Running head: THE DEVELOPMENT OF CUE ABSTRACTION PROCESSES IN JUDGMENT

Do Children Profit From Looking Beyond Looks?

From Similarity-Based to Cue Abstraction Processes in Multiple-Cue Judgment

Bettina von Helversen[1,2], Rui Mata[3] and Henrik Olsson[1]

1) Max Planck Institute for Human Development, Berlin, Germany

2) University of Basel, Department of Psychology, Switzerland

3) University of Lisbon, Portugal

Contact:

Bettina von Helversen,

University of Basel, Department of Psychology

Missionsstr. 62a, 4055 Basel, Switzerland

Phone: 0041 61 267 3528

Fax: 0041 61 267 0613

Email: Bettina.vonhelversen@unibas.ch

Abstract

We investigated 9-11 year-olds' and adults' ability to use similarity-based and rule-based processes as a function of task characteristics in a task that can be considered either a categorization task or a multiple-cue judgment task depending on the nature of the criterion (binary vs. continuous). Both children and adults relied on similarity-based processes in the categorization task. However, adults relied on cue–abstraction in the multiple-cue judgment task while the majority of children continued to rely on similarity-based processes. Reliance on cue abstraction resulted in better judgments for adults but not for children in the multiple-cue judgment task, suggesting that 9-11 year-olds may have defaulted to similarity-based processes because they were not able to efficiently employ a cue abstraction process.

Keywords: categorization; multiple-cue judgment; cognitive development; similarity; cue abstraction

Do Children Profit From Looking Beyond Looks?

From Similarity-Based to Cue Abstraction Processes in Multiple-Cue Judgment

Humans can rely on multiple sources of information and processes to make judgments, such as how dangerous a given dog is. For one, humans can rely on similarity-based processes; that is, they can use the similarity of a particular dog to previously encountered dogs to judge the threat and, ultimately, to make the decision to either flee or pet the dog in question. Similarity-based processes based on exemplar memory seem to underlie human judgment in a variety of categorization tasks (e.g., Juslin, Olsson, & Olsson, 2003; Kruschke, 1992; Nosofsky & Johansen, 2000). According to exemplar theories, objects are categorized based on their similarity to known exemplars: An object to be categorized is compared to other objects in memory and placed in the category whose members it is more similar to. The similarity between two objects is determined by the overlap of features weighted by the attention allocated to them. Alternatively, humans can use rules specifying the relationship between the cues (i.e., the characteristics of the objects) and the criterion that is judged, such as ‘dogs that bark, don’t bite’ to assess the threat posed by the animal (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994). This cue abstraction process can be described by a linear additive model such as multiple linear regression, which has been shown to capture human performance in a number of multiple-cue judgment tasks (e.g., Brehmer, 1994; Cooksey, 1996; Juslin, Olsson et al., 2003).

But can children rely on both types of processes to make judgments? The ability to make accurate judgments and decisions develops with age (e.g., Bereby-Meyer, Assor, & Katz, 2004; Davidson, 1991; Krascum & Andrews, 1993; Lafon, Chasseigne, & Mullet, 2004). While adults have been found to switch between the use of similarity-based and cue abstraction processes as a function of task characteristics (Juslin, Olsson et al., 2003; but see Karlsson, Juslin, & Olsson, 2008, for a discussion of situations where process switching is not automatic), children's ability to rely on cue abstraction processes may be constrained. For instance, it has been argued that young children tend to rely more heavily on similarity-based processes in both categorization and induction tasks (e.g., Fisher & Sloutsky, 2005; Sloutsky, Kloos, & Fisher, 2007), while complex rule use develops later in adolescence (Bunge & Zelazo, 2006; Kloos & Sloutsky, 2008). In this paper, we investigate the ontogeny of similarity-based and cue abstraction processes by having 9-11 year-olds and adults solve a task that is either a categorization task or a multiple-cue judgment task depending on the nature of the criterion (binary vs. continuous). We thus aim to assess potential factors underlying children's judgment deficits when relying on similarity-based and cue abstraction processes.

From Similarity-Based to Cue abstraction Processes

Adults can make use of both similarity-based and cue abstraction processes as a function of task structure. For example, Juslin, Olsson et al. (2003) found that feedback quality can play a major role in determining the type of process underlying multiple-cue judgments: when learning the toxicity of insects, participants receiving binary information about the toxicity of each insect (toxic vs. harmless) relied on similarity-based processes, while those receiving more fine-grained information about each insect's toxicity (i.e., level of poison) used cue abstraction processes. These results suggest that the enhanced quality of the feedback in the continuous condition allowed the abstraction of rules connecting the cues to the criterion, while the relatively poor feedback in the categorization task forced participants to rely on the similarity to the training exemplars.

Can children rely on cue abstraction processes in such judgment tasks or must they default to similarity-based processes? Piaget (1952) theorized that logical, formal operations are difficult for children younger than 11 years. Empirical work suggests that children rely heavily on similarity-based processes in both categorization and induction (Boswell & Green, 1982; Fisher & Sloutsky, 2005; Sloutsky, Kloos, & Fisher, 2007, but see Wilburn & Feeney,

2008) and have difficulties integrating information across several cues in inference tasks (e.g., Bereby-Meyer et al., 2004; Lafon et al., 2004). Possibly, cue abstraction processes require controlled processing that puts high demands on cognitive capacities such as working memory (De Caro, Thomas, & Beilock, 2008; Juslin, Jones, Olsson, & Winman, 2003; but see Juslin, Karlsson, & Olsson, 2008, for a discussion of the role of controlled processing in cue abstraction and similarity-based processes), which develops fully only in adolescence. This idea matches the view that complex rule use develops late in ontogenetic time due to late maturation of specific prefrontal brain structures (Bunge & Zelazo, 2006) and thus could be difficult to master for pre-teenage children.

