How Does Instance-Based Inference About Event Frequencies Develop? An Analysis with a Computational Process Model

Christin Schulze (cschulze@mpib-berlin.mpg.de)

Thorsten Pachur (pachur@mpib-berlin.mpg.de)

Ralph Hertwig (hertwig@mpib-berlin.mpg.de)
Max Planck Institute for Human Development, Lentzeallee 94
Berlin, 14195 Germany

Abstract

To make inferences about the frequency of events in the world (e.g., the prevalence of diseases or the popularity of consumer products), people often exploit observations of relevant instances sampled from their personal social network. How does this ability to infer event frequencies by searching and relying on personal instance knowledge develop from childhood to adulthood? To address this question, we conducted a study in which children (age 8–11 years) and adults (age 19–34 years) judged the relative frequencies of first names in Germany. Based on the recalled instances of the names in participants' social networks, we modeled their frequency judgments and the underlying search process with a Bayesian hierarchical latent-mixture approach encompassing different computational models. We found developmental differences in the inference strategies that children and adults used. Whereas the judgments of most adults were best described by a noncompensatory strategy that assumes limited and sequentially ordered search (*social-circle model*), the judgments of most children were best described by a compensatory strategy that assumes exhaustive search and information aggregation (*availability-by-recall*). Our results highlight that already children use instance knowledge to infer event frequencies but they appear to search more exhaustively for instances than adults. One interpretation of these results is that the ability to conduct ordered and focused search is a central aspect in the development of noncompensatory instance-based inference.

Keywords: child development; sampling; probabilistic inference; heuristics; availability

Introduction

The relative frequency of events in the world is an important ecological characteristic that impacts many actions and decisions. For instance, the relative frequency of other people's behaviors hints at social norms that should be followed; the number of people having bought different products may indicate differences in product quality that influence consumer choice; and the prevalence of diseases hints at health risks that may guide precautionary actions. Decision makers commonly do not have access to summary tables of these frequency statistics but need to infer them. An easily accessible but informative indicator for event frequencies in the population is their occurrence among the people one knows personally. That is, by searching for relevant instances in their personal social network people can collect a variety of information about the frequency of events in the world, and use this information to form subjective frequency judgments. In this paper, we examine how this ability to search proximate social spaces to judge the relative frequency of events develops from childhood to adulthood.

Previous work has garnered much insight into how adults make instance-based inferences. Most prominently, according to Tversky and Kahneman's (1973) availability heuristic, adults judge the frequency of events by assessing the ease with which instances of the events can be brought to mind. More recent research has elaborated the specific mechanisms guiding this search in and retrieval from mnemonic sample spaces. For instance, it has been shown that adults often restrict search to directly experienced instances in their social circles and that these social circles are searched sequentially (e.g., Hertwig, Pachur, & Kurzenhäuser, 2005; Pachur, Hertwig, & Rieskamp, 2013).

Yet, currently only little is known about how search for information in proximate social spaces develops ontogenetically. Do already children exploit their social memories to draw inferences about the frequency of events in the world? And if so, how much do they sample, in which order do they consult social circles, and how do they integrate the information to draw inferences? Existing developmental work on judgment and decision making is consistent with opposing predictions. On the one hand, working memory limitations may confine young children to using information-frugal strategies because processing and integrating large amounts of evidence may be difficult (e.g., Bereby-Meyer, Assor, & Katz, 2004). On the other hand, limitations in the ability to selectively focus attention on relevant information may lead young children to use more exhaustive but unsystematic search strategies (e.g., Davidson, 1991; Mata, von Helversen, & Rieskamp, 2011).

To disentangle these opposing predictions, we first introduce the *social-circle model*, a cognitive process model that parameterizes key components of the inference process—including search order, evidence threshold, and response noise. Second, we take a Bayesian hierarchical mixture approach to modeling the inferences of children and adults in a task in which they made judgments about the relative frequency of common first names in Germany.

