Identifying Similarity- and Rule-Based Processes in Quantitative Judgments: A Multi-Method Approach Combining Cognitive Modeling and Eye Tracking

Florian I. Seitz*[1], Rebecca Albrecht*[1], Bettina von Helversen[2], Jörg Rieskamp[1], and Agnes Rosner[3,4]

[1]Center for Economic Psychology, Department of Psychology, University of Basel

[2]Department of Psychology, Faculty of Human and Health Sciences, University of Bremen

[3]Cognitive Psychology Research Unit, Institute of Psychology, Leibniz University Hannover

[4]Center for Cognitive Psychology, Department of Psychology, University of Zurich

Author Note

* These authors contributed equally: Florian I. Seitz and Rebecca Albrecht.

Florian I. Seitz ⓘD https://orcid.org/0000-0003-3217-4083

Correspondence concerning this article should be addressed to Florian I. Seitz, Center for Economic Psychology, Department of Psychology, University of Basel, Missionsstrasse 62a, 4055 Basel, Switzerland. E-mail: florian.seitz@unibas.ch

Abstract

Quantitative judgments have been suggested to result from a mixture of similarity- and rule-based processing. People can judge an object's criterion value based on the object's similarity to previously experienced exemplars and based on a rule that integrates the object's cues like a linear regression. In order to better understand these processes, the present work combines cognitive modeling and eye tracking and tests whether people who rely more on the similarity to exemplars also look more at the exemplar locations on the screen. In two eye tracking studies, participants learned to assign each of four exemplars to a different screen corner and criterion value and then judged the criterion value of briefly presented test stimuli. Eye tracking measured participants' gazes to the now empty exemplar locations (a phenomenon called *looking-at-nothing*); cognitive modeling of the test phase judgments quantified participants' reliance on a similarity- over a rule-based process. Participants showed more similarity usage and more looking-at-nothing in the study in which the cues were linked to the criterion by a multiplicative function than in the study with an additive cue-criterion link. Focusing on the study with a multiplicative environment, participants relying more on the similarity to exemplars also showed more looking-at-nothing ($\tau = 0.25$, $p = .01$). Within trials, looking-at-nothing was usually directed at the one exemplar that was most similar to the test stimulus. These results show that a multi-method approach combining process tracing and cognitive modeling can provide mutually supportive insights into the processes underlying higher-order cognition.

Keywords: Judgment and Decision Making, Eye Tracking, Computational Modeling, Looking-At-Nothing, Exemplar

Identifying Similarity- and Rule-Based Processes in Quantitative Judgments: A Multi-Method Approach Combining Cognitive Modeling and Eye Tracking

The human mind relies on similarities and rules to infer an object's criterion—be it in categorizations (Rouder & Ratcliff, 2006; Smith & Sloman, 1994) or in quantitative judgments (Juslin et al., 2003; von Helversen & Rieskamp, 2009). People may assign similar objects to similar criterion values or apply a rule to predict an object's criterion from its cues. For example, the suitability of job applicants can be judged based on their similarity to hired employees or by integrating their various skills. People often rely both on similarity- and rule-based processes (Albrecht et al., 2020; Bröder et al., 2017; Erickson & Kruschke, 1998), and even in tasks that favor one process, the other process can affect people's responses (Hahn et al., 2010; Rosner & von Helversen, 2019; von Helversen et al., 2014). However, identifying to what extent people rely on similarity or a rule can prove difficult (cf. Hahn & Chater, 1998): The two processes make similar response predictions in many tasks (e.g., Juslin et al., 2003; Nosofsky et al., 1989; Rouder & Ratcliff, 2006) and are sometimes thought to constitute the extremes on a continuum that allows for fine-grained mixtures (e.g., Bröder et al., 2017; Pothos, 2005). The present work investigates similarity- and rule-based processes in quantitative judgments from a novel angle by combining two methodological approaches: cognitive modeling to identify the two processes at the response level and eye tracking for a better understanding of the process level.

Identifying Similarity- and Rule-Based Processes

We establish the link between cognitive modeling and eye tracking via the role memory retrieval plays when people infer an object's criterion value. Specifically, the similarity-based process relies on the object's similarity to previously experienced objects called *exemplars* (cf. Juslin et al., 2003; Nosofsky, 1986) which are retrieved from episodic memory (Hoffmann et al., 2014). The more the mind relies on similarity over rules, the more it relies on exemplar retrieval from memory. Memory retrieval, in turn, has been linked to eye movements that reinstate the encoding context: Fixating the spatial location

in which some information was previously encoded can help recalling this information (cf. Wynn et al., 2019). Accordingly, when people retrieve an exemplar from memory, they tend to fixate the spatial location that the exemplar occupied during encoding (cf. Scholz et al., 2015). This suggests a close link between cognitive modeling, which quantifies the reliance on similarity over rules, and eye tracking, which measures people's gaze proportions to the exemplars. The present work puts this alleged link to the test.

Cognitive Modeling

When similarity- and rule-based processes make distinct predictions, cognitive modeling provides a way to identify them by measuring their relative contributions to people's responses. For instance, the RulEx-J model of Bröder et al. (2017) predicts a person's judgment $\hat{c}_i$ for object i as a weighted average of the predictions $\hat{c}_i^{\text{Similarity}}$ and $\hat{c}_i^{\text{Rule}}$ of the similarity- and rule-based processes (Izydorczyk & Bröder, 2022, 2023). Formally,

$$\hat{c}_i = \alpha \cdot \hat{c}_i^{\text{Similarity}} + (1 - \alpha) \cdot \hat{c}_i^{\text{Rule}}, \tag{1}$$

where α is a free parameter (with $0 \leq \alpha \leq 1$) representing a person's relative reliance on the similarity- over the rule-based process. Specifically, α can reflect the weight with which one combines the two processes within trials or the probability with which one relies on one of the two processes in a trial (Bröder et al., 2017).

Typically, $\hat{c}_i^{\text{Similarity}}$ and $\hat{c}_i^{\text{Rule}}$ are computed based on Juslin et al. (2003, for the equations, see Appendix A): The similarity-based process predicts an object's criterion value to be the mean of the exemplars' criterion values, each weighted by the exemplar's normalized similarity to the object. Rooted in categorization (cf. Nosofsky, 1986), this similarity-to-exemplars approach successfully describes judgments (cf. Albrecht et al., 2021; Juslin et al., 2003), notably when the cues predict the criterion in a non-linear (e.g., multiplicative) way (e.g., Hoffmann et al., 2013; Juslin et al., 2008; Karlsson et al., 2007; Mata et al., 2012). The rule-based process, in turn, computes a weighted sum of the cue values like a linear model and can predict people's judgments especially in linear, additive

environments. The reason for assuming additive rules are capacity limitations that often

hinder the cognitive system from learning more complex rules (Brehmer, 1969, 1994; Juslin

et al., 2008). Compared to the similarity process, rules readily extrapolate beyond the

range of the exemplars' criterion values; therefore, the two processes can notably be

distinguished using stimuli with extreme cue values. Figure 1 shows the two processes'

prediction errors for a multiplicative environment. For the stimuli with extreme cue values,

the similarity-based predictions are too moderate (Fig. 1b), while the rule-based predictions

can even be too extreme (Fig. 1c), leading to large predictions differences (Fig. 1d).
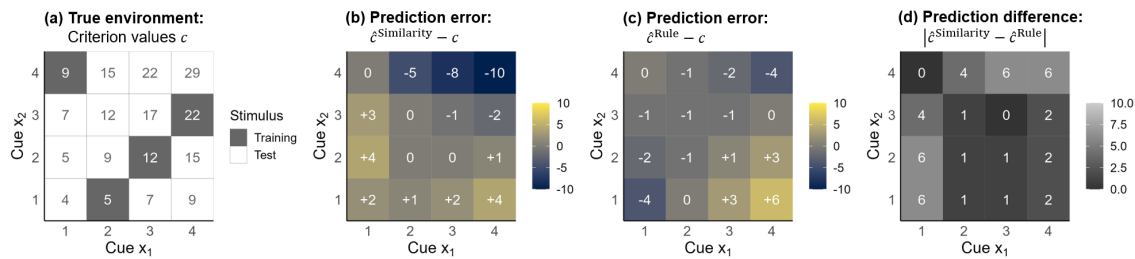
Figure 1

Multiplicative environment used in the main study. Participants estimated a numeric

criterion c of stimuli with two multivalued cues x_1 and x_2. Shown are (a) the true criterion

values resulting from $c = \lfloor (\frac{5}{3} \cdot x_1 \cdot x_2) + 2 \rceil$, the prediction errors from (b) a similarity-based

process and (c) a rule-based process, and (d) the absolute prediction differences of the two

processes. The predictions stem from processes that optimally learned the criterion values

of the four shaded cue combinations in (a) that served as exemplars in the experiment.