The existing evidence nevertheless suggests that rule use is not completely off-limits to young children (Bunge & Zelazo, 2006; Gelman & Waxman, 2007). Two-year olds have been found to use category labels in inductive inference (Gelman & Coley, 1990). Montanelli (1972) showed that third-graders can already integrate several pieces of information in a linear additive manner—albeit less efficiently than seventh- or ninth-graders. Similarly, Lafon et al. (2004) reported the use of linear additive rules by children despite wide age differences and non-standard cue relations (i.e., negative). All in all, these results suggest that age differences in judgment tasks might rather be due to deficient rule use brought about by, for example, children's difficulties in attributing correct weights to cues (Lafon et al., 2004), rather than complete reliance on similarity-based processes. In the present study we addressed these issues in a multiple cue judgment task.

The Present Study

We adapted the multiple-cue judgment task used by Juslin, Olsson et al. (2003) to explore age differences in the use of similarity-based and cue abstraction processes as a function of task structure. The specific aspect of task structure that we manipulated was the nature of the criterion, which was either binary, creating a categorization task, or continuous, creating a multiple-cue judgment task. Through formal modeling of children and adults' cognitive processes in these tasks, we wanted to find out whether children (9-11 year-olds) would rely more on similarity-based processes, even in conditions where adults typically apply cue abstraction processes (multiple-cue judgment task). We thus aimed to explore whether reliance on similarity-based vs. cue abstraction processes could account for age differences in judgment performance. Furthermore, we were interested in determining whether 9-11 year-olds were able to go beyond similarity-based processes and efficiently rely on cue abstraction processes which supposedly develop later in ontogenetic time (e.g., Lafon et al., 2004; Sloutsky et al., 2007).

Method

Participants. Fifty children (M_{age} = 10.22 years, range: 8.6 - 11.1; 52% female) and 50 adults (M_{age} = 24.92 years, range: 19-33; 54% female) participated in the study. Children were third and fourth graders recruited from Berlin elementary schools in a middle-class neighborhood. 84% of the adults were students from one of the Berlin Universities with on average 2.5 years of University education; 12% had received a master's degree and 4% d had graduated from a Realschule[1]. Most participants were Caucasian. Participation took an average of 40 minutes for adults and 1 hour for children. Participants received a performance-contingent payment (*M* = €15, including €5 show-up fee).

Design and Material. The design consisted of two between-subjects factors: task (binary vs. continuous criterion) and age group (children vs. adults). The participants' task was to find out how well fictitious characters, the Sonics, performed in a game in which each Sonic needed to catch as many Golbis as possible. In the binary task participants needed to classify each Sonic as a successful or an unsuccessful hunter. In the continuous task they needed to estimate how many Golbis a Sonic had caught. The Sonics varied on four cues (hair, nose, ears, and belly) which could be used to predict how well they performed in the game.[2] Each cue could have two features; for example, the belly was either green or blue, and the hair had spikes or dread locks. The number of Golbis a Sonic caught varied between 10 and 20 and was determined as a linear function of the cues (for a similar task see, Juslin, Olsson et al., 2003; Juslin, Jones et al., 2003):

$$C = 10 + 4c_1 + 3c_2 + 2c_3 + 1c_4 \tag{1}$$

where C is the criterion in the continuous task and c_1 to c_4 the cue values, which could be either ‘one’ or ‘zero’ (see Table 1). In the binary task, the probability with which each Sonic was categorized as ‘successful’ or ‘unsuccessful’ was determined on the basis of the continuous criterion; Sonics with criterion values above 15 were classified as ‘successful’ and those with criterion values below 15 were classified as ‘unsuccessful’. The assignment of cue weights to the four cues, as well as of the cue values to the features, was randomly determined.

Procedure. The task consisted of a training phase and a test phase. In the training phase, a training set consisting of six of the 16 Sonics was repeatedly presented (see Table 1). The 6 Sonics used in the training set were selected so that the exemplar and cue abstraction models made different predictions for the new objects in the test phase given the assumption that the two models were applied without error.[3] To minimize age differences due to learning speed and to reduce the impact of training performance on test performance, training was terminated after the 8th block if an accuracy criterion was reached. The accuracy criterion was reached if the root mean squared deviation (RMSD) between participants’ responses and the criterion values in one block sank below .5 in the binary task (corresponding to a correct classification of 5 out of 6) and below 1.5 in the continuous task (corresponding to an average error of less than 1.5 Golbis). If the accuracy criterion was not met in training blocks 8 to 13, training terminated after the 13th block.[4] In the test phase all 16 Sonics were presented twice, thus amounting to a total of 32 judgments.

Payment was performance dependent: in the binary task, participants received 10 points for a correct answer and 0 points for an incorrect answer; in the continuous task, participants received 10 points for a correct answer, 5 points if they deviated by 1, and 0 points if they deviated by more than 1. After the experiment points were converted into Euros with an exchange rate of €.5 for every 100 points. Participants were paid an extra €3 if they reached the accuracy criterion within the 13 blocks. To additionally motivate participants to try hard during the test set where no feedback was provided, we introduced an additional bonus of €4 if they reached an accuracy criterion in the test set. The cover story for this accuracy criterion was that from the 32 Sonics (each of the 16 Sonics was presented twice) evaluated in the test set the 16 Sonics judged as the most successful hunters would form a team and play against the remaining 16 Sonics. The performance of each team was calculated using Equation 1 based on the characteristics of the selected Sonics such that the team that caught more Golbis won. If their team won, they received the additional 4€.

The participants were first introduced to the Sonics in a familiarization task: One of the six training Sonics was presented and participants were asked to memorize it. All 16 possible Sonics were then presented and participants selected the memorized Sonic. The familiarization task was repeated until participants had correctly recognized each of the training Sonics twice. Afterwards, the six training Sonics were presented to the participants with their criterion values. The experimental part of the study began with the extensive training phase consisting of 8-13 blocks. In each block the training Sonics were presented in random order. In each trial of the training phase participants were asked to judge the performance of a training Sonic. After giving their response they received feedback about their performance, the correct criterion value, and the points they earned. During the test phase participants did not receive feedback. After performing the decision task, participants completed a verbal knowledge test (Lehrl, 1999) and two measures of fluid abilities (Wechsler, 1981): the digit-symbol substitution task and the digit span task (forward and backward) that we do not address in this paper.