The Social-Circle Model

To model people's inferences based on recalled instances, Pachur et al. (2013) proposed that people search sequentially through the circles of their social network—defined as self, family, friends, and acquaintances—and stop search as soon as the instance evidence in a circle allows them to make an inference. It is thus assumed that people's search for relevant instances is guided by the well-documented hierarchical structure in the ordering of discrete social groups that make up a person's social network (e.g., Hill & Dunbar, 2003; Milardo, 1992), which has also been shown to be important for search in social memory (e.g., Hills & Pachur, 2012). Adults' frequency judgments have been found to be equally well described by a model that assumes such a noncompensatory strategy and by a more exhaustive, compensatory search strategy (Pachur et al., 2013). Here, we formalize and extend the assumptions in Pachur et al.'s (2013) analysis and propose a generalized *social-circle model* (SCM) that allows for variability in the order in which circles are inspected and for probabilistic aspects in the search, stopping, and decision stages of inference.

The SCM assumes that in order to judge which of two events, A or B, is more frequent in the population, decision makers search distinct social circles, defined as self, family, friends, and acquaintances. At each inspected circle i the evidence, e_i, is represented as the difference in the number of instances recalled for each event, expressed as a proportion:

$$e_i = \frac{n_{iA}}{n_{iA}+n_{iB}} - \frac{n_{iB}}{n_{iA}+n_{iB}}. \tag{1}$$

Search Rule

The order in which the circles are inspected is represented by circle-weight parameters, one for each circle (w_i; constrained by $\sum w_i = 1$; see Bergert & Nosofsky, 2007), that can be estimated from the data. These weights represent the probability that a circle is inspected as

$$p(inspect\ circle_i) = \frac{w_i}{\sum_i^N w}. \tag{2}$$

Once a circle has been inspected, it is not considered further (i.e., the denominator is calculated only over circles that have not yet been inspected). Note that search is thus assumed to be probabilistic. The probability of following a particular search order $p(order_j)$ is given by the product of the individual probabilities of circle inspection,

$$p(order_j) = \prod_j p(inspect\ circle_i). \tag{3}$$

Stopping Rule

In the SCM it is assumed that the proportional evidence obtained from each circle is compared against a decision threshold, d. If the evidence from the recalled instances reaches or exceeds the threshold, a choice is made; if it is lower than the threshold, the next circle is inspected. The SCM implements a probabilistic version of this stopping rule by assuming normally distributed error for each circle, denoted as ε_i, generated from a normal distribution with mean zero and standard deviation σ. Specifically, it is assumed that, if the evidence in a given circle (with added error) meets or exceeds d, then the decision maker selects option A (i.e., $|e_i + \varepsilon_i| \geq d$); if the evidence meets $-d$, then the decision maker selects option B (i.e., $|e_i + \varepsilon_i| \leq -d$). Thus, the probability of making a choice after inspection of circle i is given by

$$\begin{aligned} p_i(choice) &= p(|e_i + \varepsilon_i| \geq d_i) \\ &= p(e_i + \varepsilon_i \geq d_i) + p(e_i + \varepsilon_i \leq -d_i) \\ &= \Phi\left(\frac{e_i - d_i}{\sigma}\right) + \Phi\left(\frac{-e_i - d_i}{\sigma}\right), \end{aligned} \tag{4}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Decision Rule

The probability of selecting option A based on a particular order, $p_j(A|AB)$, follows from combining the choice probabilities resulting from circle inspection in that order (cf. Rieskamp, 2008). For example, for the order j = 1,2,3,4:

$$\begin{aligned} p_{j=1,2,3,4}(A|AB) &= p_1(A|AB) + [1 - p_1(choice)] \times \\ & p_2(A|AB) + [1 - p_1(choice)] \times [1 - p_2(choice)] \times \\ & p_3(A|AB) + [1 - p_1(choice)] \times [1 - p_2(choice)] \times \\ & [1 - p_3(choice)] \times p_4(A|AB). \end{aligned} \tag{5}$$

The total probability of selecting option A is defined as the sum of all $p_j(A|AB)$, each weighted by the probability of the decision maker following the order (see Equation 3):

$$p(A|AB) = \sum_{j=1}^{N!} p_j(A|AB) \times p(order_j). \tag{6}$$

In sum, the SCM parameterizes three key components of instance-based inference: the decision maker's preferred search order (circle weight parameters, w_i), evidence threshold (d), and response noise (σ). Thus, depending on its parametrization, the model can capture various noncompensatory inference processes. In what follows, we apply the SCM to inference data from an experiment in which children and adults were asked to judge the relative frequency of common first names in Germany, and examine how well it accounts for participants' inferences compared to a compensatory strategy and a guessing strategy.