Specifically, participants learned the exemplars' criterion values in a training phase with

feedback and then judged the criterion values of all stimuli in a test phase without feedback.

Eye Tracking

Another way to distinguish similarity- from rule-based processes are process tracing

methods such as eye tracking. Specifically, when people retrieve some information from

memory they tend to look at spatial locations associated with the information during

encoding—although the information is no longer visible (called *looking-at-nothing*, Ferreira

et al., 2008; Renkewitz & Jahn, 2012; Richardson & Spivey, 2000; Richardson et al., 2009; Scholz et al., 2015; Spivey & Geng, 2001). For instance, when asked to recall a geometrical shape's orientation or color, participants fixated the blank area on the computer screen in which the shape was originally presented in 24% of all trials (Spivey & Geng, 2001). Cognitive explanations for looking-at-nothing are that eye movements to a spatial location facilitate the memory retrieval of information encoded at that location (Wynn et al., 2019) or reflect attention shifts to information in memory (Scholz et al., 2018).

Applied to criterion inference tasks, looking-at-nothing is found during exemplar retrieval and may thus be related to similarity-based processing (Rosner & von Helversen, 2019; Rosner et al., 2022; Scholz et al., 2015). For instance, compared to a condition that instructed the use of a rule, a corresponding similarity condition led people to fixate longer on blank screen areas previously associated with one exemplar each (Scholz et al., 2015). This difference was particularly pronounced when an exemplar matched the to-be-judged object: The participants in the similarity condition fixated the corresponding spatial position for about 2 s and about 1.5 s longer than the participants in the rule condition. Furthermore, looking-at-nothing can predict people's responses both for quantitative judgments (Rosner & von Helversen, 2019) and categorizations (Rosner et al., 2022), suggesting a tight link to the cognitive processes underlying human inferences.

Overview and Research Aim

The present work extends the previous findings by adding cognitive modeling to test if looking-at-nothing can distinguish similarity- from rule-based processing. Specifically, we test if parameter α reflecting a participant's reliance on a similarity process in cognitive modeling correlates with the gaze proportions to the blank exemplar locations. Additionally, we investigate the properties of such looking-at-nothing, analyzing at what time in a trial it occurs, the number of exemplars looked at per trial, and their similarity to the object being judged. To this end, we ran two eye tracking studies using a multiple-cue judgment task. Participants learned the criterion values and screen locations of four

exemplars in a training phase with feedback and then judged the criterion value of briefly presented test stimuli without feedback. One study defined the criterion by a linear, additive function; the other study by the non-linear, multiplicative function of Fig. 1.

Much more looking-at-nothing was observed in the multiplicative environment than in the additive environment. This is consistent with our predictions, as multiplicative environments tend to be associated with similarity usage and additive environments with rule usage (Juslin et al., 2008; Karlsson et al., 2007). Yet, the low looking-at-nothing rates in the additive environment limit its informative value. Therefore, we report the findings from the additive environment in Appendix B and the findings from the multiplicative environment with substantial looking-at-nothing below in the main text.

Study With a Multiplicative Environment

Method

Participants

In total, 56 students from the University of Zurich participated in a laboratory eye tracking experiment in exchange for CHF 15.00 ($\approx$ USD 17.00) or course credit. Seven participants did not complete the experiment due to eye tracking problems, and one participant was excluded for assigning larger criterion values to objects with smaller cue values—a response pattern opposite to our true environment (Fig. 1) and incompatible with our modeling framework (App. A). This led to a final sample of $N = 48$ participants (31 women; age: $M \pm SD = 26 \pm 8$ years), which corresponds to detecting a correlation of $r = 0.47$ between looking-at-nothing and parameter α from Eq. 1, assuming a type I error of .05 and a power of .95. The experiment took on average about 1 hour and was approved by the ethics board of the Faculty of Arts and Social Sciences of the University of Zurich.

Apparatus

The experiment was programmed in Expyriment (Krause & Lindemann, 2014) using Python 2.7 (Van Rossum & Drake Jr, 1995). Participants were seated in front of a 22-in.

computer screen (1920 × 1080 pixels) at a distance of 700 mm and instructed to position their head in a chin rest. The eye tracker system SMI iView RED sampled data from the right eye at 500 Hz and recorded with iView X 2.8 following a five-point calibration. Fixation detection was done with IDF Event detector 9 (SMI, Teltow) using a peak velocity threshold of 30°/s and a minimum fixation duration of 80 ms.

Design and materials

The stimuli were geometrical figures consisting of two cues (x_1 and x_2) with four possible values each (from 1 to 4) and a criterion value c defined by the multiplicative function $c = \lfloor (\frac{5}{3} \cdot x_1 \cdot x_2) + 2 \rfloor$. Figure 1 (a) shows the possible 16 cue combinations and their true criterion values. Four of the 16 stimuli were selected as exemplars (the shaded cue combinations in Fig. 1a). These stimuli were selected so that both cues take on every possible value exactly once and that some of the remaining stimuli extrapolate beyond the range of the exemplars. Figure 1 (b) and (c) show the prediction errors of the similarity- and the rule-based processes, when trained on the four exemplars. Figure 1 (d) shows that the two processes make different predictions in particular for stimuli 11 (meaning $x_1 = 1$ and $x_2 = 1$), 12, 34, and 44, but not for stimuli close to multiple exemplars. One cue was a rectangle containing 1 to 4 dots, the other cue was a circle containing 1 to 4 lines. The two cues were presented side by side on the screen; one cue was presented in green, and the other in purple (see Fig. 2 for examples). The cue-color association and the cue-side association within stimuli were randomized across participants.

Each of the four exemplars was assigned to one screen corner, with a randomized stimulus-corner association across participants. The distance from the screen center to the center of each exemplar was 9.59° of visual angle (477 pixels; 415 pixels on the x-axis and 235 pixels on the y-axis). Each exemplar had a size of 6.39° × 3.69° of visual angle (320 × 180 pixels). For the gaze analyses, we drew four rectangular areas of interest (AOIs) around the exemplar locations and one around the center of the screen where the test stimuli were presented. Each AOI had a size of 7.67° × 4.43° of visual angle (384 × 216 pixels). The

size of the AOI exceeded the outer borders of each stimulus rectangle by a factor of 0.1, equaling 0.64° of visual angle (32 pixels) on the x-axis and 0.37° (18 pixels) on the y-axis.

Procedure

Participants' task was to judge the criterion values of stimuli on a scale from 1 to 30. Participants completed a training phase in which they built a solid representation for four exemplars (location and criterion value) and a test phase in which they judged the criterion value of all 16 possible stimuli. In the test phase, the stimuli were only briefly presented, and we measured looking-at-nothing after stimulus removal by recording participants' eye movements to the four corners of the screen associated with the exemplars in the training phase. The resulting experimental procedure is illustrated in Fig. 2 and detailed below.

In the location training phase (see Fig. 2a), participants learned to assign the four exemplars to the four corners of the screen within a maximum of 15 blocks. Each block contained eight stimuli—the four exemplars and four distractors with reversed cue values (e.g., distractor 12 for exemplar 21), with random order within blocks. At the beginning of each block, participants could study the four exemplars at their associated screen locations. Then the locations turned to gray rectangles, and participants indicated for one stimulus after another the correct location by clicking on the associated rectangle. If a stimulus was a distractor, participants should click on an "unknown" button in the lower part of the screen. Location training ended after participants answered all stimuli correctly in three consecutive blocks or after 15 blocks.