Results

Participants' Performance

Training. We assessed participants' performance by the accuracy of their judgments, measured by the *RMSD* between judgments and the criterion values. Performance at the end of the training – the last block of training each participant experienced – differed between age groups, despite our attempts to equate performance by allowing children more learning trials (for means and *SD*s see Table 2; binary task: $t(48) = 1.96$, $p = .05$, $d = .53$, continuous task: $t(48) = 2.97$, $p <.01$, $d = .84$). Furthermore, age differences were more pronounced in the continuous task. In the binary task, all but one child reached the accuracy criterion. In the continuous task, 11 children and 3 adults failed to reach the criterion. An ANOVA with age group and task as between-subjects factors and the number of training blocks that each participant needed as the dependent variable showed a significant effect of age group ($F(1, 96) = 25.36$, $MS = 59.29$, $p = .001$, partial $\eta^2 = .21$), a significant effect of task ($F(1, 96) = 32.37$, $MS = 75.69$, $p = .001$, partial $\eta^2 = .25$), and an interaction between task and age group ($F(1, 96) = 8.66$, $MS = 20.25$, $p = .01$, partial $\eta^2 = .08$), indicating that age differences were greater in the continuous task. Because we found age differences in training and these could potentially affect our conclusions, we additionally conducted all subsequent analyses controlling for accuracy in the last block of training. However, for the sake of clarity, throughout the paper, we report results from these additional analyses only for the subset that showed that controlling for training performance had an effect on the results.

Test. Performance during the test set was measured as the RMSD between participants' judgments and the criterion values. Means and *SD*s are reported in Table 2. Because effects of differences in the choice of processes would be particularly expected for new test objects, we separated the test objects into old items (i.e., Sonics that had appeared during training) and new items (i.e., Sonics that had not appeared during training). In the binary task, adults were more accurate than children for the old items ($t(48) = 5.71$, $p = .001$, $d = 1.64$) but no significant age differences emerged for the new test items ($t(48) = 1.60$, $p =$

.12, *d* = .50). In the continuous task, adults were more accurate than children for old items (t(48) = 2.61, p = .01, d = .73), but the age difference in accuracy for new items only approached significance (t(48) = 1.89, p = .06, d = .54). However, these age differences were not significant once accuracy in training was added as a covariate (Old: F(1, 47) = 1.5 MS = 1.26, p = .23, partial η^2 = .03; New: F(1,47) = 2.32, MS = 1.81, p = .14, partial η^2 = .05).

Formal Modeling of Cognitive Processes

Model Fits. We adopted a formal modeling approach to determine which processes underlie participants' judgments. Specifically, we fitted an exemplar model and a cue abstraction model to the responses of each individual participant (see Appendix A for mathematical formulations of the models). For both models we conducted a leave-one-out cross validation procedure (Stone, 1974) and relied on the root mean squared deviation (RMSD) between the model prediction and the participants' response as a goodness-of-fit criterion. Cross validation is a satisfactory method to deal with the problem of overfitting because it requires prediction. Complex models can often provide better fits to data compared to simpler models even when the latter are better descriptions of the process underlying the data generation because complex models have enough flexibility to fit both systematic and random variance (Roberts & Pashler, 2000; Pitt, Myung, &, Zhang, 2002). On the other hand, simpler models usually fare better in prediction because they capture the underlying process rather than unsystematic variance in the data. Cross validation works by splitting the data into a calibration set and a validation set. Model parameters are estimated by fitting the model to the calibration set, the estimated parameters are used to make predictions concerning the validation set, and these predictions are compared to participants' actual responses to obtain a measure of prediction error.

In our study, we estimated the models' free parameters by fitting the models to 15 items of the test set and then predicted the response for the 16th object based on the estimated parameter values; this was repeated for all objects. The goodness-of-fit was determined as the

RMSD between the 16 predicted model responses and the participants' responses (averaged across the responses to the two presentations of each test object).

We estimated four free parameters for the exemplar model (an attention weight *s* for each cue, constrained to add to one, and the sensitivity parameter *h*). These were estimated by a nonlinear least squares fit, assuming the training set as a knowledge base.[5] We obtained parameter estimates for the cue abstraction model in the binary condition with a nonlinear least squares fit with the parameter values of a logistic regression as the starting values. In the continuous task, we calculated parameter values analytically by running a multiple linear regression on participants' responses.

On average, both models fit adults and children rather well. In the binary task, the exemplar model fit the participants better (M = .35 *RMSD*, SD = .15) than the cue abstraction model (M = .38 *RMSD*, SD = .18; Wilcoxon Z-test = - 2.52, p = .01). In the continuous condition, the models fit judgments equally well (Exemplar model: M = 1.89 *RMSD*, SD = .58; Cue abstraction model: M = 1.91 *RMSD*, SD = .65; Z = -.10, p = .92). Table 3 reports the average model fit by environment and age group. The *RMSDs* for all individuals can be found in Appendix B.

Participant Classification. To investigate age differences in cognitive processing, we classified participants by assigning each participant to the model that had the lower RMSD, given the difference between the model fits was higher than one standard error of the mean model fits in each environment. We introduced this threshold, because in some cases both models fit a participant about equally well.[6] We excluded the participants that could not be unambiguously classified from the further analyses. In the binary condition, 6 children and 3 adults were excluded. In the continuous condition, 5 children and 5 adults were excluded. As illustrated in Figure 1, the majority of children and adults were better described by the exemplar model in the binary task. In contrast, in the continuous task, the majority of adults but not children (albeit more children than in the binary task) were classified as relying on cue abstraction. To analyze whether the cognitive processes of children and adults changed depending on the task, we ran Chi-square tests in each age group. The tests indicated that the task affected model choice for adults ($\chi^2 = 4.63$, $p = .03$), but not children ($\chi^2 = .04$, $p = .84$).

