Mata, Rui ; von Helversen, Bettina ; Karlsson, Linnea ; Cüpper, Lutz

# Adult age differences in categorization and multiple-cue judgment

Journal Article as: peer-reviewed accepted version (Postprint)

DOI of this document*(secondary publication): https://doi.org/10.26092/elib/3448

Publication date of this document: 08/11/2024

* for better findability or for reliable citation

Recommended Citation (primary publication/Version of Record) incl. DOI:

Mata, R., von Helversen, B., Karlsson, L., & Cüpper, L. (2012). Adult age differences in categorization and multiple-cue judgment. Developmental Psychology, 48(4), 1188–1201. https://doi.org/10.1037/a0026084

# Adult Age Differences in Categorization and Multiple-Cue Judgment

Rui Mata
University of Basel

Bettina von Helversen
University of Basel
and Max Planck Institute for Human Development

Linnea Karlsson
Max Planck Institute for Human Development
and Umeå University

Lutz Cüpper
RWTH Aachen University

We often need to infer unknown properties of objects from observable ones, just like detectives must infer guilt from observable clues and behavior. But how do inferential processes change with age? We examined young and older adults' reliance on rule-based and similarity-based processes in an inference task that can be considered either a categorization or a multiple-cue judgment task, depending on the nature of the criterion (binary vs. continuous). Both older and young adults relied on rule-based processes in the multiple-cue judgment task. In the categorization task, however, the majority of older adults relied on rule-based processes while young adults preferred similarity-based processes. Moreover, older adults who relied on rule-based processes performed poorly compared with young adults who relied on the same process, suggesting that aging is associated with deficits in applying rule-based processes.

*Keywords:* categorization, multiple-cue judgment, aging, exemplar model, recognition memory

Agatha Christie's fictional characters, Hercule Poirot and Jane Marple, solved mysteries in different ways. Poirot followed deductive logic, methodically inferring from the available cues the culprits of various crimes (Christie, 1934). Miss Marple would instead gauge the similarity between current suspects to numerous acquaintances from her hometown, St. Mary Mead, to uncover motivations and ultimately determine the probability that someone committed a crime (Christie, 1957). These two investigative methods have clear resemblance to psychological theories that postulate rule-based inference processes involving cue integration (e.g., Einhorn, Kleinmuntz, & Kleinmuntz, 1979), similarity-based processes that rely on exemplar retrieval from memory (e.g., Medin & Schaffer, 1978; Nosofsky & Johansen, 2000), or both (e.g., Erickson & Kruschke, 1998). But which of these characters and theories best portrays older adults' inferences? Inference problems are ubiquitous in everyday life, and so it is crucial to evaluate whether age-related cognitive decline can limit older adults' abilities in this regard (Mata, Schooler, & Rieskamp, 2007; Thornton & Dumke, 2005). In the current article, we investigate adult age differences in the reliance on rule-based versus similarity-based processes in specific inference problems, namely, in categorization and multiple-cue judgment tasks.

Categorization and multiple-cue judgment problems represent structurally similar and common inference problems that differ in one respect—the type of criterion to be judged. Teachers can be asked to estimate whether a particular student will score above or below average on a test or, instead, to estimate the student's specific score. Likewise, a doctor may be asked to make a judgment about whether a patient will live or die or, instead, assign a specific probability of his or her recovery. Both problems involve integrating known characteristics (cues) of an object to infer some other unknown property, either a category (categorization) or a continuous criterion (multiple-cue judgment).

Juslin, Olsson, and Olsson (2003) demonstrated that the type of problem—categorization or multiple-cue judgment—can to a large extent determine the reliance on rule-based versus similarity-based processes in young adults. Specifically, when a task consisted of estimating a binary criterion (categorization), young adults primarily relied on exemplar memory, but if the criterion to be estimated was continuous (multiple-cue judgment), participants primarily relied on rule-based processes. Juslin et al. interpreted these results as showing that explicit representations of cue–criterion relations are more likely to be extracted when plenty of information is present and linear cue-criterion relations can be extracted, as in the case of the continuous feedback available in multiple-cue judgment tasks. In turn, people often default to similarity-based processes when information about the cue-criterion relations is nonlinear or scarce, such as in categorization tasks. A number of studies have since found patterns of results congruent with this view (e.g., von Helversen, Mata, & Olsson, 2010; Juslin, Karlsson, & Olsson, 2008; Karlsson, Juslin, & Olsson, 2007).

Rui Mata, Institute for Psychology, University of Basel, Basel, Switzerland; Bettina von Helversen, Institute for Psychology, University of Basel and Max Planck Institute for Human Development, Berlin, Germany; Linnea Karlsson, Max Planck Institute for Human Development, and Department of Integrative Medical Biology, Umeå University, Umeå, Sweden; Lutz Cüpper, Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, Medical Faculty, RWTH Aachen University, Aachen, Germany.

The authors would like to thank Jutta Mata for helpful discussions and Gregor Caregnato for data collection.

Correspondence concerning this article should be addressed to Rui Mata, Fakultät für Psychologie, Missionsstrasse 64A, Basel 4055, Switzerland. E-mail: rui.mata@unibas.ch

The ability to rely on similarity-based versus rule-based processes seems to change across the life span: children between the ages of 9 and 11 years appear not to select between similarity-based and rule-based processes as a function of the type of problem (von Helversen, et al., 2010). Von Helversen et al. demonstrated that young adults relied on rule–based processes in a multiple-cue judgment task but selected more often a similarity-based process in the categorization task. In contrast, the majority of children relied on similarity-based exemplar processes regardless of the type of problem. What cognitive processes do older adults favor when making categorization and multiple-cue judgments?

There are two possible—but not necessarily mutually exclusive—scenarios concerning how aging may impact older adults' preferences for rule-based versus similarity-based processes. First, given von Helversen et al.'s (2010) results suggesting that exemplar memory processes are less cognitively demanding, one could expect older adults to show an increased reliance on similarity-based processes due to reduced cognitive capacity. To the extent that exemplar-based categorization makes use of implicit or automatic processes (Hahn & Chater, 1998; Koenig et al. 2005; but see Juslin et al., 2008) that are less affected by age-related cognitive decline compared with explicit memory processes such as recognition or recall (Fleischman, Wilson, Gabrieli, Bienias, & Bennett, 2004; D. V. Howard, Howard, Dennis, LaVine, & Valentino, 2008; Jennings & Jacoby, 1993; but see also J. H. Howard, Howard, Dennis, & Kelly, 2008), older adults could be well advised to rely on the former. In this case, age-related cognitive limitations, such as deficits in processing speed or working memory, which may negatively affect execution of rule-based strategies, should be related to an increased selection of similarity-based over rule-based inference processes.

Second, age-related cognitive decline may lead to increased reliance on rule-based processes. Filoteo and Maddox (2004) used formal modeling to show that older adults may rely more often on simple rules in a perceptual categorization task and that smaller age differences in performance are observed for rule-based compared with implicit similarity-based information-integration processes (see Ashby, Noble, Filoteo, Waldron, & Ell, 2003, for a similar result). Filoteo and Maddox speculated that older adults may rely more on rule-based processes due to difficulties in memorizing visual patterns. According to this view, older adults may prefer to rely on rule-based strategies to circumvent real or perceived memory limitations. There is indeed some evidence that older adults may generally prefer cognitive strategies that do not involve high memory demands (Touron & Hertzog, 2004, 2009).

Older adults' preference for rule-based strategies does not, however, imply perfect inference performance. For example, one study on aging and category learning suggests that older adults perform similarly to young adults when category membership can be easily verbalized using a simple rule, but older adults perform considerably more poorly when more complex rules need to be carried out to arrive at a judgment (Racine, Barch, Braver, & Noelle, 2006). Research on aging and probabilistic cue learning also suggests that older adults have difficulties learning the value of cues when many competing cues are present (e.g., Chasseigne et al., 2004). Consequently, even if older adults tend to rely on rule-based processes, they may have difficulties determining the value of cues from experience. This view resonates well with the findings that smaller working memory capacity is associated with learning impairments in category learning (DeCaro, Carlson, Thomas, & Beilock, 2009) and that age-related cognitive decline may lead to learning difficulties and execution of complex inference strategies (Mata et al., 2007; Mata, von Helversen, & Rieskamp, 2010).