In the criterion training phase (see Fig. 2b), participants learned to assign the four exemplars to their criterion values. The four stimuli were presented in random order within each block for a maximum of 10 blocks or until participants answered all four stimuli correctly in three blocks. In each trial, participants studied one exemplar in its associated location without time limit (the locations of the other exemplars remained gray). Then, they pressed the left mouse button and reached a scale ranging from 1 to 30 on which they entered their numeric judgment within 2 s. A correct response was highlighted in green; an
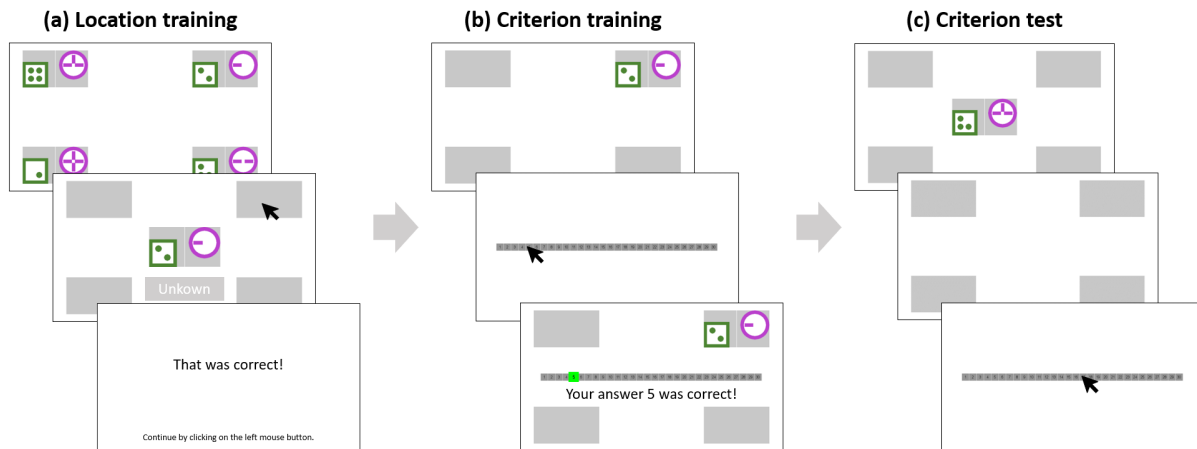
Figure 2

Experimental procedure. In location training (a), participants first saw all four exemplars at their respective location; then, they saw one stimulus per trial, clicked on the associated location or the unknown button, and got feedback. In criterion training (b), participants saw an exemplar, clicked the left mouse button to reach a scale, entered their judgment, and got feedback. In criterion test (c), participants briefly saw a stimulus in the screen center. The exemplar locations were shown as gray rectangles until participants clicked the left mouse button to enter their judgment. There was no feedback in criterion test. Eye tracking measured participants' gazes to the blank exemplar locations after the test stimulus removal (looking-at-nothing), and the criterion test responses were used for cognitive modeling.

incorrect response was highlighted in red together with the correct value highlighted in green. Additionally, verbal feedback was presented below.

In the critical criterion test phase (see Fig. 2c), participants judged the criterion value of all 16 possible stimuli in eight blocks (128 trials), with random order within blocks. In each trial, participants saw a briefly presented test stimulus in the middle of the screen and the four gray rectangles representing the locations of the exemplars, which remained visible after the test stimulus was removed from the screen. After the test stimulus removal, eye tracking measured participants' gaze durations to the blank exemplar locations until they pressed the left mouse button (without a time limit).

Participants then entered their judgment on the response scale within 2 s and got no feedback. The stimulus presentation time was individually calibrated at the beginning of the experiment in a psychophysics test, following a step-wise procedure recommended by García-Pérez (1998): In each trial of the psychophysics test, the participants saw a stimulus in the middle of the screen for a certain duration. Afterwards, four comparison stimuli were each presented in a corner of the screen, and the participants indicated which one corresponded to the original stimulus. The three non-matching alternatives were selected randomly, subject to the constraint that two of them matched the original stimulus in one cue each. The presentation duration started with 2000 ms and was decreased by 336.60 ms after four correct answers in a row and increased by 400 ms after a mistake (a 4-down-1-up staircase method). All 16 stimuli were shown with random order in three blocks (48 trials), and the mean across blocks yielded the stimulus presentation time in the test phase ($M = 412$ ms and $SD = 220$ ms across participants).

At the end of the experiment, a location test was conducted to check if the participants were still able to recall the exemplar locations. The location test had the same procedure as location training except that it comprised only one block (8 trials), that no feedback was given, and that in the beginning, no picture with all exemplars was shown.

Cognitive modeling

The RulEx-J model parameters were estimated by maximum likelihood from individual participants' test phase data using a 16-fold cross-validation and were averaged using the mean by participant across folds[1]. The RulEx-J model contained five free parameters (α from Eq. 1, two cue weights β_1 and β_2, an intercept β_0, and a standard deviation σ, see Appendix A for details and Appendix C for the resulting estimates). Appendix A also shows that α was well recovered in our task, allowing to correlate participants' α estimates with their looking-at-nothing. Appendix C further presents an

———

[1] Each fold used 15 of the 16 stimuli for parameter estimation (120 trials per participant) and the remaining stimulus for out-of-sample prediction (8 trials).

out-of-sample model comparison between the RulEx-J, a pure similarity model (implying $\alpha = 1$), and a pure rule model (implying $\alpha = 0$).

Results

The correlation analyses reported below use Kendall's τ due to normality violations in participants' looking-at-nothing and α estimates and are summarized in Table 1.

Exemplar representation

Participants learned the exemplars' screen locations and criterion values well. On average, they needed 6 blocks ($SD = 3$) to learn the locations, which they correctly recalled with 89% accuracy at the end of the experiment. Similarly, participants learned the criterion values within 7 blocks ($SD = 2$) and continued to judge the exemplars correctly in the test phase with 67% accuracy. Location accuracy and criterion accuracy were related, $\tau = .35$, $p = .002$. As location accuracy is only vaguely graded (8 responses per subject), we combined it with the criterion accuracy by computing the mean of the two measures by participant (see *exemplar accuracy* in Table 1).

Table 1

Variable distributions, normality coefficients W, and Kendall's τ correlation coefficients

				Normality		Correlations: τ and p			
Variable	M	Mdn	SD	W	p	1	2	3	4
1. Exemplar accuracy	.78	.88	.22	.88	< .001		< .001	< .001	< .001
2. Looking-at-nothing: duration	.19	.11	.20	.85	< .001	.34		< .001	.01
3. Looking-at-nothing: strength	.20	.13	.21	.86	< .001	.39	.84		.001
4. Cognitive modeling: α value	.40	.34	.32	.91	.001	.40	.25	.33	

Note. W denotes the statistic from a Shapiro-Wilk test against normality. The correlation table shows τ-coefficients below the diagonal and p-values above the diagonal. Looking-at-nothing: duration equals the gaze duration to all exemplar locations divided by the gaze duration to the exemplar and test stimulus locations; strength equals the gaze duration to the location whose exemplar is most similar to the test stimulus divided by the gaze duration to all exemplar locations.
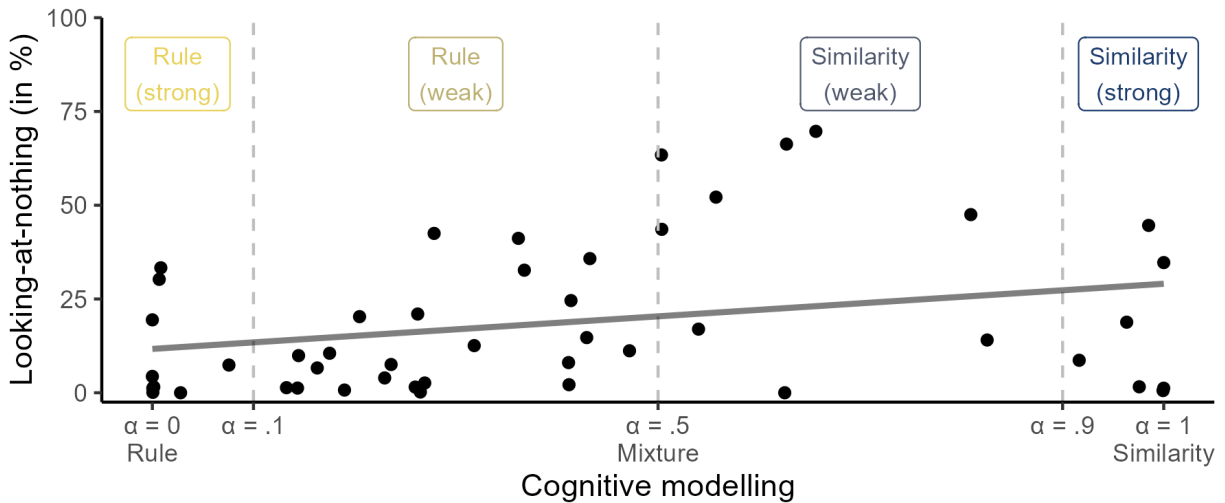
Looking-at-nothing