Cognitive Processing and Performance. Can differences in the choice of cognitive processes explain age differences in judgment performance? To answer this question we ran an ANOVA with accuracy for the new objects as the dependent variable and age group and model choice as independent variables in each task condition. We chose the new items, because here the strongest impact of model choice on judgment accuracy should be expected. While both models allow accurate learning of the training objects, they predict different judgments for the new objects (see Juslin, Olsson et al., 2003). In the binary task (Figure 2 left panel), neither age group nor model choice had a significant impact on accuracy (*p*s > .34). In the continuous task, however, there was a main effect of age group (F (1, 36) = 4.47, MS = 2.36, p = .04, partial η^2 = .11), and a significant interaction between age group and model choice (F (1, 36) = 6.79, MS = 3.61, p = .01, partial η^2 = .16) [7]. As illustrated in Figure 2 (right panel) the interaction suggests that children using the exemplar model performed as well as adults using the exemplar model (M_{Adults} = 3.39, SD = .35 vs. M_{Children} = 3.28, SD = .44; $t(19)$ = -.65, p = .52, d = .30), but children using a cue abstraction process performed worse than adults using a cue abstraction process (M_{Adults} = 2.62, SD = 1.15 vs. M_{Children} = 3.72, SD = .62, $t(17)$ = 2.44, p = .03, d = 1.01). Also, although difference in performance did not reach conventional levels of significance, children tendentially performed worse when relying on cue abstraction than on exemplar-based processes ($M_{\text{Cue abstraction}}$ = 3.72, SD = .62 vs. M_{Exemplar} = 3.28, SD = .44, $t(18)$ = 1.88, p = .08, d = .93) and adults tended to perform worse when using exemplar-based processes ($M_{\text{Cue abstraction}}$ = 2.62, SD = 1.15 vs. M_{Exemplar} = 3.39, SD = .35, $t(12.21)$ = 2.11, p = .06, d = .87).

Note that while children overall relied on similarity-based processes as successfully as adults they may have been less accurate in storing the exemplars in memory. Children showed on average a lower sensitivity parameter (19.63, *SD* = 14.45) than adults (38.58, SD = 17.59; t(48) = 4.16, p <.01, d = 1.18) suggesting children had fuzzier memory traces compared to adults (Nosofsky & Zaki, 1998).

Lastly, to investigate why children performed worse compared to adults when relying on the cue abstraction model in the continuous task, we analyzed age differences in the importance given to the cues as captured by the cue weight parameters from the cue abstraction model (Cooksey, 1996). The literature suggests that abstracting the weights might be a difficult task for children, resulting in an ineffective application of cue abstraction processes (Lafon et al., 2004; Montanelli, 1972). In line with this hypothesis we found that children relying on the cue abstraction model differed significantly from adults in the weight they assigned to the most important cue (M_{Adults} = 2.39, *SD* = 1.22 vs. M_{Children} = .63, *SD* = 1.17; $t(17) = 3.35$, $p < .01$, $d = 1.45$). The difference in the weight assigned to the second most important cue did not reach significance, but there was a tendency for adults to give more weight to the cue than did children (M_{Adults} = 1.90, *SD* = 1.00 vs. M_{Children} = .69, *SD* = 1.81; $t(17) = 1.87$, $p = .08$, $d = .70$). Children did not differ from adults in the weights assigned to the other cues (all *p*s > .18). The mean weights for the cues and the intercept are reported in Table 4. They indicate that the adults weighted the cues in the order of their actual importance, while children gave more weight to the less important cues. We conducted an analysis comparing the relative weight participants gave to the most important cue compared to the other three cues. More specifically we compared the mean absolute weight that adults gave to the most important cue (*M* = 2.44, *SD* = .96), with the mean absolute weight that they gave to the other three cues (*M* = 1.78, *SD* = .58; $t(10) = 2.01$, $p = .07$, $d = 1.26$) showing that although not significant adults tended to give more weight to the most important cue. Children on the other hand tended to give more weight to the other three cues (*M* = 1.56, *SD* = .58) than to the most important cue (*M* = .77, *SD* = 1.07; $t(7) = 1.43$, $p = .20$, $d = 1.26$).

To find out if the different weight that adults and children using a cue abstraction process gave to the most important cue would account for the age differences in the accuracy in the responses we conducted a two-step regression analysis. In the first step we regressed age group on the accuracy for the new items as a criterion; in the second step we added the parameter of the cue abstraction model representing the weight of the most important cue as a second predictor. This analysis showed that age group significantly predicted the accuracy in judgments in the first step (R^2= .26, β = -.51, t(17) = 2.44, p = .03). Adding the cue weight as a second predictor reduced the effect of age (β = -.24, t(16) = -.94, p = .36, see Table 5) and explained an additional 11% of the variance in accuracy (ΔR^2=.11, F(1, 16) = 2.79, p = .12). Thus, the ability to identify the best cue seems to have been a limitation in applying cue abstraction processes in the multiple-cue judgment task and can help explain age differences in how accurately the participants judged the new items. However, due to the small number of participants in this analysis the results need to be interpreted with caution.

Discussion

The results from the formal modeling of children's and adults' judgments suggest that while adults adapted their choice of cognitive processes to the tasks' characteristics, children did not. A majority of 9-11 year-olds relied on similarity-based processes to solve both the categorization and multiple-cue judgment tasks. In the categorization task, both adults' and children's judgments were best captured by an exemplar model. Furthermore, children performed equally well as adults in categorizing new test items, suggesting that similarity-based categorizations can already be mastered successfully by 9-11 year-olds. In contrast, in the multiple-cue judgment task, more adults relied successfully on cue abstraction processes, while the performance of the majority of 9-11 year-olds was better captured by an exemplar model.