## The Current Study

In the current study, we examined young and older adults' inferences 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). In this manner, we hoped to contrast the two views described, which suggest age-related differences in the preference for rule- versus similarity-based processes. In addition, we hoped to contribute to understanding the reasons underlying potential age differences in preferences for particular cognitive processes by assessing young and older adults' cognitive abilities and memory for task materials. To the extent that similarity-based processes are simpler or less affected by age-related cognitive decline (e.g. D. V. Howard et al. 2008; J. H. Howard et al., 2008), individual differences in cognitive abilities could be related to older adults' preferred selection of similarity- over rule-based inference processes. However, to the extent that similarity-based processes rely on memory processes that decline with age, individual differences in recognition performance could prove a predictor concerning older adults' preference for rule-based processes as suggested by Filoteo and Maddox (2004). We examined these possibilities by assessing participants' fluid cognitive abilities with two standardized cognitive tests as well as their memory for task stimuli by examining discrimination of trained exemplars and lures in a recognition task. In this manner, we hoped to assess whether participants' memory and other abilities played a role in their choice of inference processes.

## Method

### Participants

Forty-eight young adults (50% female) and 50 older adults (50% female) participated in the study (see Table 1 for participant characteristics). Most young adults were university students, and older participants were community-dwelling adults recruited through newspaper advertisements. The majority of participants were White. Participation took an average of 1.5 hr. Participants received a performance-contingent payment ($M$ = 15 euros, 5 euros as a show-up fee).

### Design and Material

The design included two between-subjects factors, task (binary vs. continuous criterion) and age group (young vs. older). We modeled our task on von Helversen et al. (2010; for similar tasks see also Juslin et al., 2003, 2008; Karlsson et al., 2007). The participants' goal was to learn how well fictitious cartoon characters performed in a game in which each needed to catch as many "golbis" as possible. In the binary task, participants needed to classify each character as a successful or an unsuccessful hunter. In the continuous task, they needed to estimate how many golbis a character had caught.

Table 1
*Participant Characteristics*

| Variable | Young adults ($N$ = 48) | | Older adults ($N$ = 50) | | | |
|---|---|---|---|---|---|---|
| | $M$ | $SD$ | $M$ | $SD$ | $t$ | $p$ |
| Age | 24.5 | 2.6 | 68.3 | 4.2 | | |
| Education (years)[a] | 16.4 | 2.9 | 16.1 | 5.1 | 0.45 | .65 |
| Vocabulary[b] | 30.8 | 5.6 | 33.7 | 2.2 | 3.44 | .001 |
| Processing speed[a] | 63.8 | 13.1 | 43.9 | 10.8 | 8.22 | <.001 |
| CRT[b] | 1.83 | 1.11 | 0.80 | 0.95 | 4.93 | <.001 |
| Recognition ($d'$) | 0.90 | 0.70 | .46 | 0.66 | 3.17 | .002 |
| Recognition (β) | .66 | .16 | .91 | .50 | 3.31 | .001 |

*Note*. Vocabulary = Spot-a-Word (score range 0–35; Lehrl, 1999); processing speed = Digit–Symbol Substitution (score range 0–93; Wechsler, 1981); CRT = Cognitive Reflection Test (score range 1–3;Frederick, 2005).
[a] $df$ = 96. [b] $df$ = 95.

The task consisted of a rather strict training phase and a test phase. In the training phase, a training set consisting of 10 of the 16 cartoon characters was repeatedly presented (see Table 2). The rationale for not including all items in the training set is that when all items are learned during the training phase, the cue abstraction and similarity-based models are indistinguishable (see Juslin et al., 2003, for a full rationale for excluding items from the training set). Excluding items from the training set ensures that the cue abstraction and exemplar models can in principle make different predictions for new test items: The effects of similarity can best be observed when judgments of new items are biased by the occurrence of similar items in memory. The characters used in the training set were selected so as to generate differences between the predictions of the exemplar and cue abstraction model for the new test items (see Appendix A for details). In the test phase, all 16 characters were presented four times, thus amounting to a total of 64 judgments.

The characters varied on four cues (hair, nose, ears, and belly), which could be used to predict how well they performed in the game. Each cue had one of two possible features, for example, the belly was either green or blue, and the hair had spikes or dread locks. The number of golbis a character caught varied between 10 and 20 and was determined as a linear function of the cues:

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

where $C$ is the criterion in the continuous task and $c_1$ through $c_4$ are the cue values, which could be either 1 or 0 (see Table 2). Equation 1 represents a linear function of the sort used in previous studies on estimation and categorization processes with similar experimental designs to those used here (e.g., von Helversen et al.,

Table 2
*Task Structure: Cue Values for Training and Test Characters*

| 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 | Training |
| 0 | 0 | 1 | 1 | 13 | 0 | Training |
| 0 | 1 | 0 | 0 | 13 | 0 | Training |
| 0 | 1 | 0 | 1 | 14 | 0 | Training |
| 0 | 1 | 1 | 0 | 15 | .5 | Test |
| 0 | 1 | 1 | 1 | 16 | 1 | Training |
| 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 | Training |
| 1 | 1 | 0 | 0 | 17 | 1 | Test |
| 1 | 1 | 0 | 1 | 18 | 1 | Test |
| 1 | 1 | 1 | 0 | 19 | 1 | Test |
| 1 | 1 | 1 | 1 | 20 | 1 | Training |

*Note.* Training characters appeared during training and test. Test characters appeared only during test. Objects with a continuous criterion = 15 have probability of .5 of being classified into the successful or unsuccessful group.

2010; Juslin et al., 2003). In the binary task, each character was categorized as successful or unsuccessful on the basis of the continuous criterion; characters with criterion values above 15 were classified as successful and those with criterion values below 15 were classified as unsuccessful. Please note that Equation 1 implies that some items have similar continuous criterion values despite having different cue profiles (see Table 2). The assignment of cue weights to the four pictorial cues (hair, nose, ears, and belly), as well of the cue values (positive vs. negative) to the features (spiky hair vs. dread locks), were randomly determined and varied across participants. For example, while for one participant, hair could be the most important cue (4), followed by nose (3), ears (2), and belly (1), another participant could experience a different ranking: belly (4), ears (3), nose (2), and hair (1). Also, the first participant may have learned that a spiky hair, relative to dread locks, was associated with more golbis being caught, while the opposite was true for another participant.

To facilitate learning for older adults who as a group seem to have difficulties in learning from probabilistic feedback (e.g., Mata et al., 2010), we excluded items that could not be deterministically classified (criterion = 15) from the training set (see Table 2). In addition, we wanted to ensure that young adults would not over-learn the training stimuli relative to older adults whom we expected would require more time to learn the training stimuli. Consequently, training was terminated after the eighth block if a rather strict accuracy criterion was reached. The accuracy criterion was reached if the root-mean-square deviation (RMSD) between participants' responses and the criterion values in one block sank below 0.5 in the binary task (corresponding to a correct classification of eight out of 10) and below 1.5 in the continuous task. If the accuracy criterion was not met in Training Blocks 8–16, training terminated after the 16th block (which amounted to about 1 hr of training in the continuous condition, which we had determined from previous testing should be the upper time limit to ensure participants' remained engaged in the task).

Regarding test performance, participants' judgments for testing items in the two task conditions (binary vs. continuous) were evaluated in respect to the criterion given by Equation 1. Note that while the values used were arbitrarily defined because they correspond to an artificial task structure, the use of Equation 1 allowed us to assess accuracy at training by setting a criterion for novel stimuli that followed the same structure as that experienced by participants during training. Finally, 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 their answer deviated by 1, and 0 points if their answer deviated by more than 1. At the end of the experiment, points were converted into euro with an exchange rate of 1 euro for every 100 points.