Experiment

Method

Participants Forty children (age 8–11 years; 18 female) and 40 adults (age 19–34 years; 19 female) who were recruited via the subject pool of the Max Planck Institute for Human Development participated in the experiment. The data of five additional children were excluded from the analysis because the children showed insufficient reading-comprehension (two children aged 7 years); did not recall any or only one instance for each name in the same social circle, yielding a guessing prediction for instance-based models on every trial (two children); or terminated the experiment prematurely (one child). Participants received a performance-based payment (earning 0.04 EUR for each correct inference but losing the same amount for each incorrect inference; 1 EUR ≈ 1.12 USD at the time of the experiment), and an additional flat fee of 10 EUR.

Materials Table 1 lists the 22 first names (11 female) that were used in the experiment. Because no census data about the frequency and distribution of first names in Germany was available, we approximated a frequency ranking by weighting popular baby names between 1911 and 2010 (Bielefeld, 2016) with each cohort's proportion in the population to date (Statistisches Bundesamt, 2014).[1] We constructed a set of all possible 231 paired comparisons of the names, and informed participants that the accuracy of their inferences was judged on the basis of the available data. Participants were instructed to ignore the particular spelling of each name and to judge the relative frequency of names by taking possible variants of a name into account.

Procedure The experiment consisted of two tasks, an inference and a retrieval task, that were completed by all participants in this order. In the *inference task*, participants were asked to judge which of two first names is more frequent in Germany for each of the 231 name pairs. The pairs were presented sequentially on a computer screen in blocks of 23 pairs (24 pairs in the final block). The order in which name pairs were presented was randomized across participants; the order of names in each pair was predetermined so that correct and incorrect inferences (according to our statistics) were distributed equally across the two response alternatives. Each trial started with the display of a fixation cross at the center of the screen, followed by the presentation of two black silhouettes (either male or female) which were labeled with the respective names in the comparison (see Figure 1A). Participants made a selection by pressing one of two designated keys on the keyboard. After each choice, the selected name's silhouette was shown on a podium at the center of the screen to confirm the selection to the participant. There was no trial-by-trial feedback about the accuracy of decisions. Participants were encouraged to make as many correct judgments as possible. There was a self-paced pause after each block and participants completed two training trials with fictitious names before the start of the inference task. In the *retrieval task*, participants were asked to recall how many people with each of the 22 names shown in Table 1 they knew personally. For each name, participants counted each person among their family, friends, and acquaintances with that name by dragging and dropping pictorial representations of family members, friends, and acquaintances on a black silhouette labeled with the respective name (see Figure 1B). Following the retrieval of a person, participants were also asked to indicate their contact frequency with that person on a scale from one (less than once every six months) to five (multiple times per week). Additionally, participants could allocate a pictorial person labeled "self" to indicate the shown name was their own. Each recalled person was listed on the screen and counted toward an overall tally of persons with a particular name also shown on the screen. Before the start of the retrieval task, a training trial familiarized participants with the controls of this task. At the end of the experiment, participants were informed about their overall accuracy on the inference task and paid in cash by the experimenter.

Table 1: The 22 first names used in the experiment, their approximated frequency rank in Germany, and the total number of instances children and adults recalled from their own social networks.

Name	Gender	Rank	Total number of recalled instances	
			Children	Adults
Michael	m	1	35	66
Thomas	m	2	34	72
Peter	m	3	29	45
Andreas	m	4	34	65
Jan	m	5	40	67
Hans	m	6	22	26
Christian	m	7	29	76
Karin	f	8	14	24
Ursula	f	9	4	15
Julia	f	10	34	78
Anna	f	11	41	70
Sabine	f	12	29	44
Stefanie	f	13	24	58
Renate	f	14	19	20
Helga	f	15	18	17
Günter	m	16	11	16
Tim	m	17	40	43
Horst	m	18	11	17
Angelika	f	19	16	27
Lukas	m	20	39	46
Hannah	f	21	42	44
Gertrud	f	22	6	9