After the test stimulus removal, a participant looked at blank exemplar locations in $M = 38\%$ of the trials ($SD = 32$ percentage points) and clicked to reach the response scale after $M = 3,760$ ms ($SD = 2,822$ ms). We computed looking-at-nothing as the summed gaze duration to the exemplar locations divided by the summed gaze duration to the exemplar and test stimulus locations (*looking-at-nothing duration* in Table 1)[2]. The mean looking-at-nothing is .19 and varies considerably across participants ($SD = .20$), but is comparable when the test stimulus was old ($M = .17$) or new ($M = .19$), $\tau = .75$, $p < .001$. Looking-at-nothing is quite stable within participants, with stimulus-wise splithalf correlations between odd and even trials across participants of $M = .81$ ($SD = .05$). Furthermore, looking-at-nothing is associated with a higher exemplar accuracy after training, $\tau = .34$, $p < .001$ (see Table 1). Thus, especially the participants who learned the exemplars well gazed back at them in the test phase.

Association between looking-at-nothing and cognitive modeling

The estimated α values vary substantially ($SD = .32$) with a tendency for more rule usage ($M = .40$), $t(47) = -2.17$, 95% CI $= [.31, .49]$, $p = .04$ (see Table 1; for the remaining parameter estimates, see Table C1). Figure 3 shows that participants' mean α values correlate positively with their mean looking-at-nothing: A participant who relied more on similarity over rules according to cognitive modeling (a larger α) also displayed more looking-at-nothing, $\tau = 0.25$, $p = .01$ (see Table 1). The association was particularly pronounced for the 32 participants who weakly relied on a rule or similarity ($.1 < \alpha < .9$), $\tau = 0.43$, $p < .001$, with a drop in looking-at-nothing among the participants with $\alpha > .9$.

On the one hand, this might suggest a non-linear relationship between α and looking-at-nothing: Participants who strongly rely on similarity might have encoded the

---

[2] The results remain qualitatively unchanged if, instead, normalization is carried out over the total trial duration from removing the test stimulus to reaching the response scale. The two measures of looking-at-nothing are strongly correlated $\tau = .91$, $p < .001$.

Figure 3

Association between parameter α and looking-at-nothing. Each point shows the mean estimated α value and the mean looking-at-nothing duration of one participant. A larger α is associated with more looking-at-nothing (dark gray regression line), in particular for the participants who only weakly rely on one of the two processes (.1 < α < .9).

exemplars in a particularly robust way and no longer need to look at their locations during retrieval. However, exemplar accuracy did not differ between the 7 participants with $\alpha >$ .9 ($M = 90\%$) and the 9 participants with $.5 < \alpha < .9$ ($M = 88\%$), tie-corrected asymptotic Wilcoxon-Mann-Whitney Test $W = 28$, $p = .75$.

On the other hand, the result may have been influenced by a subgroup of four participants who generalized the exemplars' criterion values to all objects with the same value on one cue, regardless of the other cue value. Our modeling described these participants by an almost maximal α ($M = .99$) with the weight β_n of the ignored cue n approaching zero[3]. Such a single-cue strategy no longer requires retrieving complete exemplars, but only the combination of a cue value and the criterion value, relaxing the dependency on exemplar memory. Accordingly, the four participants displayed rather low

───────

[3] The α estimates are not perfectly 1 because the four participants deviated from this single-cue strategy in $M = 1\%$ of the trials.

looking-at-nothing ($M = 10\%$)—removing them increased the correlation between α and looking-at-nothing to $\tau = 0.37$, $p < .001$.

We further corroborated our analyses by comparing the 32 rule users with $\alpha < .5$ and the 16 similarity users with $\alpha > .5$ (see Table 2). The similarity users tended to display more looking-at-nothing ($M = 30\%$) than the rule users ($M = 13\%$), $W = 160.5$, $p = .04$, in particular when the four similarity users who ignored one cue are excluded ($M = 37\%$), $W = 77.5$, $p = .003$. Furthermore, the two groups responded differently to the critical stimuli with extreme cue values that distinguish best between the two processes (stimuli 11, 12, 34, and 44, see Fig. 1). Relative to the rule users, the similarity users made much more moderate judgments, that do not extrapolate beyond the range of the exemplars' criterion values learned during training (see Table 2).

Table 2

Looking-at-nothing and mean responses to critical stimuli for rule users and similarity users

| Group | N | Looking-at-nothing | | Mean responses to critical stimuli | | | |
		Duration	Strength	11	12	34	44
Rule users	32	.13	.13	3.13	5.70	19.56	25.30
				(-0.87)	(+0.70)	(-2.44)	(-3.70)
Similarity users	16	.30	.34	4.64	8.15	13.73	21.78
				(+0.64)	(+3.15)	(-8.27)	(-7.22)

Note. Rule users are defined by $\alpha < .5$; similarity users by $\alpha > .5$. As in Table 1, the looking-at-nothing duration considers participants' gaze duration to all exemplar locations, the strength only the gaze duration to the location of the most similar exemplar. The critical stimuli discriminate between rule- and similarity-based processes (see Fig. 1).
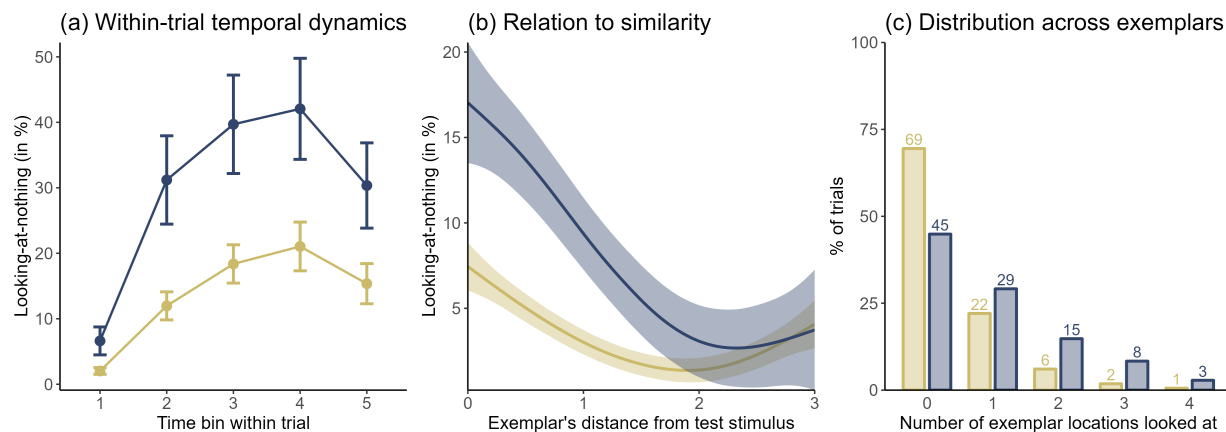
Properties of looking-at-nothing

Beyond analyzing the quantitative association between α and looking-at-nothing, our multi-method approach of cognitive modeling and eye tracking allows for qualitative insights into people's judgment formation process. Figure 4 shows the key results.

Panel (a) shows how looking-at-nothing unfolded within trials after the test stimulus was removed. To this end, the remaining trial time was split into five bins of equal length ($M = 756$ ms and $SD = 1,119$ ms across trials), and participants' mean looking-at-nothing duration was computed for each bin (further aggregated across participants grouped into similarity users, $\alpha > .5$, and rule users, $\alpha < .5$, in Fig. 4). Looking-at-nothing peaked halfway through a trial (bin 4; $M_{\text{similarity}} = 42\%$ among the similarity users $> M_{\text{rule}} = 21\%$ among the rule users, $W = 147.5$, $p = .02$) and was lowest in bin 1, where participants still fixated the screen center where the test stimulus had just disappeared ($M_{\text{similarity}} = 7\% \approx M_{\text{rule}} = 2\%$, $W = 191$, $p = .16$). Participants tended to gaze back at the screen center shortly before they reached the response scale (bin 5; $M_{\text{similarity}} = 30\% \approx M_{\text{rule}} = 15\%$, $W = 168.5$, $p = .06$); however, the difference to bin 4 was not significant, $W = 953$, $p = .15$ across all participants.