## Procedure

The study began with the extensive training phase consisting of between eight and 16 blocks. In each block, the training characters were presented in random order. In each trial of the training phase, participants were asked to judge the performance of a training character. After giving their response, they received feedback about their performance, the correct criterion value, and the points they earned. During the test phase, participants were asked to make a categorization or give a continuous estimate regarding each character but did not receive feedback. They were also asked to indicate if they recognized the character from the training set or not. Afterwards, participants completed processing speed, cognitive reflection, and vocabulary measures (see Table 1) and a number of additional measures that are not the focus of this article. The cognitive reflection test is a three-item measure (e.g. "A bat and a ball cost $1.10 in total; the bat costs $1.00 more than the ball; how much does the ball cost?") and is thought to measure one's ability to engage in effortful inference processes and avoid judgment biases (Frederick, 2005; Oechssler, Roider, & Schmitz, 2009).

# Results

We first provide an overview of participants' training and test performance. We then provide an account of the cognitive processes underlying categorization and estimation judgments of young and older adults. Finally, we report participants' memory for task materials and cognitive abilities and relate these to choice of cognitive process in young and older adults.

## Training and Test Performance

Table 3 reports performance in training and test by task and age group as the RMSD between participants' responses to an item and the item's criterion. RMSD is a common measure of judgment accuracy and model fit (see later section) in estimation and categorization tasks (von Helversen et al., 2010; Juslin et al., 2003; Karlsson et al., 2007). In the continuous condition, the RMSD preserves the scale used (from 10 to 20) such that values can be interpreted as a measure of average deviation from the actual criterion values. To facilitate interpretation of RMSD in the binary criterion, we provide the following guidelines: in the training

Table 3
*Performance (Root-Mean-Square Deviation) in Training and Test by Age Group and Task Condition*

| | Young adults | | | | Older adults | | | |
|---|---|---|---|---|---|---|---|---|
| | Binary ($n$ = 23) | | Continuous ($n$ = 25) | | Binary ($n$ = 25) | | Continuous ($n$ = 25) | |
| Performance | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ |
| Training (last block) | 0.24 | 0.22 | 1.45 | 0.73 | 0.49 | 0.21 | 2.50 | 1.06 |
| Test | 0.36 | 0.10 | 1.65 | 0.63 | 0.47 | 0.10 | 2.35 | 0.46 |

phase, a correct classification of nine out of 10, eight out of 10, and seven out of 10 correspond to a RMSD of .32, .45, and .55, respectively; in the test phase, a correct classification of 15 out of 16, 14 out of 16, and 13 out of 16 correspond to a RMSD of .25, .35, and .43, respectively. Please note that due to the different nature of the criterion in the two task conditions (continuous: from 10 to 20 vs. binary: 0 or 1), it is not possible to meaningfully compare RMSD between the two conditions. For this reason, we report separate analyses for the continuous and binary conditions whenever using RMSD.

We conducted analyses of variance (ANOVAs) with judgment accuracy (RMSD) in the final training block as the dependent variable and age group as the independent variable. We found effects of age group in both the continuous, $F(1, 48) = 16.64$, $p < .001$, $\eta_p^2 = .26$, and binary tasks, $F(1, 46) = 16.85$, $p < .001$, $\eta_p^2 = .27$. Because one cannot use RMSD to compare training performance in the two task conditions, we analyzed the frequency with which participants reached the accuracy criterion during training. In the continuous condition, significantly more young adults (21 out of 25) than older adults (eight out of 25) reached the learning criterion, $\chi^2(1) = 11.82$, $p < .001$. In the binary condition, all young adults (23) and two thirds of older adults (17 out of 25) reached the criterion, $\chi^2(1) = 6.68$, $p < .01$. Moreover, the results suggest that while young adults were equally likely to reach the learning criterion in the two task conditions (21 vs. 23), $\chi^2(1) = 2.91$, $p = .14$, older adults were less likely to reach criterion in the continuous relative to the binary condition (eight vs. 17), $\chi^2(1) = 5.12$, $p = .02$. Overall, these results suggest that older adults had more learning difficulties than did young adults and that older adults had more learning difficulties in the continuous than in the binary task condition.

Concerning test performance, we assessed age differences in judgment at test by conducting ANOVAs with judgment accuracy in the test phase (RMSD) as the dependent variable and age group as the independent variable separately for the continuous and binary conditions. In addition, we conducted these analyses while controlling for training performance by adding performance in the last training block as a covariate in the ANOVA with the goal of assessing whether age differences in test performance can be explained by individual differences in learning abilities. In the continuous condition, we found a significant effect of age group, $F(1, 48) = 19.55$, $p < .001$, $\eta_p^2 = .29$. This effect was reduced but remained significant when controlling for accuracy in the last training block, $F(1, 47) = 4.18$, $p = .05$, $\eta_p^2 = .08$. In the binary condition, we also found a significant effect of age group, $F(1, 46) = 14.86$, $p < .001$, $\eta_p^2 = .24$, but this effect was no longer significant when controlling for accuracy in the last training block, $F(1, 45) = 1.83$, $p = .18$, $\eta_p^2 = .04$. Overall, the results suggest that older adults had significant difficulties making categorization and estimation judgments compared with young adults. However, age differences in performance at test seem partly due to differences in training performance. In the following, we turn to identifying the cognitive processes underlying young and older adults' judgments.

## Formal Modeling of Cognitive Processes

Our goal was to test the hypothesis that young and older adults differ in their preferences for different inference strategies. Consequently, we first used formal models to identify the processes underlying participants' judgments and then classified participants so as to be able to compare young and older adults' preferences for specific cognitive processes.

**Model fits.** We fitted an exemplar model and a cue abstraction model to the responses of each individual participant; these models have been shown to capture well participants' judgments in estimation and categorization tasks (e.g., von Helversen et al., 2010; Juslin et al., 2003; Karlsson et al., 2007; see Appendix A for mathematical formulations of the models). We conducted a leave-one-out cross validation procedure and relied on the RMSD between the model prediction and the participants' response as a goodness-of-fit criterion. Specifically, we estimated the models' free parameters for each individual participant by fitting the models to 15 items of the test set and then predicting the response for the 16th object on the basis of the estimated parameter values. This process was repeated for all objects. The goodness of fit was determined as the RMSD between the 16 predicted model responses and the participant's responses (averaged across the responses to the four 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 1, and the sensitivity parameter $h$) using a nonlinear least squares fit, assuming the training set as a knowledge base. We obtained parameter estimates for the cue abstraction model in the binary condition with a nonlinear least squares (iterative) procedure 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. Table 4 reports the average model fit by environment and age group. Individual model fits can be found in Appendix B.

**Participant classification.** 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 fit of the two models (cue abstraction

Table 4
*Model Fits (Root-Mean-Square Deviation) by Age Group and Task Condition*

| | Young adults | | | | Older adults | | | |
|---|---|---|---|---|---|---|---|---|
| | Binary ($n$ = 23) | | Continuous ($n$ = 25) | | Binary ($n$ = 25) | | Continuous ($n$ = 25) | |
| Performance | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Exemplar-based model | 0.19 | 0.13 | 1.22 | 0.27 | 0.30 | 0.12 | 1.59 | 0.50 |
| Cue abstraction model | 0.21 | 0.12 | 1.11 | 0.50 | 0.19 | 0.11 | 1.47 | 0.57 |

model [CAM] and exemplar-based model [EBM]) in each condition (binary vs. continuous). We introduced this threshold because in some cases, both models fit a participant about equally well (cf. von Helversen et al., 2010). We excluded those participants that could not be unambiguously classified from further analyses. In the continuous condition, four young adults and seven older adults were excluded. In the binary condition, two young adults and four older adults were excluded (for individual classifications, see Appendix B). The proportions of participants classified as EBM and CAM users are presented in Figure 1.