Our findings resonate well with the idea that cue abstraction processes have higher cognitive demands than similarity-based judgments (DeCaro et al., 2008; Juslin, Jones et al., 2003), and that such demands can only be met later in ontogenetic time (Bunge & Zelazo, 2006; Kloos & Sloutsky, 2008). Our study also matches others showing the importance of similarity-based processes for decision making and inference in children (Boswell & Green, 1982; Krascum & Andrews, 1993; Sloutsky, 2003; Sloutsky et al., 2007). In addition, our results extend previous work by suggesting that children aged 9-11 rely on similarity-based processes in a task where adults typically rely on cue abstraction processes. Paradoxically, while adults relying on cue abstraction performed considerably better than children relying on cue abstraction to solve the multiple-cue judgment task, children using an exemplar-based process performed as well as adults using an exemplar-based process. Thus, while adults profited from "looking beyond looks", children did not.

One potential reason why children may not profit from relying on cue abstraction is that they are limited by strategy-utilization deficits (e.g., Miller, 1994; Siegler, 2000). Under this hypothesis 9-11 year-olds may be starting to rely on both, similarity-based and cue abstraction processes, but may not be able to apply cue abstraction processes as efficiently as adults. In line with this hypothesis, our results suggest that children performed worse when relying on cue abstraction processes because they had problems identifying and focusing on the most important pieces of information (Lafon et al., 2004; Mata, von Helversen, & Rieskamp, 2009; Miklich & Gillis, 1975; Montanelli, 1972), which is crucial to correctly apply the cue abstraction model.

In turn, children's failure to identify the correct cue weights may be related to learning deficits. Children had more difficulties learning to criterion in the training set and performed worse compared to adults when judging both old and new items in the test phase. Controlling for individual differences in performance at the end of the training phase reduced these age differences at test which suggests that the successful mastering of the training phase through learning was indeed an important factor in determining judgment performance. Future studies using similar training procedures could likely profit from using "rolling regression" methods (Lagnado, Newell, Kahan, & Shanks, 2006) to uncover children's learning processes and possible limitations thereof.

In this paper we took a computational modeling approach to investigate the cognitive processes underlying multiple-cue judgments and categorization in two different age groups. Our approach allowed us to distinguish between participants relying on similarity-based and cue abstraction processes and, in a next step, to investigate how reliance on these processes may determine age-related differences in performance. Computational modeling thus enabled us to go beyond directly observable behavior and examine the cognitive processes underlying estimation and categorization. As such, our results suggest that computational modeling can help uncover differences between the cognitive processes of children and adults and is a powerful method for developmental research.

One limitation of our study is that we have relied on a fairly large age range of children (9-11 years) but have not considered different age groups within our sample due to our small sample size. Most likely, there is considerable development in the use of similarity-based and cue abstraction processes from 9 to 11 years of age that we were not able to explore. A promising avenue for future research involves examining the development of categorization and inference processes from childhood to adolescence by considering separate age groups. We hope that the insights and methodology advanced here can help guide these future efforts.

In sum, 9-11 year-olds seem to rely more often on similarity-based processes compared to adults. This preference for similarity-based processing was possibly due to difficulties in applying cue abstraction processes, implying that the ability to rely on complex rule-based processes develops during adolescence.

Footnotes

1) The Realschule is type of secondary school in Germany that does not qualify for a University education.
2) In a pilot test 66 participants (Mean Age = 13.5) rated the similarity of all possible pairs of Sonics. To see if the cues were equally salient we calculated the mean similarity of all pairs of Sonics that only differed on the respective cue. The mean (*SD*) similarity for the four cues was 5.85 (.26), 5.52 (.27), 5.30 (.33) and 5.33 (.31), indicating that the cues were perceived as approximately equally salient.
3) The model predictions were generated by fitting the free parameters of the models to the training set, reflecting an error free learning of the training set. Then the estimated parameter values were used to generate model predictions for the test set.
4) We restricted training to 13 blocks because the pilot test showed that more training trials did not lead to better performance in children. This was probably due to attention problems as children could not concentrate on the task any longer and their performance decreased.
5) We also tested a version of the exemplar model with an additional free response scaling parameter γ (Nosofsky & Zaki, 2002). However, this version of the exemplar model performed worse in cross-validation. Thus, we report only the results for the exemplar model with $\gamma = 1$.
6) To check for the robustness of the classification we also considered thresholds of two and three standard errors. These classifications provided similar results, although with somewhat higher *p*-values due to the increasingly smaller sample sizes. This indicates that our findings based on the classification are robust. As a further variant we conducted a model classification only based on the new test items. This model classification was highly correlated to the classification based on the complete test set, ($r(70) = .75$). The

results from this classification indicated the same pattern, but did not reach significance, probably due to the higher variance in the participants' responses to the new test items.

7) Controlling for the accuracy during the training reduced the main effect of age group to insignificance, but did not affect the interaction.

Appendix A

Mathematical Models

Exemplar Memory

The exemplar model assumes that the judgment is the average of the criterion values, weighted by their similarity to the probe.

$$\hat{y}_p = \frac{\sum_{i=1}^{I} S(p,i) \cdot x_i}{\sum_{i=1}^{I} S(p,i)} \tag{A1}$$

where $\hat{y}_p$ is the estimated criterion value for the probe *p*; *S* is the similarity of the probe to the stored exemplars; x_i is the criterion value of the exemplar *i*; and *I* is the number of stored exemplars in memory. The similarity *S* between the stored exemplar and the probe is calculated by the similarity rule of the *generalized context model* (GCM, Nosofsky, 1984):

