices might recruit additional cognitive capacities, such as our ability to understand, represent, and reason about others' mental states (theory of mind).

As explained earlier, our main goal here was not to directly explain how people infer preferences (e.g., predicting the degree to which an agent will actually like an item), but to explain how people use observed choice data to guide their future choices. Our model does this by computing anticipated choice probability given the inferred structure of social groups, rather than appealing to mental representations such as preferences; however, it is reasonable to imagine that once a latent social structure is inferred in a given domain, people might naturally attribute abstract, stable, and systematic choice structures that we informally refer to as "preferences," and even assign explicit labels to refer to such groups (e.g., punk rock fans). In future work, we hope to apply our model to the kinds of social learning in which these mental state inferences allow people to go beyond observed choice data.

Building on prior work on simple social influence, our framework provides a broader account of how, and what, we learn about others from their choices, and how such learning informs our own choices. Prior research in social and developmental psychology has found a strong influence of group membership on choice (Frazier et al., 2012; Shutts et al., 2009, 2010), showing how existing groups can shape our preferences via an expectation that certain utilities are shared among group members. However, in the real world, we simultaneously belong to multiple groups with varying levels and scopes. Furthermore, we routinely encounter individuals with rich and heterogeneous preferences across a wide range of domains, such that their group memberships cannot be clearly identified. Our work provides an important step toward explaining how, despite these challenges, people can use everyday observations of choices to inform future decisions. People are capable of generating novel groups based on the perceived pattern of choices; these groups not only affect our choices but also our social lives (e.g., whom to affiliate given overlaps in choices; Vélez, Bridgers, & Gweon, 2016) and reflect the dynamic nature of cultural phenomena at a larger scale, such as fads, trends, and customs. Our account is an important step toward a more precise understanding of our tendency to affiliate with those who share our preferences, and how grouping facilitates social learning and reasoning in large-scale societies.

9.1. Related work

Our work is related to several lines of research in social and developmental psychology. The effect of interpersonal similarity on social influence has been studied extensively (see Simons et al., 1970, for a review). For example, Brock (1965) demonstrated that a paint salesman was more effective at inducing a customer to switch to a higher price level if the salesman reported a similar magnitude of paint consumption. Developmental studies have revealed that children prefer objects endorsed by other children of their own age and gender (Shutts et al., 2010), or adults speaking in their native language (Shutts et al., 2009). While there are existing models that formalize the effect of similarity on social influence (e.g., Akerlof, 1997), these models do not explain how *latent* groups can structure social influence, as we report in Experiments 2 and 3.

Recently, a number of studies have explored the multinomial logit model, originally developed in econometrics (McFadden, 1973), as a theory of preference learning (Bergen et al., 2010; Jern, Lucas, & Kemp, 2011; Lucas et al., 2014). This model parameterizes the utility function as a linear combination of object features and assumes that choice probabilities are a nonlinear function of utilities. Our current model is agnostic to features of choices. Accordingly, in our experiments, agents' choice behaviors had no systematic relationship to the such features (e.g., movie genre, cast, etc.); agents simply chose the same items as the participant on varying proportions of trials. This was a deliberate design choice to isolate the effect of structure learning. Although participants may have also used feature information in their reasoning, this would not predict our behavioral results.

The key innovation of our model, missing from the multinomial logit model, is the incorporation of structure learning, which allows it to discover latent groupings of individuals who share similar choice patterns. We emphasize that such groupings are essential for understanding the results of Experiment 2. An interesting future direction of research will be to modify the multinomial logit model to incorporate structure learning. This can be accomplished by allowing the utility functions to arise from a *mixture distribution*, where each mixture component corresponds to a latent group.

More broadly, our work is related to theories of structure learning in other domains (see Gershman & Niv, 2010, for a review). The pioneering work in this area was carried out by John Anderson in his analysis of category learning (Anderson, 1990 1991), which posited that the distributional structure of categories was specified by a set of latent mixture components. Because the number of mixture components is unknown, Anderson defined a nonparametric model that would discover them in an unsupervised manner; this model was later recognized to be identical to the CRP (Sanborn, Griffiths, & Navarro, 2010). Variants of this model have been applied to many other inductive problems with combinatorial latent structure, ranging from classical conditioning (Gershman et al., 2010) to word segmentation (Goldwater, Griffiths, & Johnson, 2009). The model has also been generalized in a number of significant ways, such as allowing simultaneously active latent features and nested structures (Austerweil et al., 2015). However, until now these structure learning principles have so far not found extensive application in social learning.

9.2. Future directions

We envision a number of other directions for future research that will enrich the theoretical framework presented here. First, we have assumed that all individuals within the same group share a similar set of utility functions, or preferences (on average). Clearly humans have heterogenous preferences even within particular groups. We can accommodate this additional source of heterogeneity by defining a hierarchical model in which individuals possess unique preferences whose central tendency is group-specific. This will allow us to capture both the normative influence of group membership and the existence of individuality.