We conducted a logistic regression with participant classification (EBM vs. CAM) as the dependent variable, and age group, task condition (continuous vs. binary), and their interaction as independent variables, Nagelkerke's $R^2 = .17$, $p = .01$. The logistic regression revealed an interaction between age group and condition, $\exp(B) = .12$, $p = .04$, but no effect of age group, $\exp(B) = .85$, $p = .80$, or condition, $\exp(B) = 2.17$, $p = .22$. As can be seen in Figure 1, young and older adults seem to have relied to a similar extent on cue abstraction processes in the continuous condition. However, in the binary condition, older adults preferred cue abstraction processes to a larger extent than young adults did. We also conducted the same analysis including performance in the last training block as an independent variable to test whether training performance had a significant impact on strategy choice, but performance in the last training block was not significantly associated with strategy choice, $\exp(B) = .99$, $p = .97$. In sum, while young adults were more likely to rely on similarity-based processes in the categorization relative to the multiple-cue judgment task, the majority of older adults relied on rule-based inference processes when making both categorization and multiple-cue judgments.

**Participant classification and judgment performance.** The test phase included judgments about "old" characters presented at training and "new" characters that had not yet been presented (see Table 2). In the following, we focus on new characters because the CAM and EBM differ more markedly for new characters (see Appendix B; Juslin et al., 2003; von Helversen et al., 2010). Our results suggest that older participants performed overall worse relative to younger adults (see Figure 2). We conducted ANOVAs

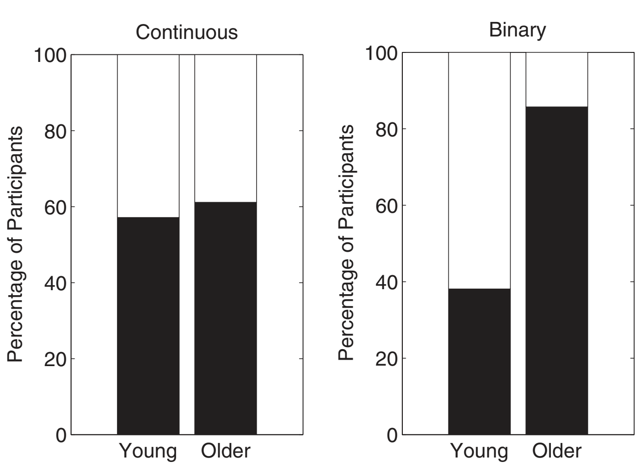


*Figure 1.* Percentage of participants best described by the exemplar-based model (EBM; white) and the cue abstraction model (CAM; black) as a function of task conditions (continuous, binary).

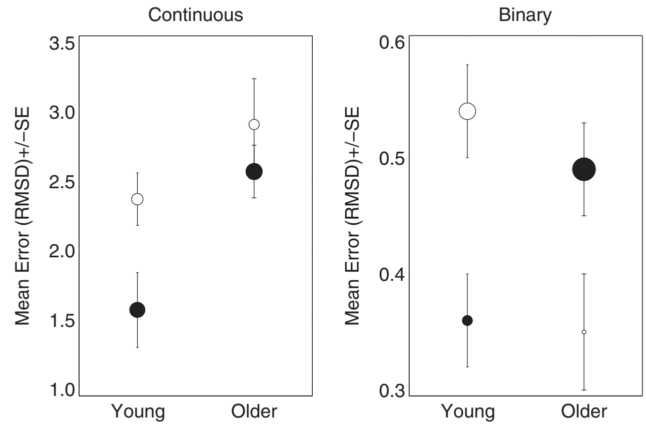


*Figure 2.* Root-mean-square deviation (RMSD) for new test items for young and older adults classified as using an exemplar-based model (EBM; white) or cue abstraction model (CAM; black) as a function of task conditions (continuous, binary). Circle diameters are proportional to the percentage of participants classified as CAM or EBM (see also Figure 1). Error bars represent standard errors of the mean.

with judgment performance for new characters as dependent variable and age group (young, older) and participant classification (EBM, CAM) as independent variables. In a second step, we also included performance in the last training block as an independent variable with the goal of assessing whether potential age differences at test may be due to individual differences in learning.

In the continuous condition, we found an effect of age group, $F(1, 35) = 9.23$, $p = .004$, $\eta^2_p = .21$, and model, $F(1, 35) = 5.17$, $p = .03$, $\eta^2_p = .13$, but not an age group by model interaction, $F(1, 35) = 0.84$, $p = .37$, $\eta^2_p = .02$. We further assessed age differences in performance only for those participants classified as CAM (the mode for both young and older adults) and found a significant effect of age group, $F(1, 21) = 8.90$, $p = .007$, $\eta^2_p = .30$. Consequently, these results suggest that age-related differences in judgment performance cannot simply be attributed to model selection but that older adults had more difficulties applying cue abstraction processes. We further controlled for performance in the last training block and found that age was no longer significant, $F(1, 20) = 2.20$, $p = .15$, $\eta^2_p = .10$, suggesting that age differences in applying the CAM model were related to individual differences in learning.

In the binary condition, we found no main effects of age group, $F(1, 38) = 0.32$, $p = .57$, $\eta^2_p < .01$, nor model, $F(1, 38) = 0.16$, $p = .91$, $\eta^2_p < .01$, but an age group by model interaction emerged, $F(1, 38) = 8.74$, $p = .005$, $\eta^2_p = .19$. Note, however, that only three older adults were classified as using the exemplar model in the binary condition. Consequently, we further assessed age differences in performance only for those participants classified as CAM (the mode for older adults) and found a significant effect of age, $F(1, 24) = 4.56$, $p = .04$, $\eta^2_p = .16$, which remained significant after controlling for performance in the last training block, $F(1, 23) = 8.76$, $p = .007$, $\eta^2_p = .28$. These results show that older CAM users in the binary condition made poorer judgments than young CAM users, even when the effect of training performance was controlled.

## Cognitive Abilities and Choice of Cognitive Process

One of our goals in the study was to determine whether individual differences in cognitive ability played a role in participants' preferences for rule-based versus similarity-based inference processes. For this purpose, we conducted standardized measures of cognitive ability (processing speed, cognitive reflection) as well as task-related recognition test. We measured recognition memory for the task stimuli by asking participants at each object presentation during the test phase whether they recognized the object from the training phase. We calculated hit and false alarm rates and the respective signal detection parameters, discriminability ($d'$), and bias ($\beta$) for each participant (see Appendix C). As can be seen in Table 1, young adults scored higher than older adults on the cognitive reflection and processing speed measures. Also, older adults had more difficulties discriminating between new and old items in the recognition test relative to young adults. Could difficulties in discriminating new and old characters or other cognitive limitations be responsible for reliance on rule- versus similarity-based processes?

We assessed the link among processing speed, cognitive reflection, recognition memory ($d'$), and participants' classification using logistic regression. To avoid biasing our results due to mean age group differences in the cognitive ability measures (cf. Hofer & Sliwinski, 2001), we conducted analyses separately for each age group. In the following analyses, we collapsed across conditions so as to increase power but included a dummy variable to control for task effects (similar results are obtained when task conditions are analyzed separately). We found no link between individual differences in the cognitive measures and strategy choice for young adults, $d'$, $\exp(B) = .26$, $p = .52$; cognitive reflection, $\exp(B) = 1.48$, $p = .23$; processing speed, $\exp(B) = 1.01$, $p = .67$; or task condition, $\exp(B) = 1.97$, $p = .32$. We also found no significant link between individual differences in cognitive measures and strategy choice for the sample of older adults, $d'$, $\exp(B) = .30$, $p = .12$; cognitive reflection, $\exp(B) = 1.03$, $p = .95$; processing speed, $\exp(B) = .94$, $p = .18$; or task condition, $\exp(B) = 0.97$, $p = .17$. Future studies with larger sample sizes and a more comprehensive battery of tests may be needed to better determine the relation of individual differences in cognitive abilities to the selection of rule-based versus similarity-based inference processes. Overall, the results do not provide support for the idea that older participants relied on cue abstraction processes due to limitations in recognition memory or other cognitive abilities.