Bayesian Hierarchical Mixture Modeling Based on the instances of names that each participant recalled from their social network in the retrieval task, we modeled each participant's decisions in the inference task with a Bayesian latent-mixture approach (see, e.g., Bartlema, Lee, Wetzels, & Vanpaemel, 2014). Hierarchical mixture modeling allows

[1] We scored the top 30 male and top 30 female first names between 1911 and 2010 in Germany (Bielefeld, 2016) on a scale from 30 (for the most popular male/female name in a year) to zero (for names not listed during a year). These scores were then weighted, for each gender separately, by the proportion of people in the German population who belong to the cohort (Statistisches Bundesamt, 2014). We selected the most popular male and female name in each decade based on the summed raw scores each name received across these ten-year periods. In addition to these 20 most popular names from each decade, we selected the most frequent male and female name in the population (that was not already in the list) based on the total sum of the weighted scores across all years. Finally, the 22 selected names were ranked based on the sum of their weighted scores across all years.

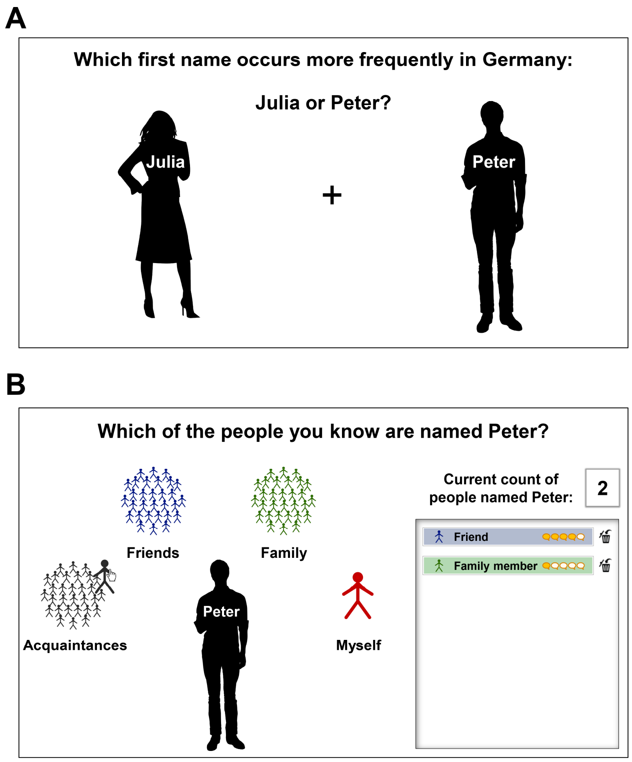

Figure 1: Illustration of the task screen and controls used during the inference task (A) and retrieval task (B).

us to simultaneously estimate discrete classes of participants who use categorically different inference strategies and to robustly model variation within each group of strategy-users, thus combining the advantages of pooling continuous individual differences hierarchically and assuming discrete differences among groups of individuals. We assumed three latent subgroups of participants, each using a different inference strategy: (a) the social-circle model, (b) *availability-by-recall*, which assumes a compensatory process (Hertwig et al., 2005; Pachur, Hertwig, & Steinmann, 2012), and (c) a random guessing strategy.

According to availability-by-recall, all instances of an event are tallied across the entire social network and the option with the larger summed instance-evidence is chosen. For comparability, we applied the same response noise mechanism as for the SCM, which gives the probability of choosing option A as