Second, here we have considered the social influence in just one direction: from "other individuals (A, B, and C)" to the participants' choices. We also did not explicitly specify whether these other individuals were aware of the participants' choices. However, to the extent that people can observe each other's choices, the social influence should be bidirectional, and the individuals whose choices are widely broadcasted should exert stronger influence on others' choices. This is consistent with the intuition that in real-world social environments, there are individuals whose preferences and choices have a huge impact on many people's behaviors (e.g., fashion choices of Hollywood stars). One interesting extension of our work is to directly manipulate the degree of public availability of individuals' choices and ask whether our model also can also capture this intuition.

Third, we have assumed a single grouping structure based on the entire set of choices; however, evidence suggests that humans can learn cross-cutting categories, which allow multiple layers of groupings based on different subsets of object features (Shafto, Kemp, Mansinghka, & Tenenbaum, 2011). In the domain of choice, cross-cutting groupings are very natural: you might follow one friend's choices for movies but follow a different friend's choices for restaurants, depending on who is grouped together with you in what domain. We can extend our model to capture multiple groupings by simultaneously learning the structure of individuals and choice sets, analogous to the work of Shafto et al. (2011). This approach has a particularly interesting implication for future developmental research. Unlike adults, young children might initially expect that two agents who share preferences in one domain must also share preferences in other domains, and gradually overcome this bias with more data. Future computational and developmental work should test this prediction of something akin to "mutual exclusivity" (Markman & Wachtel, 1988) in preference learning.

Finally, although we have ignored explicit cues to salient social groups (e.g., age, gender, and language), we note that it is possible to generalize our prior to favor groupings based on these explicit cues. Indeed, these cues still do exert certain influences on our real-world choice behaviors, and one interesting empirical question is whether participants will show an initial inductive bias toward these biologically and culturally specified groups that can be overridden by additional data.

9.3. Conclusion

We have presented a novel computational model that formalizes how individuals learn the latent structure of social influence from observations of others' choice behaviors. Our model predictions are consistent with behavioral data from five experiments, and they also offer an elegant and parsimonious explanation for various existing empirical work including developmental shifts in children's understanding of others' preferences. These results raise exciting questions for future research, and they provide the first steps toward understanding the nature of inferential processes and representations that allow us to learn from and influence others.

Acknowledgments

We are grateful to Jason Mitchell for helpful suggestions and to Walid Bendris for assistance with data collection. This research was supported by start-up funds from Harvard University and Stanford University, and the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216, and Varieties of Understanding grant from the John Templeton Foundation.

References

Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica: Journal of the Econometric Society*, *65*, 1005–1027.

Aldous, D. (1985). Exchangeability and related topics. In *École d'Été de probabilités de Saint-Flour xiii* (pp. 1–198). Berlin: Springer.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Psychology Press.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Austerweil, J., Gershman, S., Tenenbaum, J., & Griffiths, T. (2015). Structure and exibility in Bayesian models of cognition. In J. Busemeyer, J. Townsend, Z. Wang, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology* (pp. 514–522). Oxford, UK: Oxford University Press.

Bergen, L., Evans, O. R., & Tenenbaum, J. B. (2010). Learning structured preferences. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 853–858). Austin, TX: Cognitive Science Society.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, *100*(5), 992–1026.

Brock, T. C. (1965). Communicator-recipient similarity and decision change. *Journal of Personality and Social Psychology*, *1*, 650–654.

Burnkrant, R. E., & Cousineau, A. (1975). Informational and normative social influence in buyer behavior. *Journal of Consumer Research*, *2*, 206–215.

Carey, S. (2009). *The origin of concepts*. Oxford, UK: Oxford University Press.

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends Cognitive Sciences*, *3*, 57–65.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591–621.

Cohen, J. B., & Golden, E. (1972). Informational social influence and product evaluation. *Journal of Applied Psychology*, *56*, 54–59.

Cook, M., & Mineka, S. (1989). Observational conditioning of fear to fear-relevant versus fear-irrelevant stimuli in rhesus monkeys. *Journal of Abnormal Psychology*, *98*, 448.

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow preschoolers prefer nondissenters as informants. *Psychological Science*, *20*, 372–377.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*, 629–636.

Doan, T., Denison, S., Lucas, C., & Gopnik, A. (2015). Learning to reason about desires: An infant training study. In D.C. Noelle et al. (Eds.), *Proceedings for the Annual Meeting of the Cognitice Science Society* (pp. 578–583). Austin, TX: Cognitive Science Society.

Fawcett, C. A., & Markson, L. (2010). Children reason about shared preferences. *Developmental Psychology*, *46*, 299–309.