# Discussion

People often need to infer unknown properties of objects from observable ones, just like detectives must infer guilt from observable clues and behavior. We asked whether older adults were more likely to rely on rule-based or similarity-based processes when making inferences. Specifically, young and older adults made inferences in a task that can either be considered a multiple-cue judgment task or a categorization task depending on whether the criterion is continuous or binary, respectively. In the multiple-cue judgment task, in which participants had to estimate a continuous criterion, the majority of both young and older adults were best fit by a rule-based model that assumes that participants identify the predictive value of each relevant cue and linearly integrate cues to arrive at judgments (CAM). More remarkably, in the categorization task, the majority of older adults were best fit by a CAM, while the majority of younger adults were best fit by an EBM that assumes that participants make judgments on the basis of similarity between the item to be categorized and retrieved exemplars from memory. In sum, while young adults relied on different processes as a function of the task, older adults preferred cue abstraction processes in both multiple-cue judgment and categorization tasks.

Overall, the finding that older adults relied more on rule-based processes relative to young adults in at least one task condition undermines the suggestion that age-related deficits in cognitive abilities, such as working memory or processing speed, foster implicit similarity-based inference. In turn, the idea that older adults rely more on rule-based processes because of memory limitations (Filoteo & Maddox, 2004) also receives limited support: Older adults tended to rely on rule-based strategies, but we did not find a link between individual differences in cognitive ability, including recognition memory, and participants' reliance on rule-based versus similarity-based processes.

Our findings raise the possibility that older adults' increased reliance on cue abstraction processes relative to young adults was not due to objective cognitive limitations. Previous work on strategy selection and aging has found that subjective factors may play an important role in strategy selection (Touron & Hertzog, 2004, 2009): Older adults' are often reluctant to select strategies that rely on memory due to lack of confidence in these abilities. Future studies should evaluate the role that confidence in memory abilities plays in individuals' choosing to use specific inference strategies, such as those that purportedly rely on memory (Juslin et al., 2003). In sum, while our study is limited in its ability to assess the specific mechanism underlying age differences in inference, one promising avenue for future work is to study how meta-cognitive factors, rather than objective cognitive limitations, impact strategy choice in the inference domain.

Older adults had more difficulties reaching the arbitrary learning criteria in the continuous relative to the binary task, while young adults were equally likely to reach criterion in the two tasks. This may indicate that, at least for older adults, the continuous (multiple-cue judgment) task is more demanding than the binary (categorization) task. Nonetheless, the differences between young and older adults' choice of rule-based strategies were only evident in the latter (categorization task). Consequently, one may infer that the age differences in performance in the continuous (multiple-cue judgment) task are not simply due to reliance on different judgment processes. Also, even when only those participants who relied on the rule-based strategy are considered, older adults showed poorer judgment performance in both the multiple-cue judgment and categorization tasks relative to their young counterparts. These results are congruent with the view that aging leads to deficits in strategy execution (e.g., Dunlosky & Hertzog, 1998). Nevertheless, because these age differences were no longer significant or reduced when controlling for performance in the last block of training, it is reasonable to assume that age differences at test were at least partly related to learning deficits, for example, learning cue polarity (the direction of the correlation between a cue and the criterion) and cue weights (the strength of the correlation between a cue and the criterion). Such findings match previous ones showing that aging is associated with deficits in learning the value of cues (e.g., Chasseigne et al., 2004), decision options (Wood, Busemeyer, Coling, Cox, & Davis, 2005), and strategies (Mata et al., 2010).

Our study did not allow us to test whether participants were adaptive in the sense of selecting the best performing strategy given their cognitive abilities. For this purpose, studies could rely on a "no-choice-of-strategy" procedure (Lemaire & Siegler, 1995) in which experimenters instruct participants to apply cue abstraction and exemplar-based processes on specific trials. The application efficacy with which participants use each strategy could then be linked to participants' choice of cognitive process in an independent task. While it is thought that older adults often adopt such adaptive and compensatory behaviors (Baltes, 1997), a thorough test of this premise in the inference domain is still lacking. In addition, it could be informative to test whether age differences in strategy choice can also be found in other domains in which no age differences or even superior performance for older adults have been reported, such as social and consumer judgments in which older adults have significant experience (e.g., Blanchard-Fields, 2007; Hess, Leclerc, Swaim, & Weatherbee, 2009).

Older adults' performance deficits in our task make clear the importance of understanding age differences in inference tasks. But what task conditions can favor successful cue abstraction learning and application by older adults? Research shows that providing information about task structure or context can provide a helpful boost to performance (e.g., Heit, 1997), for example, because context and causal information make clear the relation between cues and criterion (Newell, Weston, Tunney, and Shanks, 2009). Future work should evaluate how providing causal or contextual information to participants by allowing participants to rely on previous knowledge or providing real-world scenarios can boost older adults' performance in these inference tasks.

Finally, the behavior of older adults in our study contrasts starkly with that of 9- to 11-year-olds in a similar experiment (von Helversen et al., 2010). Older adults in our study relied on rule-based processes regardless of task condition, while most children in von Helversen et al. relied on similarity-based strategies in both multiple-cue judgment and categorization tasks. Taken together, the results of the two studies emphasize the importance of going beyond performance measures to understand life span differences in inference behavior: A comparison of children and older adults' performance, for example, in terms of monetary payoff relative to that of young adults, would suggest that the children and older adults are equivalent in their categorization and judgment behavior. Only by relying on the computational modeling of participants' inferences were we able to distinguish between rule- and similarity-based processes and identify differences between children and older adults. Future studies will need to investigate how meta-cognitive considerations as well as objective cognitive limitations contribute to such stark discontinuity between the two ends of the life span.

In conclusion, the idea that different cognitive processes can underlie judgments is crucial to understanding age differences in inference processes. Our results provide evidence that preferences for rule-based processes change across the life span and make clear the need to understand how to foster successful selection and application of rule- and similarity-based inference processes by older adults.

## References


Ashby, F. G., Noble, S., Filoteo, V., Waldron, E., & Ell, S. W. (2003). Category learning deficits in Parkinson's disease. *Neuropsychology, 17,* 115–124. doi:10.1037/0894-4105.17.1.115

Baltes, P. B. (1997). On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as foundation of developmental theory. *American Psychologist, 52,* 366–380. doi:10.1037/0003-066X.52.4.366

Blanchard-Fields, F. (2007). Everyday problem solving and emotion: An adult developmental perspective. *Current Directions in Psychological Science, 16,* 26–31. doi:10.1111/j.1467-8721.2007.00469.x

Chasseigne, G., Ligneau, C., Grau, S., Lafon, P., Le Gall, A., Roque, M., & Mullet, E. (2004). Aging and probabilistic learning in single- and multiple-cue tasks. *Experimental Aging Research, 30,* 23–45. doi:10.1080/03610730490251469

Christie, A. (1934). *Murder on the Orient Express.* London, England: Collins Crime Club.

Christie, A. (1957). *4:50 from Paddington.* London, England: Collins Crime Club.