Furthermore, looking-at-nothing increased with the exemplar's similarity to the test stimulus—in particular for the participants classified as similarity users (see panel b). Looking-at-nothing peaked when an exemplar matched the test stimuli (a difference of 0 in Fig. 4; $M_{\text{similarity}} = 20\% > M_{\text{rule}} = 8\%$, $W = 160$, $p = .04$) and decreased the more an exemplar differed from the test stimulus in terms of the city-block distance, $\tau = -.22$, $p < .001$ for the similarity users and $\tau = -.10$, $p < .001$ for the rule users. Similarity users showed more looking-at-nothing than rule users for exemplars whose city-block distance from the test stimulus was lower than 1 ($M_{\text{similarity}} = 11\% > M_{\text{rule}} = 4\%$, $W = 144.5$, $p = .02$) but not for exemplars with a city-block distance larger than 1 ($M_{\text{similarity}} = 4\% \approx M_{\text{rule}} = 2\%$, $W = 194.5$, $p = .18$). Thus, the similarity users particularly focused on exemplars similar to the test stimulus. This finding is corroborated by the so-called

Figure 4

Properties of looking-at-nothing for similarity users ($\alpha > .5$, shown in blue) and rule users ($\alpha < .5$, shown in yellow). Panel (a) shows the looking-at-nothing (mean and standard error) when the trial time after the removal of the test stimulus is evenly split into five bins. Panel (b) shows the smoothed looking-at-nothing (mean and 95% CI) as a function of the exemplar's weighted city-block distance from the test stimulus (using each participant's median estimated cue weights β_1 and β_2, normalized by participant to sum to 1). Panel (c) shows how many exemplar locations were looked at in how many percent of the trials.

looking-at-nothing strength (i.e., the gaze duration to the most similar exemplar divided by the summed gaze duration to all exemplar locations): The looking-at-nothing strength is larger for the similarity users ($M = 34\%$) than for the rule users ($M = 13\%$), $W = 137.5$, $p = .01$ (see Table 2), and positively correlates with α, $\tau = .33$, $p = .001$ (see Table 1).

Finally, panel (c) shows that looking-at-nothing was present in $M = 55\%$ of trials among the similarity users, but only in $M = 31\%$ of trials among the rule users, $\chi^2(1) = 338.74$, $p < .001$. In these looking-at-nothing trials, participants usually only looked at one exemplar location, $M_{\text{rule}} = 72\% > M_{\text{similarity}} = 53\%$, $\chi^2(1) = 92.72$, $p < .001$. The location being looked at often corresponded to the exemplar most similar to the test stimulus, especially for the similarity users ($M = 76\%$) and, to a lesser extent, for the rule users ($M = 46\%$), $\chi^2(1) = 128.94$, $p < .001$. Even when participants looked at multiple

exemplar locations within trials, most looking-at-nothing time was directed at a single location ($M = 66\%$, $SD = 15$ percentage points) which belonged to the most similar exemplar in $M = 48\%$ of the trials ($M_{\text{similarity}} = 53\% > M_{\text{rule}} = 41\%$, $\chi^2(1) = 10.21$, $p = .001$). This suggests that people often judge an object's criterion based on a single, similar exemplar, in line with recent evidence (Albrecht et al., 2020).

General Discussion

This work used a multi-methodological approach to investigate the cognitive processes involved when people infer an object's unknown criterion from its cues. Analyzing process and behavioral data from a multiple-cue judgment task, we found a clear correspondence between the results obtained from cognitive modeling and eye tracking: The more cognitive modeling indicated reliance on the similarity to exemplars, the more participants looked at the blank locations where the exemplars were previously encoded (looking-at-nothing, Scholz et al., 2015). Further analyses showed that looking-at-nothing peaked halfway through a trial, increased for exemplars more similar to the test stimulus being judged, and focused mostly one exemplar per trial.

Comparing the two studies, we found substantially more looking-at-nothing and more similarity usage in the multiplicative environment reported in the main text than in the additive environment reported in Appendix B. This is consistent with a cognitive perspective: Additive environments can be perfectly learned by rule-based processes (which assume an additive relation between cues and criterion, see App. A) but not by similarity-based processes. Accordingly, the human mind has a strong inclination to use a rule in an additive environment (Hoffmann et al., 2016; Juslin et al., 2008; Karlsson et al., 2007) and may therefore display little looking-at-nothing (Scholz et al., 2015).

While relying on the similarity to exemplars seems to be clearly associated with looking back at the blank exemplar locations, the present work reported only correlational evidence. The causal direction thus remains unclear: Previous research has shown that inducing either a rule or a similarity process affects looking-at-nothing accordingly (Scholz

et al., 2015); however, more looking-at-nothing might also lead to more exemplar retrieval and similarity usage. Future research could test this hypothesis by experimentally manipulating people's eye movements to the blank exemplar locations.

Furthermore, while our application of cognitive modeling filled a gap in previous literature on looking-at-nothing and exemplar retrieval, the frequentist modeling framework we used can bias α towards the rule process in the presence of much noise (Bröder et al., 2017; Izydorczyk & Bröder, 2022). We accepted this as we focused on the association between α and looking-at-nothing, which remains unaffected by linear transformations in the α estimates. Also, parameter α was well recovered in our task even in the presence of noise (see Figure A1), further corroborating the validity of our approach.

Zooming out, our results suggest that cognitive modeling and process tracing provide mutually supportive insights into the processes underlying higher-order cognition. Process tracing may be particularly beneficial when different cognitive processes lead to similar response predictions and thus cannot be distinguished at the behavioral level (e.g., when test stimuli do not extrapolate beyond the range of the exemplars' criterion values). Note that other process tracing methods than eye tracking can provide important insights into the judgment formation processes too (e.g., verbal reports, Steiner et al., 2021). One avenue for future research could be to broaden up the multi-method approach of this paper by comparing multiple process tracing methods with cognitive modeling.

Conclusion

We found that the process people rely on according to cognitive modeling to make inferences is reflected in their eye movements. Fixating the locations of previously encoded exemplars is associated with a cognitive process that relies on the similarity to exemplars. Additionally, the synergy of cognitive modeling and process tracing brings qualitative insights into the exemplar retrieval. Ultimately, the multi-method approach of the present work sheds light on the processes underlying human cognition in a way that either approach in isolation would be hardly able to do.

Declarations and Open Practices Statement

Funding

This work was supported by three grants of the Swiss National Science Foundation: No. 195389 provided to Florian I. Seitz, No. 157432 provided to Bettina von Helversen, and No. 186032 provided to Agnes Rosner.

Conflicts of interest / competing interests

None.

Ethics approval

The experiments complied with the Ethical Principles of Psychologists and Code of Conduct of the American Psychological Association (APA) and were approved by the ethics board of the Faculty of Arts and Social Sciences of the University of Zurich.

Consent to participate

In both studies, participants provided informed consent.

Consent for publication

In both studies, participants provided consent for publication of their anonymized data.

Availability of data, code, and materials

The anonymized data, the complete analysis code, and the materials are available at https://osf.io/k82yt/?view_only=d413e412902f4ca0a704d7c9b13ec017. None of the experiments was preregistered.

Authors' contributions

All authors conceptualized the project under the supervision of BvH, JR, and AR. RA designed and programmed the experiments. AR took care of the data collection. FS and RA conducted the analyses. FS wrote the manuscript. BvH and AR provided significant revisions to the manuscript.