The similarity *S(p, i)* between exemplars is found by transforming the distance between them. The distance between a probe *p* and an exemplar *i* is

$$d_{pi} = h\left[\sum_{j=1}^{J} s_j \left| c_{pj} - c_{ij} \right|\right], \tag{A2}$$

where c_{pj} and c_{ij}, respectively, are the cue values of the probe *p* and an exemplar *i* on cue dimension *j*, *h* is a sensitivity parameter (changed from the usual *c* to avoid confusion with the cue values *c*) that reflects overall discriminability in the psychological space (Nosofsky & Zaki, 1998) and, the parameters s_j are the attention weights associated with cue dimension *j*. Attention weights vary between 0 and 1 and are constrained to sum to 1. The similarity *S(p, i)* between a probe *p* and an exemplar *i* is a nonlinearly decreasing function of their distance (d_{pi}),

$$S(p,i) = e^{-d_{pi}}, \tag{A3}$$

Cue abstraction

The cue abstraction model assumes that the judgment $\hat{y}$ of an object p is the sum of the weighted cue values $c_1..c_j$. plus an intercept k.

$$\hat{y}_p = k + \sum_{1=j}^{J} w_j \cdot c_{j,} \tag{A 4}$$

where the intercept k and the weights w are free parameters. If k = 10, w_1 = 4, w_2 = 3, w_3 = 2 and w_4 = 1, equations A4 is identical to the function determining the continuous criterion and the model produces perfect judgments.

In the binary task, we assume a decision rule assuming that all objects p with the criterion C < 15 are classified into group A and all objects with C > 15 into B and objects with C = 15 have probability of .5 in being classified into A or B. The proportion of classifications into B $p(b$ =1) was modeled by a smoother logistic function to take into account random error (c.f. Juslin, Olsson et al, 2003):

$$\hat{p}\left(\hat{b} = 1\right) = \frac{e^{k+\sum W_i \cdot c_i}}{1 - e^{k+\sum W_i \cdot c_i}}, \tag{A 5}$$

where W_i are the cue weights and k the intercept.

Appendix B

Individual Model Fits

Table B1

Individuals Models Fits (RMSD) by Age Group and Environment

		Task			
		Binary		Continuous	
Age Group	No. participant	EM	CAM	EM	CAM
Children	1	0.39	0.40	1.26	1.25
	2	0.35	0.42	2.09	1.48
	3	0.32	0.35	2.02	2.41
	4	0.43	0.57	2.51	2.44
	5	0.33	0.23	1.44	1.74
	6	0.30	0.32	2.58	3.05
	7	0.33	0.28	1.61	1.82
	8	0.40	0.30	1.91	1.26
	9	0.43	0.60	1.11	1.93
	10	0.36	0.38	2.20	2.01
	11	0.46	0.34	2.37	3.12
	12	0.43	0.45	2.37	2.31
	13	0.41	0.66	1.46	1.58
	14	0.50	0.58	2.23	1.52
	15	0.52	0.00	1.15	1.63
	16	0.26	0.36	2.06	2.06
	17	0.46	0.51	1.65	1.16
	18	0.46	0.56	1.57	1.28
	19	0.38	0.39	2.27	2.69
	20	0.44	0.46	2.53	2.70
	21	0.25	0.29	1.05	1.66
	22	0.33	0.45	0.99	1.15
	23	0.42	0.44	2.76	2.12
	24	0.43	0.35	1.84	1.57
	25	0.33	0.18	2.10	2.10
Adults	1	0.29	0.38	1.53	2.58
	2	0.00	0.00	2.86	2.84
	3	0.30	0.31	1.35	2.49
	4	0.00	0.00	1.93	2.03
	5	0.57	0.61	1.68	1.41
	6	0.13	0.45	2.19	2.51
	7	0.46	0.52	3.26	3.18
	8	0.58	0.66	2.23	1.93
	9	0.21	0.18	1.30	1.26
	10	0.53	0.71	1.54	0.97
	11	0.18	0.45	1.83	2.18
	12	0.33	0.18	2.15	2.09
	13	0.55	0.00	1.71	1.56
	14	0.44	0.52	1.63	1.51
	15	0.37	0.58	1.46	1.94

16	0.42	0.18	1.16	1.14
17	0.57	0.61	1.26	0.55
18	0.21	0.37	2.55	2.22
19	0.01	0.36	1.79	2.23
20	0.40	0.33	3.78	3.04
21	0.35	0.45	1.38	0.90
22	0.20	0.42	1.78	1.17
23	0.35	0.43	1.94	2.72
24	0.00	0.18	1.23	1.96
25	0.28	0.43	1.73	1.03

Note: EM = Exemplar model; CAM = Cue abstraction model. Participants were classified, if the difference between model fits exceeded a threshold, which was set at one standard error of the mean model fits. The threshold in the binary condition was .023, in the continuous condition .087.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442-481.

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

Boswell, D. A., & Green, H. F. (1982). The abstraction and recognition of prototypes by children and adults. *Child Development, 53*, 1028-1037.

Brehmer, B. (1994). The psychology of linear judgment models. *Acta Psychologica, 87*, 137-154.

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science, 15*, 118-121.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. San Diego, CA: Academic Press.

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

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition, 107*, 284-294.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*, 107-140.

Fisher, A. V., & Sloutsky, V. M. (2005). When induction meets memory: Evidence for gradual transition from similarity-based to category-based induction. *Child Development, 76*, 583-597.

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*, 796-804.

Gelman, S. A., & Waxman, S. R. (2007). Looking beyond looks. *Psychological Science, 18*, 554-555.

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 924-941.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition, 106*, 259-298.

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General, 132*, 133-156.

Karlsson, L., Juslin, P., & Olsson. H. (2008). Exemplar-based inference in multi-attribute decision making: Contingent, not automatic, strategy shifts? *Judgment and Decision Making, 3,* 244-260.

Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General, 137*, 52-72.

Krascum, R. M., & Andrews, S. (1993). Feature-based versus exemplar-based strategies in preschoolers' category learning. *Journal of Experimental Child Psychology, 56*, 1-48.

Kruschke, J. K. (1992). Alcove: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.