DeCaro, M. S., Carlson, K. D., Thomas, R. D., & Beilock, S. L. (2009). When and how less is more: Reply to Tharp & Pickering. *Cognition, 111,* 415–421. doi:10.1016/j.cognition.2009.03.001

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 968–986. doi:10.1037/0278-7393.23.4.968

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging, 13,* 597–607. doi:10.1037/0882-7974.13.4.597

Einhorn, H. J., Kleinmuntz, D. N.,& Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review, 86,* 465–485. doi:10.1037/0033-295X.86.5.465

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127,* 107–140. doi:10.1037/0096-3445.127.2.107

Filoteo, J. V., & Maddox, W. T. (2004). A quantitative model-based approach to examining aging effects on information-integration category learning. *Psychology and Aging, 19,* 171–182. doi:10.1037/0882-7974.19.1.171

Fleischman, D. A., Wilson, R. S., Gabrieli, J. D. E., Bienias, J. L., & Bennett, D. A. (2004). A longitudinal study of implicit and explicit memory in old persons. *Psychology and Aging, 19,* 617–625. doi:10.1037/0882-7974.19.4.617

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19,* 25–42. doi:10.1257/089533005775196732

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition, 65,* 197–230. doi:10.1016/S0010-0277(97)00044-9

Heit, E. (1997). Knowledge and concept learning. In K. Lamberts & D. R. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 7–41). Hove, England: Psychology Press.

Hess, T. M., Leclerc, C. M., Swaim, E., & Weatherbee, S. R. (2009). Aging and everyday judgments: The impact of motivational and processing resource factors. *Psychology and Aging, 24,* 735–740. doi:10.1037/a0016340

Hofer, S. M., & Sliwinski, M. J. (2001). Understanding ageing: An evaluation of research designs for assessing the interdependence of ageing-related changes. *Gerontology, 47,* 341–352. doi:10.1159/000052825

Howard, D. V., Howard, J. H., Jr., Dennis, N. A., LaVine, S., & Valentino, K. (2008). Aging and implicit learning of an invariant association. The

*Journals of Gerontology, Series B: Psychological Sciences,* 63, P100–P105.

Howard, J. H., Jr., Howard, D. V., Dennis, N. A., & Kelly, A. J. (2008). Implicit learning of predictive relationships in three-element visual sequences by young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1139–1157. doi:10.1037/a0012797

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging, 8,* 283–293. doi:10.1037/0882-7974.8.2.283

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General, 132,* 133–156. doi:10.1037/0096-3445.132.1.133

Juslin, P. Karlsson, L., & Olsson, H. (2008). Information integration in multiple-cue judgment: A division of labor hypothesis. *Cognition, 106,* 259–298. doi:10.1016/j.cognition.2007.02.003

Karlsson, L., Juslin, P., & Olsson, H. (2007). Adaptive changes between cue abstraction and exemplar memory in a multiple-cue judgment task with continuous cues. *Psychonomic Bulletin & Review, 14,* 1140–1146. doi:10.3758/BF03193103

Koenig, P., Smith, E. E., Glosser, G., DeVita, C., Moore, P., & McMillan, C., . . . Grossman, M. (2005). The neural basis for novel semantic categorization, *NeuroImage, 24,* 369–383. doi:10.1016/j.neuroimage.2004.08.045

Lehrl, S. (1999). *Mehrfachwahl-Wortschatz-Intelligenztest: Manual mit Block MWT-B* [Multiple-choice vocabulary intelligence test]. Balingen, Germany: Spitta.

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General, 124,* 83–97. doi:10.1037/0096-3445.124.1.83

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide.* Mahwah, NJ: Erlbaum.

Mata, R., Schooler, L., & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging, 22,* 796–810. doi:10.1037/0882-7974.22.4.796

Mata, R., von Helversen, B., & Rieskamp, J. (2010). Learning to choose: Cognitive aging and strategy selection learning. *Psychology and Aging, 25,* 299–309. doi:10.1037/a0018923

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238. doi:10.1037/0033-295X.85.3.207

Newell, B. R., Weston, N. J., Tunney, R. J., & Shanks, D. R. (2009). The effectiveness of feedback in multiple-cue probability learning. *The Quarterly Journal of Experimental Psychology, 62,* 890–908. doi:10.1080/17470210802351411

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114. doi:10.1037/0278-7393.10.1.104

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

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization, 72,* 147–152. doi:10.1016/j.jebo.2009.04.018

Racine, C. A., Barch, D. M., Braver, T. S., & Noelle, D. C. (2006). The effect of age on rule-based category learning. *Aging, Neuropsychology, and Cognition, 13,* 411–434. doi:10.1080/13825580600574377

Thornton, W. J. L., & Dumke, H. (2005). Age differences in everyday problem solving and decision making effectiveness: A meta-analytic review. *Psychology and Aging, 20,* 85–99. doi:10.1037/0882-7974.20.1.85

Touron, D. R., & Hertzog, C. (2004). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging, 19,* 452–466. doi:10.1037/0882-7974.19.3.452

Touron, D. R., & Hertzog, C. (2009). Age differences in strategic behavior during a computation-based skill acquisition task. *Psychology and Aging, 24,* 574–585. doi:10.1037/a0015966

von Helversen, B., Mata, R., & Olsson, H. (2010). Do children profit from looking beyond looks? From similarity-based to cue-abstraction processes in multiple-cue judgment. *Developmental Psychology, 46,* 220–229. doi:10.1037/a0016690

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised: Manual.* New York, NY: Psychological Corp.

Wood, S., Busemeyer, J., Koling, A., Cox, C. R., & Davis, H. (2005). Older adults as adaptive decision makers: Evidence from the Iowa gambling task. *Psychology and Aging, 20,* 220–225. doi:10.1037/0882-7974.20.2.220

## Appendix A

### Mathematical Models

#### Exemplar Model

The exemplar model assumes that the judgment is the average of the criterion values $c$, 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. Equation A1 can be applied in the two task conditions (binary vs. continuous), with only the criterion for the items varying as a function of condition, specifically, being either binary (0 or 1) or continuous (from 10 to 20). 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, 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 Model

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

$$\hat{y}_p = k + \sum_{1=J}^{J} w_j \cdot c_{j,} \tag{A4}$$

where the intercept $k$ and the weights $w$ are free parameters. If $k = 10$, $w_1 = 4$, $w_2 = 3$, $w_3 = 2$, and $w_4 = 1$, Equation 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, all objects with $C > 15$ are classified into Group B, and objects with $C = 15$ have probability of .5 in being classified into Group A or Group 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 et al, 2003):

$$\hat{p}(\hat{b} = 1) = \frac{e^{k + \Sigma W_i \cdot c_i}}{1 - e^{k = \Sigma W_i \cdot c_i}}, \tag{A5}$$

where $\mathrm{W_i}$ are the cue weights and k the intercept.

#### Model Predictions

We created 16 items (stimuli) using Equation 1 but excluded six from the training set so that participants would only face a subset of items during training (Table 2). Excluding items from the training set ensures that the CAM and EBM can, in principle, make different predictions for new test items (see Juslin et al., 2003, for a full rationale). When all items are learned during the training phase, the models cannot make different predictions; however, when the training set is constrained, the model predictions can diverge: Whereas the cue abstraction model makes judgments according to the linear additive rule, the exemplar model is sensitive to similarity between new items and the items stored in memory. Consequently, EBM's predictions for new items are biased by similarity between these and the items stored in memory, that is, the items in the training set.