$$p_{AbR}(A|AB) = \Phi\left(\frac{n_A - n_B}{\sigma}\right), \tag{7}$$

where n_A denotes the number of instances recalled for event A across all circles and σ is a response noise parameter. For the guessing strategy, we assumed that participants randomly selected one of the two names in each pair with probability .50. With this approach, we can estimate the proportion of participants using each strategy based on inference and recall data while taking into account the uncertainty surrounding such a classification. We modeled participants' inferences for all paired comparisons on which a participant's instance knowledge allowed each strategy to make an unambiguous prediction. The two instance-based strategies did not make a prediction, if a participant recalled no or equal numbers of instances for both names in a comparison. The posterior distributions of model parameters were estimated via Gibbs sampling methods implemented in JAGS (Plummer, 2003). We used reasonably uninformative priors: For the w_i and d parameters of the SCM we assumed uniform priors on the group-level mean (beta distributions with shape parameters of 1) and gamma priors (with a shape parameter of 1.1051 and a scale parameter of 0.01051; see Bartlema et al., 2014) on the group-level precision. For the σ parameters of the SCM and availability-by-recall we assumed uniform distributions constrained between 0.01–40 on the group-level mode and standard deviation. For the latent-mixture indicator variable we assumed a categorical prior that assigned equal weight to each strategy.[2] To ensure efficient mixing, we used pseudo-priors that approximate the posterior density for the individual-level parameters. These pseudo-priors were obtained from an initial Bayesian hierarchical estimation procedure that was performed separately for each model (without a mixture component). In the model estimation, 16 chains each with 50,000 samples drawn from the posterior distributions were run after an initial burn-in period of 2000 samples. Gelman–Rubin statistics and visual inspections of the four chains indicated adequate chain convergence.

Results

Behavioral Data We found differences between the age groups in inferential accuracy, $t(78) = 5.17$, $p < .001$, $d = 1.16$, $BF_{10} = 8362$, and in reported instance knowledge, $t(60.00) = 4.68$, $p < .001$, $d = 1.05$, $BF_{10} = 1456$. On average, adults picked the more frequent first name more often than children ($M = .64$ vs. $M = .57$) and recalled more people with any of the 22 first names in their social network ($M = 23.63$ vs. $M = 14.28$; see also Table 1). One possible reason for children's lower inferential accuracy is that the instances they reported were less valid indicators of the actual frequency distribution of first names in the population (possibly because they recalled fewer instances overall). That is, for adults, there was a significant rank correlation between reported instances and actual frequency ranks, $r_S(20) = .524$, $p = .012$, $BF_{10} = 4.99$. For children, however, no such correlation was found, $r_S(20) = .203$, $p = .364$, $BF_{10} = 0.39$.[3]

[2] For few participants, this resulted in the mixture collapsing on the SCM. For these participants, we used a prior that assigned low initial weight to the SCM (e.g., .001) and equal weight to the other two strategies. To ensure unbiased estimation of latent group-membership, these unequal priors were taken into account in the calculation of membership probabilities.

[3] Yet children's inferences were well calibrated to their cohort's instances. Evaluating inferences based on a ranking derived from children's reported number of instances, flips the accuracy pattern such that children significantly outperform adults, $t(78) = -2.40$, $p = .019$, $d = -.536$, $BF_{10} = 2.70$.

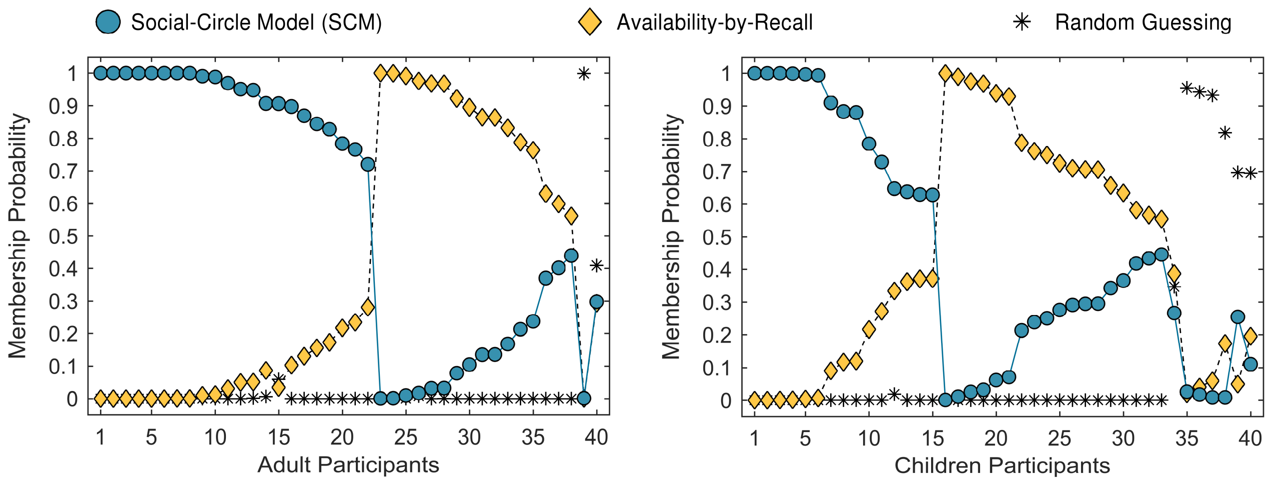

Figure 2: Allocation of adult and children participants to three latent subgroups of strategy users.