References

Albrecht, R., Hoffmann, J. A., Pleskac, T. J., Rieskamp, J., & von Helversen, B. (2020). Competitive retrieval strategy causes multimodal response distributions in multiple-cue judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(6), 1064–1090. https://doi.org/10.1037/xlm0000772

Albrecht, R., Jenny, M. A., Nilsson, H., & Rieskamp, J. (2021). The similarity-updating model of probability judgment and belief revision. *Psychological Review*, *128*(6), 1088–1111. https://doi.org/10.1037/rev0000299

Brehmer, B. (1969). Cognitive dependence on additive and configural cue-criterion relations. *The American Journal of Psychology*, *82*(4), 490–503. https://doi.org/10.2307/1420442

Brehmer, B. (1994). The psychology of linear judgement models. *Acta Psychologica*, *87*(2-3), 137–154. https://doi.org/10.1016/0001-6918(94)90048-5

Bröder, A., Gräf, M., & Kieslich, P. J. (2017). Measuring the relative contributions of rule-based and exemplar-based processes in judgment: Validation of a simple model. *Judgment and Decision Making*, *12*(5), 491–506.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*(2), 107–140. https://doi.org/10.1037/0096-3445.127.2.107

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11), 405–410. https://doi.org/10.1016/j.tics.2008.07.007

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. *Vision Research*, *38*(12), 1861–1881. https://doi.org/10.1016/s0042-6989(97)00340-4

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? exhaustive? empirically distinguishable? *Cognition*, *65*(2), 197–230. https://doi.org/10.1016/s0010-0277(97)00044-9

Hahn, U., Prat-Sala, M., Pothos, E. M., & Brumby, D. P. (2010). Exemplar similarity and rule application. *Cognition*, *114*(1), 1–18. https://doi.org/10.1016/j.cognition.2009.08.011

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). Deliberation's blindsight: How cognitive load can improve judgments. *Psychological Science*, *24*(6), 869–879. https://doi.org/10.1177/0956797612463581

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2014). Pillars of judgment: How memory abilities affect performance in rule-based and exemplar-based judgments. *Journal of Experimental Psychology: General*, *143*(6), 2242–2267. https://doi.org/10.1037/a0037989

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2016). Similar task features shape judgment and categorization processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1193–1217. https://doi.org/10.1037/xlm0000241

Izydorczyk, D., & Bröder, A. (2022). Measuring the mixture of rule-based and exemplar-based processes in judgment: A hierarchical Bayesian approach. *Decision*, *10*(4), 347–371. https://doi.org/10.1037/dec0000195

Izydorczyk, D., & Bröder, A. (2023). What is the airspeed velocity of an unladen swallow? Modeling numerical judgments of realistic stimuli. *Psychonomic Bulletin & Review*, 1–15. https://doi.org/10.3758/s13423-023-02331-0

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*(1), 259–298. https://doi.org/10.1016/j.cognition.2007.02.003

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and
multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*(1),
133–156. https://doi.org/10.1037/0096-3445.132.1.133

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*(6), 1140–1146.
https://doi.org/10.3758/bf03193103

Krause, F., & Lindemann, O. (2014). Expyriment: A python library for cognitive and
neuroscientific experiments. *Behavior Research Methods*, *46*(2), 416–428.
https://doi.org/10.3758/s13428-013-0390-6

Mata, R., von Helversen, B., Karlsson, L., & Cüpper, L. (2012). Adult age differences in
categorization and multiple-cue judgment. *Developmental Psychology*, *48*(4), 1188.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization
relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.
https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization,
identification, and recognition. *Journal of Experimental Psychology: Learning,
Memory, and Cognition*, *15*(2), 282–304.
https://doi.org/10.1037//0278-7393.15.2.282

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral and Brain
Sciences*, *28*(1), 1–14. https://doi.org/10.1017/S0140525X05000014

Renkewitz, F., & Jahn, G. (2012). Memory indexing: A novel method for tracing memory
processes in complex cognitive tasks. *Journal of Experimental Psychology: Learning,
Memory, and Cognition*, *38*(6), 1622. https://doi.org/10.1037/a0028073

Richardson, D. C., Altmann, G. T., Spivey, M. J., & Hoover, M. A. (2009). Much ado
about eye movements to nothing: A response to Ferreira et al.: Taking a new look

at looking at nothing. *Trends in Cognitive Sciences*, *13*(6), 235–236. https://doi.org/10.1016/j.tics.2009.02.006

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and hollywood squares: Looking at things that aren't there anymore. *Cognition*, *76*(3), 269–295. https://doi.org/10.1016/S0010-0277(00)00084-6

Rosner, A., Schaffner, M., & von Helversen, B. (2022). When the eyes have it and when not: How multiple sources of activation combine to guide eye movements during multiattribute decision making. *Journal of Experimental Psychology: General*, *151*(6), 1394–1418. https://doi.org/10.1037/xge0000833

Rosner, A., & von Helversen, B. (2019). Memory shapes judgments: Tracing how memory biases judgments by inducing the retrieval of exemplars. *Cognition*, *190*, 165–169. https://doi.org/10.1016/j.cognition.2019.05.004

Rouder, J. N., & Ratcliff, R. (2006). Comparing exemplar- and rule-based theories of categorization. *Current Directions in Psychological Science*, *15*(1), 9–13. https://doi.org/10.1111/j.0963-7214.2006.00397.x

Scholz, A., Klichowicz, A., & Krems, J. F. (2018). Covert shifts of attention can account for the functional role of "eye movements to nothing". *Memory & Cognition*, *46*, 230–243. https://doi.org/10.3758/s13421-017-0760-x

Scholz, A., von Helversen, B., & Rieskamp, J. (2015). Eye movements reveal memory processes during similarity-and rule-based decision making. *Cognition*, *136*, 228–246. https://doi.org/10.1016/j.cognition.2014.11.019

Seitz, F. I. (2024). Overcoming the empirical challenges of geometric similarities: A model-based review. *Manuscript in preparation.*

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*(1), 54–87. https://doi.org/10.1016/0022-2496(64)90017-3

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243

Smith, E. E., & Sloman, S. A. (1994). Similarity-versus rule-based categorization. *Memory & Cognition*, *22*(4), 377–386. https://doi.org/10.3758/BF03200864

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*(4), 235–241. https://doi.org/10.1007/s004260100059

Steiner, M. D., Seitz, F. I., & Frey, R. (2021). Through the window of my mind: Mapping information integration and the cognitive representations underlying self-reported risk preference. *Decision*, *8*(2), 97–122. https://doi.org/10.1037/dec0000127

Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

von Helversen, B., Herzog, S. M., & Rieskamp, J. (2014). Haunted by a doppelgänger: Irrelevant facial similarity affects rule-based judgments. *Experimental Psychology*, *61*(1), 12–22. https://doi.org/10.1027/1618-3169/a000221

von Helversen, B., & Rieskamp, J. (2009). Models of quantitative estimations: Rule-based and exemplar-based processes compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 867–889. https://doi.org/10.1037/a0015501

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Wynn, J. S., Shen, K., & Ryan, J. D. (2019). Eye movements actively reinstate spatiotemporal mnemonic content. *Vision*, *3*(2), 21. https://doi.org/10.3390/vision3020021

Appendix A

Cognitive Models

We performed cognitive modeling within the RulEx-J framework of Bröder et al. (2017), which predicts people's multiple-cue judgment to be a weighted average of the predictions resulting from a similarity- and a rule-based process (see Eq. 1). We now outline the formal similarity- and rule-based processes we used, which are based on Juslin et al. (2003).

Similarity-Based Process

The similarity-based process predicts an object's criterion value based on its similarity to previously experienced objects called exemplars. Similar exemplars affect the object's predicted criterion value more than dissimilar exemplars. Specifically, the similarity-based process predicts object i's criterion value $\hat{c}_i^{\text{Similarity}}$ to be the mean of the criterion values across exemplars, each weighted by its normalized similarity to i. Formally,

$$\hat{c}_i^{\text{Similarity}} = \frac{\sum_j s_{ij} \cdot c_j}{\sum_j s_{ij}}, \tag{A1}$$

where s_{ij} is the similarity between object i and exemplar j with criterion value c_j.

The similarity s_{ij}, in turn, is computed from the summed cue value differences between object i and exemplar j. A larger difference results in a smaller similarity and vice versa. We modeled the negative relation between the similarity s_{ij} and the summed difference d_{ij} with Shepard (1987) universal law of generalization $s_{ij} = \exp\left(-d_{ij}\right)$ and computed d_{ij} with the weighted city-block distance $d_{ij} = \sum_{n=1}^{N} \beta_n \cdot |x_{in} - x_{jn}|$, leading to

$$s_{ij} = \exp\left(-\sum_{n=1}^{N} \beta_n \cdot |x_{in} - x_{jn}|\right), \tag{A2}$$

where x_{in} is object i's value x on cue n and β_n is a free model parameter denoting the weight given to the differences between object i and exemplar j with respect to cue n (with $\beta_n \geq 0$). Eq. A2 thus computes a weighted sum of the cue value differences and transforms it into an inversely related measure of similarity. Other distance-based formalizations for s_{ij} exist (cf. Nosofsky, 1986; Seitz, 2024)—our formalization has been shown to describe similarity relations for stimuli like ours with highly separable cues (Shepard, 1964).