Frazier, B. N., Gelman, S. A., Kaciroti, N., Russell, J. W., & Lumeng, J. C. (2012). Ill have what shes having: The impact of model characteristics on childrens food choices. *Developmental Science*, *15*, 87–98.

Galef, B. G., & Giraldeau, L.-A. (2001). Social influences on foraging in vertebrates: Causal mechanisms and adaptive functions. *Animal Behaviour*, *61*, 3–15.

Gergely, G., Bekkering, H., & Király, I. (2002). Developmental psychology: Rational imitation in preverbal infants. *Nature*, *415*, 755–755.

Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*, 1–12.

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology*, *20*, 251–256.

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, *117*, 197–209.

Goldin, G., & Darlow, A. (2013). Turkgate (version 0.4. 0) [software].

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, *112*, 21–54.

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. *Psychological Review*, *118*, 110–119.

Gweon, H., & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science*, *332*, 1524–1524.

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, *107*(20), 9066–9071.

Hu, J., Whalen, A., Buchsbaum, D., Griffiths, T. L., & Xu, F. (2015). Can children balance the size of a majority with the quality of their information? In D.C. Noelle et al. (Eds.), *Proceedings of the 37th Annual Cognitive Science Society* (pp. 956–961). Austin, TX: Cognitive Science Society.

Hu, J., Lucas, C. G., Griffiths, T. L., & Xu, F. (2015). Cognitive development. *Cognitive Development*, *36*, 93–102.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The Naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20* (8), 589–604.

Jern, A., & Kemp, C. (2015). A decision network account of reasoning about other peoples choices. *Cognition*, *142*, 12–38.

Jern, A., Lucas, C. G., & Kemp, C. (2011). Evaluating the inverse decision-making approach to preference learning. In J. Shawe-Taylor et al. (Eds.), *Advances in neural information processing systems* (Vol. 24, pp. 2276–2284), Red Hook, NY: Curran Associates.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*, 10687–10692.

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, *21*, 1134–1140.

Latané, B. (1981). The psychology of social impact. *American Psychologist*, *36*, 343–356.

Liberman, Z., Kinzler, K. D., & Woodward, A. L. (2014). Friends or foes: Infants use shared evaluations to infer others' social relationships. *Journal of Experimental Psychology: General*, *143*, 966–971.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309–332.

Lucas, C. G., Griffiths, T. L., Xu, F., Fawcett, C., Gopnik, A., Kushnir, T., Markson, L., & Hu, J. (2014). The child as econometrician: A rational model of preference understanding in children. *PloS One*, *9*, e92160.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: John Wiley and Sons.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*, 838–850.

Pincus, S., & Waters, L. (1977). Informational social influence and product quality judgments. *Journal of Applied Psychology*, *62*, 615–619.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., ... Laland, K. N. (2010). Why copy others? Insights from the social learning strategies tournament. *Science*, *328*, 208– 213.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14-and 18-month-olds. *Developmental Psychology*, *33*, 12–21.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*, 1144–1167.

Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*, 382–389.

Shafto, P., Kemp, C., Mansinghka, V., & Tenenbaum, J. B. (2011). A probabilistic model of cross-categorization. *Cognition*, *120*, 1–25.

Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, *86*, 87–123.

Shutts, K., Kinzler, K. D., McKee, C. B., & Spelke, E. S. (2009). Social information guides infants' selection of foods. *Journal of Cognition and Development*, *10*, 1–17.

Shutts, K., Banaji, M. R., & Spelke, E. S. (2010). Social categories guide young children's preferences for novel objects. *Developmental Science*, *13*, 599–610.

Simons, H. W., Berkowitz, N. N., & Moyer, R. J. (1970). Similarity, credibility, and attitude change: A review and a theory. *Psychological Bulletin*, *73*, 1–16.

Steiniger, F. (1950). Beiträge zur soziologie und sonstigen biologie der wanderratte. *Zeitschrift für Tierpsychologie*, *7*, 356–379.

Tanford, S., & Penrod, S. (1984). Social influence model: A formal integration of research on majority and minority influence processes. *Psychological Bulletin*, *95*, 189–225.

Tenenbaum, J. (1996). Learning the structure of similarity. *Advances in Neural Information Processing Systems*, *9*, 3–9.

Tenenbaum, J., Kemp, C., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285.

Vélez, N., Bridgers, S., & Gweon, H. (2016). Not all overlaps are equal: Social affiliation and rare overlaps of preferences. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.) *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1985–1990). Austin, TX: Cognitive Science Society.

Weitzman, M. (1965). Utility analysis and group behavior: An empirical study. *The Journal of Political Economy*, *73*(1), 18–26.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541.

Whittler, T. E., & DiMeo, J. (1991). Viewers' reactions to racial cues in advertising stimuli. *Journal of Advertising Research*, *31*, 37–46.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1–34.