Figure A1 provides an illustration of how the exemplar model's predictions are influenced by using a constrained training set. To obtain these predictions, we fitted the EBM and CAM to the training set by minimizing the RMSD between the model prediction and the actual criteria as a goodness-of-fit measure. We then used the estimated parameter values to make predictions for the full test set. We estimated four free parameters for the exemplar model (an attention weight $s$ for each cue, constrained to add to 1, and the sensitivity parameter $h$) using a nonlinear least squares fit, assuming the training set as a knowledge base. We obtained parameter estimates for the cue abstraction model in the binary condition with a nonlinear least squares (iterative) procedure with the parameter values of a logistic regression as the starting values. Parameters referring to parameter weights were bound between −10 and 10 to match the fitting procedure used for individual participants (see main text). In the continuous task, we calculated

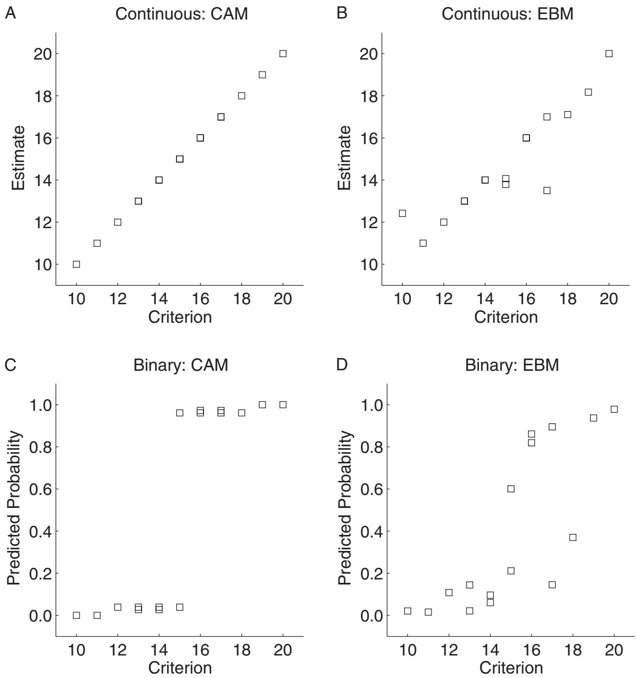


*Figure A1.* Model predictions in the continuous and the binary condition. CAM = cue abstraction model; EBM = exemplar-based model.

parameter values analytically by running a multiple linear regression on participants' responses. The parameter values for the regression model in the continuous condition were $w_1 = 4$, $w_2 = 3$, $w_3 = 2$, $w_4 = 1$, $k = 10$, and in the binary condition, $w_1 = 6.79$, $w_2 = 6.42$, $w_3 = 6.78$, $w_4 = 0.0001$, $k = -10$. The estimated parameter values for the exemplar model in the continuous condition were $s_1 = .25$, $s_2 = .25$, $s_3 = .25$, $s_4 = .25$, $h = 41.52$, and in the binary condition, $s_1 = .28$, $s_2 = .28$, $s_3 = .30$, $s_4 = .14$, $h = 8.27$.

As can be seen in Figure A1A, the CAM is able to perfectly estimate the criteria for all test items because the estimation of cue weights and intercept matches the structure of the environment (Equations 1 and A4). In turn, as can be seen in Figure A1B, the exemplar models' predictions diverge for some items that were not experienced before. For example, one extreme item with criterion 10 is omitted in the constrained training set (cue values: 0, 0, 0, 0: criterion = 10 or 0 for continuous and binary conditions, respectively). When presented with the extreme option, an addition of the cue values as implied by cue abstraction will produce the most extreme response, 10. Moreover, the estimate for the most extreme item is more extreme than the estimate for the second-to-most extreme item (with a criterion of 11). In other words, the estimate for the item (0, 0, 0, 0) is lower than for item (0, 0, 0, 1). The exemplar model, on the other hand, implies the opposite pattern. On average, the response proportion for the extreme item (0, 0, 0, 0) will be higher than for item (0, 0, 0, 1) The estimate for (0, 0, 0, 1) is determined almost exclusively by retrieval of the identical exemplar learned during training (criterion = 11), and other exemplars receive marginal impact (i.e., because of the multiplicative similarity rule in Equation A2). For the new item (0, 0, 0, 0), the estimate cannot be dominated by the identical exemplar because this has not been learned during training. Consequently, other exemplars in memory, all with criteria larger than 10, bias its estimation. This is a well-known difference between the cue abstraction and exemplar models: The rules of the CAM afford extrapolation, while an EBM is unable to extrapolate (DeLosh, Busemeyer, & McDaniel, 1997; Erickson & Kruschke, 1998; Juslin et al., 2003). Figure A1C shows a good categorization performance by the regression model with the exception of the items with a criterion of 15. In turn, for the exemplar model, and as can be seen in Figure A1D, the estimates for some new items, in particular, those with criterion values 17 and 18, are influenced by the similarity to poor items (criterion = 0). In sum, these results suggest that the two models are able to make different predictions given the constrained training set we employed.

# Appendix B

## Model Fitting and Participant Classification

### Individual Model Fits and Participant Classification

We fitted the EBM and CAM to each individual participant and classified the participant accordingly. 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 fit of the two models (CAM, EBM) in each condition (binary vs. continuous). Table B1 presents the model fits and respective classification for each individual participant.

### Average Judgments

We also explored whether average estimates for those participants classified as either CAM or EBM users matched the qualitative expectations derived from the formal models. For this purpose, we compared model predictions assuming perfect learning of

Table B1
*Model Fits for Individual Participants*

| Task condition/participant | Young adults EBM | Young adults CAM | Young adults Classification | Older adults EBM | Older adults CAM | Older adults Classification |
|---|---|---|---|---|---|---|
| Continuous | | | | | | |
| 1 | 1.10 | 0.23 | CAM[a] | 1.24 | 1.30 | UN |
| 2 | 1.26 | 1.83 | EBM | 1.19 | 1.40 | EBM |
| 3 | 1.23 | 1.00 | CAM[a] | 1.85 | 1.97 | EBM |
| 4 | 1.21 | 0.38 | CAM[a] | 1.80 | 1.85 | UN |
| 5 | 1.07 | 1.24 | EBM | 1.05 | 1.44 | EBM[a] |
| 6 | 0.87 | 0.90 | UN[a] | 1.96 | 0.53 | CAM |
| 7 | 1.50 | 1.50 | UNa | 1.90 | 2.26 | EBM |
| 8 | 1.01 | 1.36 | EBM[a] | 2.41 | 0.98 | CAM |
| 9 | 1.52 | 1.59 | UN | 1.04 | 1.77 | EBM[a] |
| 10 | 0.93 | 1.41 | EBM[a] | 2.25 | 2.24 | UN |
| 11 | 1.08 | 1.80 | EBM[a] | 1.80 | 2.28 | EBM |
| 12 | 0.92 | 1.01 | EBM[a] | 1.49 | 1.33 | CAM[a] |
| 13 | 1.31 | 1.56 | EBM[a] | 1.78 | 1.77 | UN |
| 14 | 1.28 | 1.06 | CAM[a] | 1.41 | 1.41 | UN[a] |
| 15 | 1.31 | 1.23 | CAM[a] | 1.14 | 1.41 | EBM[a] |
| 16 | 1.65 | 1.98 | EBM[a] | 1.38 | 1.38 | UN |
| 17 | 1.05 | 1.00 | UN[a] | 0.74 | 0.49 | CAM |
| 18 | 1.30 | 0.64 | CAM[a] | 1.66 | 0.73 | CAM |
| 19 | 0.96 | 0.73 | CAM[a] | 2.66 | 2.44 | CAM |
| 20 | 1.10 | 0.53 | CAM[a] | 1.13 | 1.03 | CAM |
| 21 | 1.76 | 1.39 | CAM | 2.48 | 2.38 | CAM[a] |
| 22 | 1.81 | 0.97 | CAM[a] | 1.47 | 1.32 | CAM[a] |
| 23 | 0.95 | 1.47 | EBM[a] | 1.46 | 1.41 | UN |
| 24 | 0.97 | 0.09 | CAM[a] | 1.05 | 0.92 | CAM[a] |
| 25 | 1.42 | 0.90 | CAM[a] | 1.51 | 0.78 | CAM |
| Binary | | | | | | |
| 1 | 0.16 | 0.24 | EBM[a] | 0.31 | 0.33 | UN |
| 2 | 0.00 | 0.04 | EBM[a] | 0.40 | 0.14 | CAM[a] |
| 3 | 0.32 | 0.24 | CAM[a] | 0.43 | 0.21 | CAM |
| 4 | 0.26 | 0.01 | CAM[a] | 0.43 | 0.18 | CAM |
| 5 | 0.09 | 0.19 | EBM[a] | 0.44 | 0.32 | CAM[a] |
| 6 | 0.24 | 0.25 | UNa | 0.25 | 0.30 | EBM[a] |
| 7 | 0.15 | 0.26 | EBM[a] | 0.43 | 0.30 | CAM[a] |
| 8 | 0.43 | 0.51 | EBM[a] | 0.21 | 0.23 | UN[a] |
| 9 | 0.06 | 0.38 | EBM[a] | 0.31 | 0.04 | CAM[a] |
| 10 | 0.14 | 0.20 | EBM[a] | 0.36 | 0.29 | CAM[a] |
| 11 | 0.43 | 0.32 | CAM[a] | 0.25 | 0.15 | CAM[a] |
| 12 | 0.22 | 0.11 | CAM[a] | 0.36 | 0.04 | CAM[a] |
| 13 | 0.00 | 0.04 | EBM[a] | 0.18 | 0.20 | UN |
| 14 | 0.00 | 0.18 | EBM[a] | 0.38 | 0.22 | CAM |
| 15 | 0.25 | 0.18 | CAM[a] | 0.16 | 0.01 | CAM[a] |
| 16 | 0.24 | 0.16 | CAM[a] | 0.43 | 0.22 | CAM |
| 17 | 0.42 | 0.24 | CAM[a] | 0.21 | 0.27 | EBM[a] |
| 18 | 0.11 | 0.15 | EBM[a] | 0.00 | 0.04 | EBM[a] |
| 19 | 0.21 | 0.41 | EBM[a] | 0.38 | 0.07 | CAM[a] |
| 20 | 0.21 | 0.21 | UN[a] | 0.41 | 0.40 | UN |
| 21 | 0.14 | 0.19 | EBM[a] | 0.15 | 0.12 | CAM[a] |
| 22 | 0.32 | 0.11 | CAM[a] | 0.34 | 0.11 | CAM[a] |
| 23 | 0.07 | 0.18 | EBM[a] | 0.29 | 0.20 | CAM[a] |
| 24 | | | | 0.21 | 0.14 | CAM |
| 25 | | | | 0.21 | 0.11 | CAM[a] |