Rule-Based Process

In contrast, the rule-based process uses a main effects linear regression model to predict an object's criterion value from its cues. Specifically, the rule-based process predicts object i's criterion value $\hat{c}_i^{\text{Rule}}$ as a weighted sum of i's cue values

$$\hat{c}_i^{\text{Rule}} = \beta_0 + \sum_{n=1}^{N} \beta_n \cdot x_{in}, \tag{A3}$$

where β_0 represents an intercept (i.e., the predicted criterion value when all cue values equal zero) and β_n represents the weight for cue n. We constrained the weight β_n in the rule-based process to equal β_n in the similarity-based process, as this model led to a higher out-of-sample log likelihood.
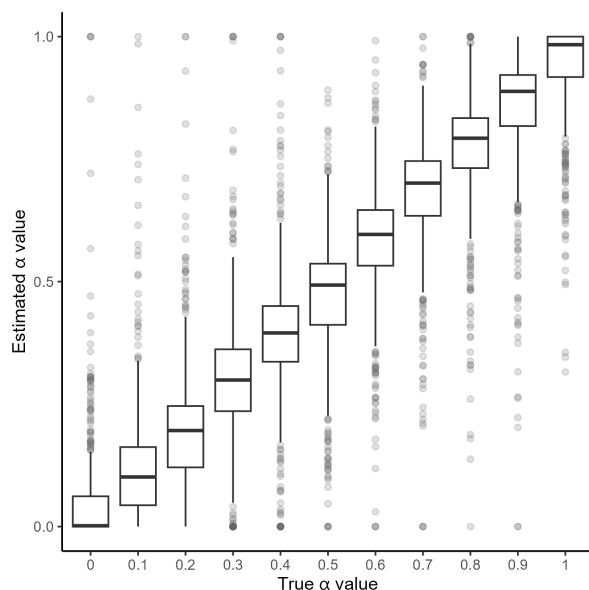
Combining Similarity- and Rule-Based Processes: Parameter Recovery of α

We combined the similarity- and rule-based processes using the RulEx-J framework of Bröder et al. (2017), $\hat{c}_i = \alpha \cdot \hat{c}_i^{\text{Similarity}} + (1 - \alpha) \cdot \hat{c}_i^{\text{Rule}}$ (Eq. 1), and assumed that people's judgments are sampled from a normal distribution, $R_i \sim \mathcal{N}(\hat{c}_i, \sigma)$, where the mean equals the combined prediction $\hat{c}_i$ and the standard deviation $\sigma \geq 0$ is a free parameter.

Ultimately, we aimed to associate the estimated α values with participants' eye movement data. To test whether α was recoverable in our experimental task, we simulated multiple-cue judgments based on various parameter value combinations and checked to what extent the estimated α values when fitting the models correspond to those used to simulate the data. Specifically, in each out of 500 iterations, we randomly sampled a parameter value combination (excluding α) and made predictions for the similarity- and rule-based processes as per Eqs. A1 and A3. Then, for each α level from 0 to 1 in steps of .1, the combined prediction $\hat{c}_i$ was computed and used to simulate noisy responses $R_i \sim \mathcal{N}(\hat{c}_i, \sigma)$ for all 16 stimuli i repeated eight times as in the actual experiments. The model was then refit to the simulated responses using maximum likelihood.

Parameter α was recovered well in our task (see Figure A1). The mean correlation between the true and estimated α across iterations was $r = 0.92$. The mean absolute error

was low (MAE $= .09$), and the mean signed difference was close to zero (MSD $= .01$), indicating no estimation bias. An α of .5 (denoting equal reliance on the similarity- and rule-based processes) was well recovered, $M = .47$ across iterations, and a stronger reliance on one process was almost perfectly recovered in a qualitative way: The true and estimated values of α diverged from .5 in the same direction in 93% of cases. This means that we can classify participants into rule and similarity users with confidence based on their α value.

Figure A1

Parameter recovery. For each true α level, the α values estimated from simulated responses in our experimental task are shown as box plots over the 500 simulation iterations.

Summary

Our cognitive modeling combines a similarity- and a rule-based process (see Juslin et al., 2003) using the RulEx-J framework of Bröder et al. (2017). In our task with $N = 2$ cues, this yields five free model parameters: the intercept β_0, the weights β_1 and β_2, α denoting the reliance on similarity over rules, and the standard deviation σ of the normal distribution relating predictions to responses. Given that in our task the feature values ranged from 1 to 4 and the criterion values from 0 to 30, we constrained the parameters as follows: $-10 \leq \beta_0 \leq 10$, $0 \leq \beta_n \leq 10$ for $n \in \{1, 2\}$, $0 \leq \alpha \leq 1$, and $0 \leq \sigma \leq 10$.

Appendix B

Study with an Additive Environment

This section shows the result of the study using an additive function to link the cues to the criterion. Compared to the multiplicative environment, the additive environment yielded only little looking-at-nothing and little reliance on similarity, which is in line with previous research (e.g., Hoffmann et al., 2016; Juslin et al., 2008; Karlsson et al., 2007).

Method

The method is the same as in the multiplicative environment, except that (i) the additive function $c = 5 \cdot x_1 + x_2$ was used; (ii) the stimuli 13, 24, 32, and 41 were selected as exemplars; and (iii) the final sample included $N = 18$ participants (12 women; age: $M = 21$ years, $SD = 7$ years; one additional participant did not complete the experiment due to eye tracking problems). The experiment was discontinued at $N = 18$ participants, because only little looking-at-nothing and similarity usage were observed. More variability in both measures would have been needed for reliable estimates of the association between them.

Results

Exemplar representation

As in the multiplicative environment, participants learned the exemplars well, the locations within $M = 4$ blocks ($SD = 1$) and the criterion values within $M = 6$ blocks ($SD = 2$). The accuracy remained high after training, with $M = 94\%$ ($Mdn = 100\%$, $SD = 12$ percentage points) for the locations and $M = 80\%$ ($Mdn = 95\%$, $SD = 28$ percentage points) for the criterion values. As accuracy was generally very high, no association between location and criterion accuracy was found, $\tau = $ -.05, $z = $ -0.22, $p = $ .83.
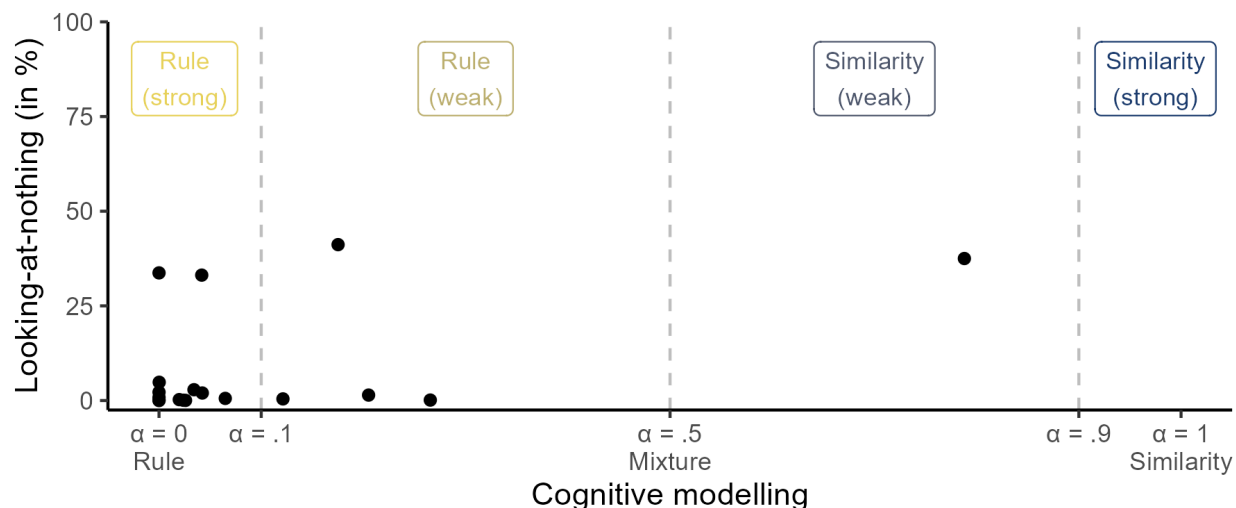
Looking-at-nothing

The test stimuli were removed after $M = 371$ ms ($SD = 222$ ms), and participants reached the response scale after another $M = 3,184$ ms ($SD = 2,436$ ms). Only $M = 20\%$

of the trials (SD = 31 percentage points) contained looking-at-nothing[4], and a participant's average looking-at-nothing duration across trials was only M = 9% (Mdn = 1%, SD = 15 percentage points across participants). Looking-at-nothing was comparable for old and new test stimuli (M_{old} = 10% ≈ M_{new} = 9%, τ = .65, z = 3.58, p < .001) and was stable within participants, with stimulus-wise splithalf correlations across participants of M = .81 (SD = .12). Thus, most participants consistently showed no looking-at-nothing.