Lafon, P., Chasseigne, G., & Mullet, E. (2004). Functional learning among children, adolescents, and young adults. *Journal of Experimental Child Psychology, 88*, 334-347.

Lagnado, D. A., Newell, B., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple cue learning. *Journal of Experimental Psychology: General, 135*, 162-183.

Lehrl, S. (1999). *Mehrfachwahl-Wortschatz-Intelligenztest: Manual mit Block MWT-B*. Balingen, Germany: Spitta.

Mata, R., von Helversen, B., & Rieskamp, J. (2009). *When easy comes hard: The development of adaptive strategy selection.* Manuscript Submitted for publication.

Miklich, D. R., & Gillis, J. S. (1975). Interaction of age and cue validities in multiple-cue probability learning by children. *Psychological Reports, 37*, 235-240.

Miller, P. H. (1994). Individual differences in children's strategic behaviors: Utilization deficiencies. *Learning and Individual Differences, 6*, 285-307.

Montanelli, D. S. (1972). Multiple-cue learning in children. *Developmental Psychology, 7*, 302-312.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 104-114.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin and Review, 7*, 375-402.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101*, 53-79.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 248-255.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 924-940.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.

Pitt, M., Myung, I., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109,* 472-491.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*, 358-367.

Siegler, R. (2000). The rebirth of children's learning. *Child Development, 71*, 26-35.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7*, 246-250.

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science, 18*, 179-185.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B, 36,* 111-147.

Wilburn, C., & Feeney, A. (2008). Do development and learning really decrease memory? On similarity and category-based induction in adults and children. *Cognition, 106*, 1451-1464.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale: Revised manual (WAIS-R)*. New York: Psychological Corporation.

Authors' Note

Bettina von Helversen, Max Planck Institute for Human Development, Berlin, Germany; University of Basel, Switzerland. Rui Mata, University of Lisbon, Portugal; Henrik Olsson, Max Planck Institute for Human Development, Berlin, Germany. We would like to thank Yee lee Shing and Linnea Karlsson for their helpful comments on an earlier version of the manuscript. We are grateful to Julia Schooler for editing a draft of this manuscript. This work has been supported by a doctoral fellowship of the International Max Planck Research School LIFE and by a research grant (RI 1226/5) of the German Research Foundation to the first author. Correspondence concerning this article should be addressed to Bettina von Helversen.

Contact:

Bettina von Helversen,

University of Basel, Department of Psychology

Missionsstr. 62a, 4055 Basel

Switzerland

Phone: 0041 61 267 3528

Fax: 0041 61 267 0613

Email: Bettina.vonhelversen@unibas.ch

Table 1

Structure of the Task

Cue 1	Cue 2	Cue 3	Cue 4	Criterion continuous	Criterion binary	Training/Test
0	0	0	0	10	0	Test
0	0	0	1	11	0	Training
0	0	1	0	12	0	Test
0	0	1	1	13	0	Test
0	1	0	0	13	0	Training
0	1	0	1	14	0	Test
0	1	1	0	15	.5	Test
0	1	1	1	16	1	Test
1	0	0	0	14	0	Training
1	0	0	1	15	.5	Test
1	0	1	0	16	1	Training
1	0	1	1	17	1	Test
1	1	0	0	17	1	Test
1	1	0	1	18	1	Training
1	1	1	0	19	1	Training
1	1	1	1	20	1	Test

Note: Training items appeared during training and test. The test items only appeared during test.

Table 2

Participants Performance (Mean and SD) in Training and Test by Age Group and Task

	Age group			
	Adult's Task		Children's Task	
	Binary (*N* = 25)	Continuous (*N* = 25)	Binary (*N* = 25)	Continuous (*N* = 25)
Training	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Number of blocks	8.04 (.20)	8.88 (1.69)	8.68 (1.41)	11.32 (2.12)
RMSD: Last Training Block	.07 (.15)	.76 (1.14)	.17 (.22)	1.80 (1.32)
Test				
RMSD: Old	.14 (.19)	1.87 (1.05)	.46 (.20)	2.63 (1.02)
RMSD: New	.54 (.11)	3.04 (1.06)	.60 (.13)	3.51 (.64)
RMSD: Total	.45 (.09)	2.72 (.90)	.56 (.12)	3.25 (.63)

Note: RMSD = Root Mean Squared Deviation

Table 3

Average Model Fits (Mean RMSD and SD) by Age group and Task

	Task			
	Binary		Continuous	
	Model		Model	
	Exemplar	Cue abstraction	Exemplar	Cue abstraction
Children	.39 (.07)	.40 (.15)	1.88 (.53)	1.92 (.57)
Adults	.31 (.19)	.37 (.20)	1.89 (.65)	1.90 (.72)

Table 4

Cue Weights (Mean and SD) for the Participants classified as using the Cue Abstraction Model in the Continuous Condition

	Intercept	Cue 1	Cue 2	Cue 3	Cue 4
Children	15.00 (1.67)	.63 (1.17)	.69 (1.81)	1.25 (1.36)	-1.61 (1.35)
Adults	12.41 (2.11)	2.39 (1.11)	1.90 (1.00)	1.07 (1.73)	-.51 (1.82)

Note: $n_{\text{Children}} = 9$, $n_{\text{Adults}} = 11$.

Table 5

Regression Analysis on the Accuracy for the New Objects for Participants Classified as Using the Cue Abstraction Model in the Continuous Condition

	β	t (df)	p-value
Model 1			
Age group	-.51	-2.44 (17)	.03
Model 2			
Age group	-.24	-.93 (16)	.36
Cue weight	-.43	-1.67 (16)	.12

Note: R^2 = .37, N = 18. Cue weight denotes the parameter value for the cue with the highest weight as estimated from the cue abstraction model

Figure Captions

Figure 1. Percentage of participants best described by the exemplar model or the cue abstraction model by age group and task. The results for children are reported in the left panel, the results for adults in the right panel.

Figure 2. Accuracy on the new objects in the test set by model choice and age group. The results for the binary condition are shown in the left panel, the results for the continuous condition in the right panel.