*Note.* CAM = cue abstraction model; EBM = exemplar-based model; UN = unclassified.
[a] Learning criterion reached.

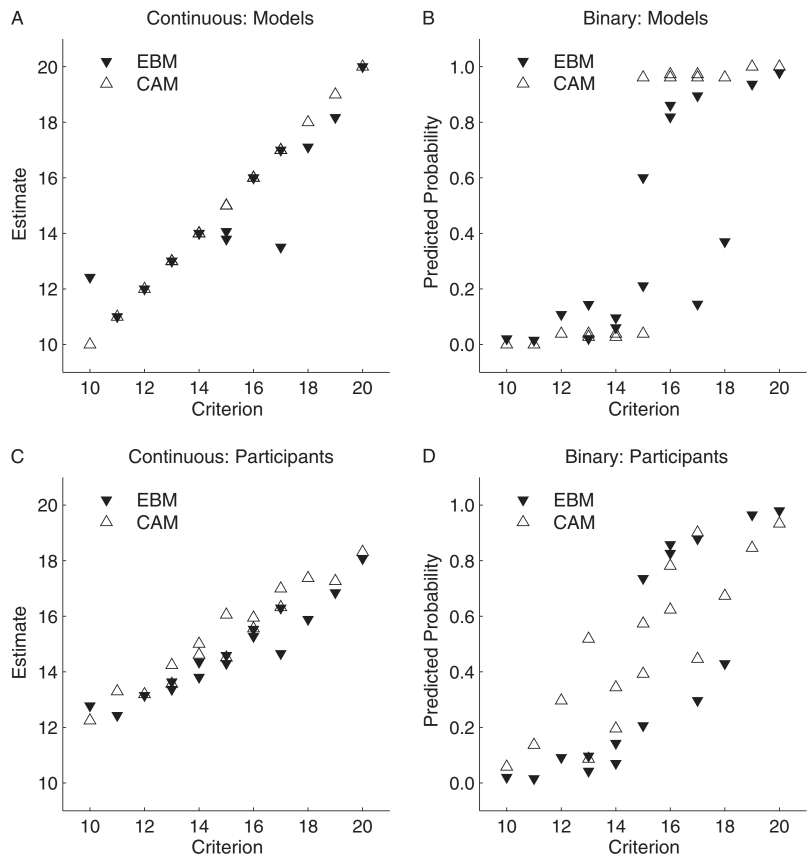


*Figure B1.* Top panels: Model judgments using optimal parameter values. Bottom panels: Model judgments using parameter estimates obtained from fitting participants' judgments. CAM = cue abstraction model; EBM = exemplar-based model.

the training set (see Appendix A) to model judgments obtained using the parameters estimated from the responses of cue abstraction and exemplar users. A good qualitative match between model predictions using optimal weights and model judgments using fitted parameter values would provide additional support for our participant classification. To obtain parameter estimates for exemplar and cue abstraction users, we averaged the estimates or judgments of those classified as EBM or CAM users for each item and fitted the models to the averaged data. In the continuous condition, the parameter estimates for participants classified as EBM users were $s_1$=.30, $s_2$ = .25, $s_3$ =.26, $s_4$ =.19, and $h$ = 5.16, and for those classified as CAM users were $w_1$= 2.76, $w_2$ = 1.32, $w_3$ =0.94, $w_4$ = 1.05, and $k$ = 12.25. In the binary condition, the parameter estimates for participants' classified as EBM users were $s_1$=.28, $s_2$ = 40, $s_3$ =.26, $s_4$ =.06, and $h$ = 8.55, and for those classified as CAM users were $w_1$=2.14, $w_2$ = 0.43, $w_3$ =1.92, $w_4$ = 0.94, and $k = -2.78$.

The top panels in Figure B1 present model predictions as computed in Appendix A; The bottom panels in Figure B1 present model judgments using the parameter values estimated from participants' average data. Overall, and as expected, the judgments based on participants' data qualitatively match the models' predictions. For example, in the continuous condition, the judgments are less extreme than predicted by the models, a common pattern in estimation tasks and one that suggests that there may be noise in the execution of the cue abstraction and exemplar processes (e.g., Juslin et al., 2003). Nevertheless, the judgments based on EBM users' data do not show evidence of extrapolation for the extreme item (criterion = 10) while those of CAM users do. In the binary condition, judgments based on EBM users' data show the predicted pattern of generally good categorization performance but difficulties regarding items with a criterion of 17 and 18. In turn, the pattern for CAM users seems qualitatively different. Note that while the pattern for CAM user does not strictly match the step function predicted by error-free model, this is to be expected from a noisy execution of the cue abstraction process, which can be captured by the regression model by assigning nonoptimal values to the cue weightings. In sum, these results provide support for both our classification procedure and our approach of conducting analyses on new items for which the two models make more distinct predictions.

## Appendix C

### Signal Detection Analysis

We calculated two signal detection theory parameters, discriminability ($d'$) and bias ($\beta$), on the basis of each participant's hit ($H$) and false alarm ($FA$) rates (cf. Macmillan & Creelman, 2005). When hits or false alarm rates were zero, 0.5 was added to the total number of hits or false alarms, whereas 0.5 was subtracted from the total number of hits or false alarms, when hit rates or false alarm rates amounted to 1. Discriminability $d'$ was calculated for each participant as $d' = z(H) - z(FA)$. The function z( ) transforms probabilities into real values that are normally distributed with mean $\mu = 0$ and standard deviation $\sigma = 1$. Bias $\beta$ was calculated as $\exp(d' \cdot c)$ with $c$ being the relative criterion calculated as $c = -\frac{1}{2} \cdot [z(FA) + z(H)]$. $\beta$ reflects an observer's bias to say "yes" or "no" with the unbiased observer having a value around 1.0. As the bias to say yes increases, $\beta$ approaches 0. As the bias to say no increases, $\beta$ increases over 1.0 on an open-ended scale.