Association between looking-at-nothing and cognitive modeling

The estimated values for α were low (M = .10, SD = .19), indicating strong rule usage, V = 2, p < .001. Six participants had an α = 0 (pure rule users), and only one participant had an α > .5. This similarity user also showed substantial looking-at-nothing (M = 38%). However, across participants, there was no credible association between their α estimates and mean looking-at-nothing, τ = 0.10, p = .60. One reason for this is the insufficient variability both in α and looking-at-nothing, illustrated in Fig. B1.

Figure B1

No association between α and looking-at-nothing in the additive experiment. Each point shows the mean estimated α value and the mean looking-at-nothing of one participant.

---

[4] As in the main study, looking-at-nothing was computed as the duration spent looking at any exemplar location divided by the summed duration spent looking at the exemplar plus test stimulus locations.

Appendix C

Model Comparison

To further test the validity of the α parameter combining a rule and a similarity-based process, we compared the RulEx-J model to a pure similarity model and a pure rule model. Both models were identical to RulEx-J, except that we fixed $\alpha = 0$ for the rule model and $\alpha = 1$ for the similarity model (which thus additionally did not have the intercept parameter β_0). As for RulEx-J, all parameters were estimated by maximum likelihood using the 16-fold cross-validation, in which each test stimulus was once used for out-of-sample prediction. The parameter estimates were averaged by participant across the 16 folds with the mean, and Table C1 shows the distribution of the resulting values across participants.

Table C1

Mean model parameter estimates (standard deviation in brackets) across participants.

Model	Additive Environment ($N = 18$)					Multiplicative Environment ($N = 48$)				
	α	β_0	β_1	β_2	σ	α	β_0	β_1	β_2	σ
RulEx-J	0.10	-0.99	4.72	1.81	1.56	0.40	-5.58	3.72	3.72	2.79
	(0.19)	(3.57)	(1.35)	(1.10)	(1.37)	(0.32)	(5.68)	(1.84)	(2.02)	(1.67)
Rule	-	-0.47	4.71	1.57	1.60	-	-3.93	3.73	2.81	3.40
		(3.47)	(1.18)	(1.02)	(1.39)		(5.08)	(1.57)	(1.22)	(1.59)
Similarity	-	-	8.64	0.06	2.86	-	-	5.71	6.03	3.94
			(2.65)	(0.20)	(1.69)			(3.01)	(2.76)	(1.87)

Note. The standard deviations are computed across participants, using each participant's mean parameter estimates across folds. The similarity model does not contain parameter β_0 and implicitly assumes $\alpha = 1$, while the rule model implicitly assumes $\alpha = 0$.

To study the out-of-sample performance, we computed the log likelihoods from the out-of-sample predictions of the RulEx-J model, the similarity model, and the rule model.

For each participant and model, we transformed the median out-of-sample log likelihood across folds into a model evidence strength (i.e., Akaike weight, Wagenmakers & Farrell, 2004). Table C2 reports the mean evidence strengths across participants and additional fit indices; Figure C1 displays the individual evidence strengths. For almost every participant, there is substantial evidence for RulEx-J, but also for one of the alternative models, namely the rule model for the participants with $\alpha < .5$ on the left, and the similarity model for the participants with $\alpha > .5$ on the right (this concerns only one participant in the additive environment, panel a, but multiple participants in the multiplicative environment, panel b).
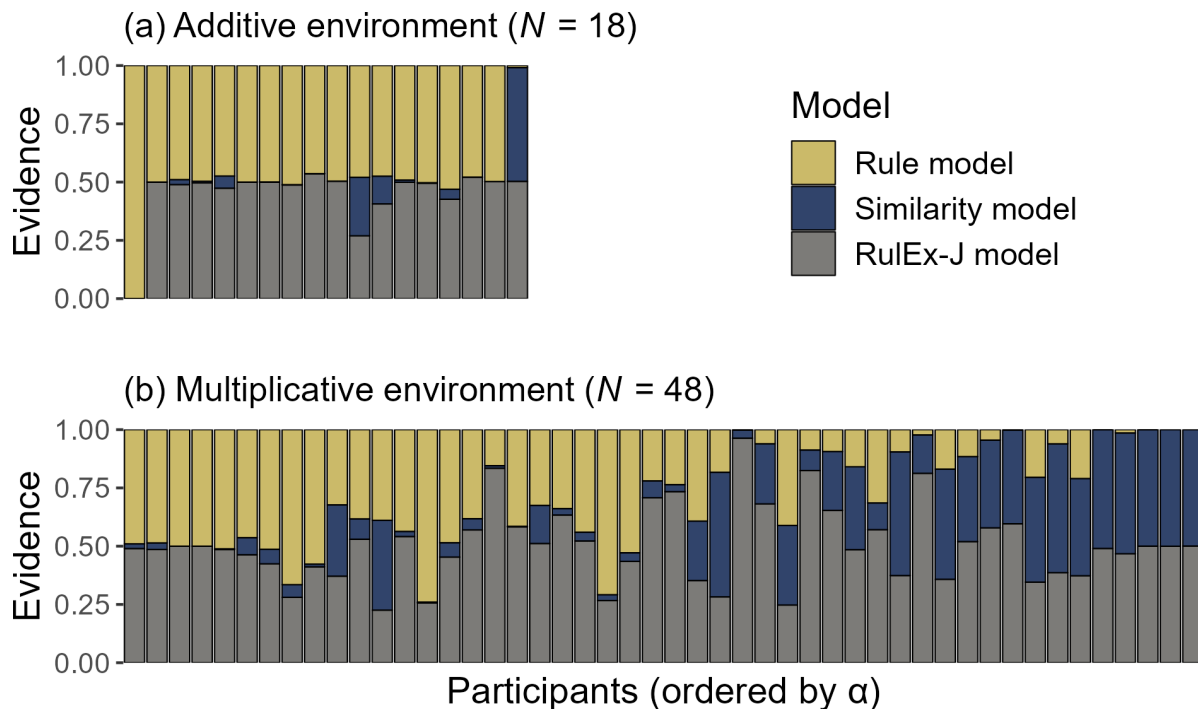
Figure C1

Evidence strengths (Akaike weights) of RulEx-J, a rule model, and a similarity model in the (a) additive and (b) multiplicative studies. Each participant is one bar, and the evidence of the different models is stacked on top of each other, summing to 1 for each participant. The x-axis orders participants by their α value from smallest (left) to largest (right).

Table C2

Descriptive out-of-sample model fit measures aggregated across participants.

Model	Additive Environment ($N = 18$)					Multiplicative Environment ($N = 48$)				
	w_{AIC}	M_ℓ	SD_ℓ	*MAE*	*RMSE*	w_{AIC}	M_ℓ	SD_ℓ	*MAE*	*RMSE*
RulEx-J	.45	-3.67	28.77	1.19	1.71	.50	-15.83	11.65	2.47	3.19
Rule	.49	-3.41	30.61	1.20	1.72	.29	-19.75	4.74	3.19	3.91
Similarity	.06	-16.71	4.55	2.22	2.90	.21	-18.12	11.74	3.19	4.22

Note. Fit measures are computed at the aggregate level: w_{AIC} = mean evidence strength based on Akaike weights, ℓ = participant-wise median out-of-sample log likelihood across folds (summarized across participants with the mean M_ℓ and the standard deviation SD_ℓ), *MAE* = mean absolute error, *RMSE* = root mean square error.