Computational Modeling Figure 2 shows the membership probability of each adult (left panel) and each child (right panel) in each group of strategy-users, as derived from the posterior distribution of the latent-mixture variable. The figure shows that the judgments of most adults were best described by the SCM (55% of adults compared to 38% of children). By contrast, the judgments of most children were best described by availability-by-recall (48% of children compared to 40% of adults). Only few participants were best described by the guessing strategy. Overall, there was greater uncertainty in the classification of children to latent groups than in the classification of adults. This was partly due to the lower number of instances children recalled resulting in poorer discriminability between the models.

Next, we compared children's and adults' search and decision processes by evaluating their group-level SCM parameter estimates. As shown in Figure 3, children and adults weighted the different circles in their social network similarly (although adults showed greater inter-individual variability in the weighting of different circles), applied similar decision thresholds, and did not differ on the response noise parameter (for all parameters, 95% HDIs overlapped). Children's lower inferential accuracy was thus not due to a more error prone execution of an instance-based inference strategy. This also held for inferences described by availability-by-recall.

Discussion

Our results suggest that already children systematically exploit their instance knowledge to make inferences about the frequency of events in the world. However, they do so differently than adults. Whereas the judgments of most adults were best described by a strategy that assumes limited information search, the judgments of most children were best accounted for by a strategy that assumes exhaustive search. This finding echoes previous research on multi-attribute choice and cue-based inference which has found young children to use more exhaustive but unsystematic search strategies (e.g., Davidson, 1991; Mata et al., 2011). A possible explanation for why children use more information-intensive strategies is that they have difficulties to selectively attend to relevant and diagnostic information (cf. Betsch, Lehmann, Lindow, Lang, & Schoemann, 2016). In young children, this inability to effectively focus search may be driven by the required executive control functions being not yet fully developed (see Best & Miller, 2010). In light of

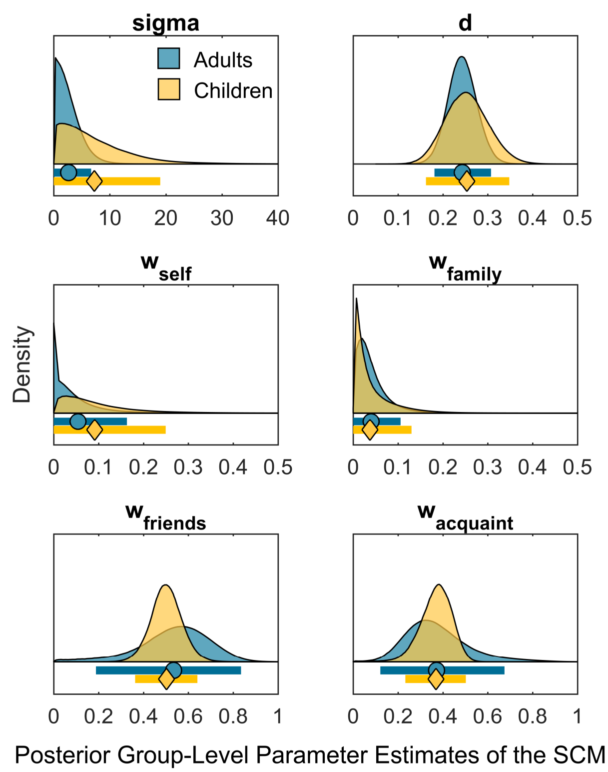

Figure 3: Posterior distributions of the group-level parameters of the SCM. Small circles and diamonds below the density plots show the posterior means for adults and children, respectively; lines show 95% HDIs.

children's more limited and less ecologically valid instance knowledge, their greater tendency to adopt exhaustive sampling strategies might represent an adaptive response to these limiting factors. However, it should also be noted that, due to children's lower instance knowledge, the discriminability between models was lower, which might have contributed to the more balanced strategy classification in children as well.