Figure 1

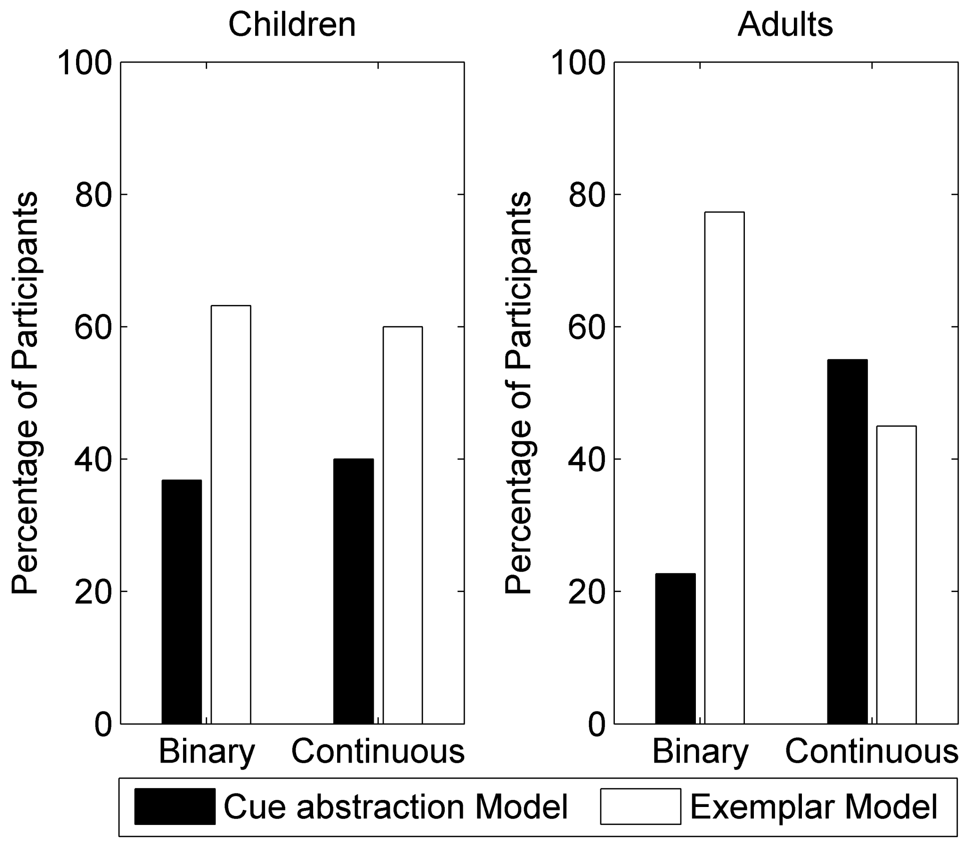
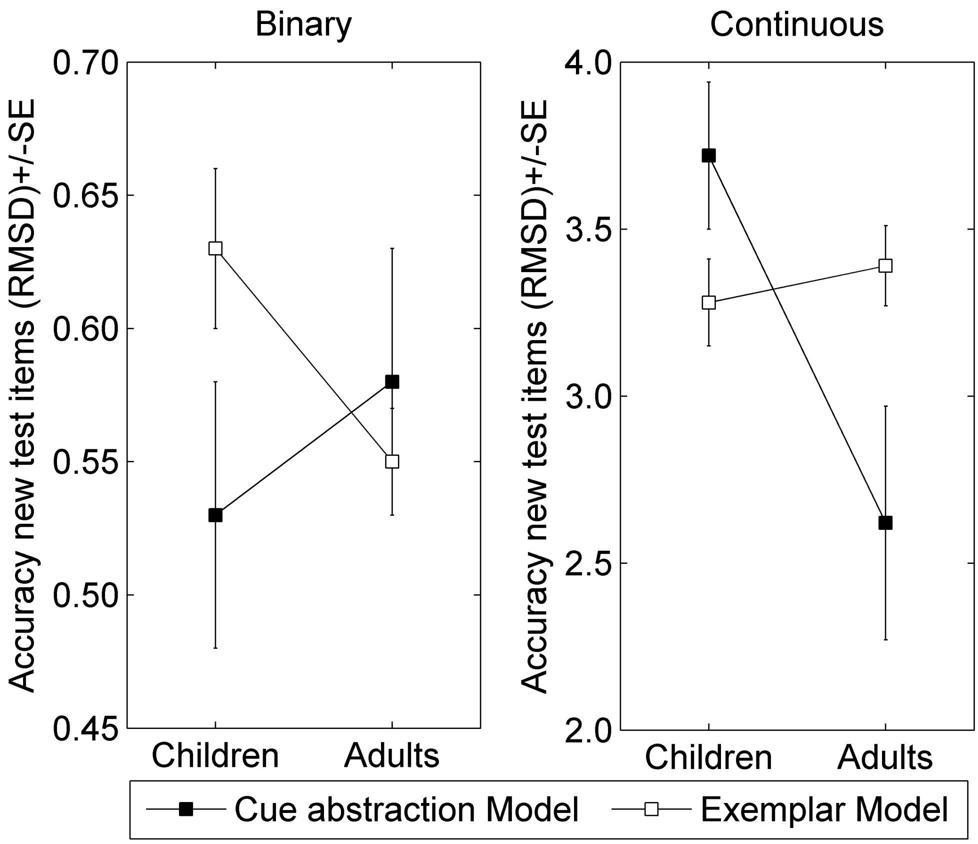

Children
Adults
Percentage of Participants
100
80
60
40
20
0
Binary
Continuous
Cue abstraction Model
Exemplar Model

Figure 2


Binary
Continuous
Accuracy new test items (RMSD)+/-SE
0.70
0.65
0.60
0.55
0.50
0.45
4.0
3.5
3.0
2.5
2.0
Children
Adults
Cue abstraction Model
Exemplar Model