Our results extend previous research that has found children to use availability as a cue for judging the relative frequency of and their own memory for names (Davies & White, 1994; Geurten, Willems, Germain, & Meulemans, 2015). This prior work, however, did not use cognitive modeling to formalize and quantitatively analyze the development and use of instance-base inference strategies. By taking a formal computational modeling-based approach, our analysis enabled us to simultaneously detect developmental differences in the use of discrete strategies and parameterize the specific mechanisms underlying search for instances in memory. This approach highlighted that children search for instances more exhaustively but weight the subgroups in their social network similarly as do adults. The analysis also revealed substantial individual differences in the process of search for instances in memory among both age groups. In this respect, the social-circle model that we applied provides an advantage over previously proposed models of instance-based inference (e.g., Tversky & Kahneman, 1973), which are silent regarding the specific mechanisms and order of instance sampling.

We conclude that the social-circle model provides an effective tool for capturing and illuminating individual and group differences in the cognitive processes that underlie instance-based inference. The insights gained with this model are consistent with the finding that search in social memory is guided by factors such as social proximity (Hills & Pachur, 2012) and suggest that one important factor in the development of information-frugal strategies for judging frequencies is the ability to limit and selectively focus search on relevant instance knowledge.

References

Bartlema, A., Lee, M., Wetzels, R., & Vanpaemel, W. (2014). A Bayesian hierarchical mixture approach to individual differences: Case studies in selective attention and representation in category learning. *Journal of Mathematical Psychology*, *59*, 132–150.

Bereby-Meyer, Y., Assor, A., & Katz, I. (2004). Children's choice strategies: The effects of age and task demands. *Cognitive Development*, *19*, 127–146.

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 107–129.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*, 1641–1660.

Betsch, T., Lehmann, A., Lindow, S., Lang, A., & Schoemann, M. (2016). Lost in search: (Mal-)adaptation to probabilistic decision environments in children and adults. *Developmental Psychology*, *52*, 311–325.

Bielefeld, K. (2016). *Statistiken der häufigsten Vornamen für jeden Geburtsjahrgang* [*Statistics of the most common first names for every cohort*; data set]. Retrieved from http://www.beliebte-vornamen.de/jahrgang

Davidson, D. (1991). Children's decision-making examined with an information-board procedure. *Cognitive Development*, *6*, 77–90.

Davies, M., & White, P. A. (1994). Use of the availability heuristic by children. *British Journal of Developmental Psychology*, *12*, 503–505.

Geurten, M., Willems, S., Germain, S., & Meulemans, T. (2015). Less is more: The availability heuristic in early childhood. *British Journal of Developmental Psychology*, *33*, 405–410.

Hertwig, R., Pachur, T., & Kurzenhäuser, S. (2005). Judgments of risk frequencies: Tests of possible cognitive mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 621–642.

Hill, R. A., & Dunbar, R. I. M. (2003). Social network size in humans. *Human Nature*, *14*, 53–72.

Hills, T. T., & Pachur, T. (2012). Dynamic search and working memory in social recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 218–228.

Mata, R., von Helversen, B., & Rieskamp, J. (2011). When easy comes hard: The development of adaptive strategy selection. *Child Development*, *82*, 687–700.

Milardo, R. M. (1992). Comparative methods for delineating social networks. *Journal of Social and Personal Relationships*, *9*, 447–461.

Pachur, T., Hertwig, R., & Rieskamp, J. (2013). Intuitive judgments of social statistics: How exhaustive does sampling need to be? *Journal of Experimental Social Psychology*, *49*, 1059–1077.

Pachur, T., Hertwig, R., & Steinmann, F. (2012). How do people judge risks: Availability heuristic, affect heuristic, or both? *Journal of Experimental Psychology: Applied*, *18*, 314–330.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*.

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1446–1465.

Statistisches Bundesamt. (2014). *Bevölkerung nach Geburtsjahren und Geschlecht für Deutschland* [*German population by year of birth and gender*; data set]. Retrieved from http://www.zensus2011.de

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.