Memory shapes judgments: Tracing how memory biases judgments by inducing the retrieval of exemplars

Rosner, Agnes ; von Helversen, Bettina

DOI: https://doi.org/10.1016/j.cognition.2019.05.004

Originally published at:
Rosner, Agnes; von Helversen, Bettina (2019). Memory shapes judgments: Tracing how memory biases judgments by inducing the retrieval of exemplars. Cognition, 190:165-169.
DOI: https://doi.org/10.1016/j.cognition.2019.05.004

This paper is an accepted manuscript and not the copy of record and may not exactly replicate the final, authoritative version of the article. The final article is available at: https://www.sciencedirect.com/science/article/pii/S0010027719301283?dgcid
Supplemental materials can be found at: https://ars.els-cdn.com/content/image/1-s2.0-S0010027719301283-mmc1.pdf

Cite as:
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

Memory shapes judgments:

Tracing how memory biases judgments by inducing the retrieval of exemplars

Agnes Rosner and Bettina von Helversen

University of Zurich

Author Note

Correspondence concerning this article should be addressed to Agnes Rosner, Department of Psychology, University of Zurich, Binzmuehlestrasse 14 Box 19, 8050 Zurich, Switzerland. E-mail: agnes.rosner@psychologie.uzh.ch

Declarations of interest: Both authors gratefully acknowledge the support of the Swiss National Science Foundation, Switzerland (grant no. 157432).

Highlights

- Eye movements to "nothing" can be used to observe exemplar retrieval from memory
- We study whether memory biases in judgment can be explained by exemplar retrieval
- We show that eye movements to "nothing" predict judgments
- Eye movements mediate the effect of biasing cues on judgments
- The study quantifies the biasing effect of memory on judgments

Abstract

When making judgments (e.g., about the quality of job candidates) decision makers should ignore salient, but unrepresentative information (e.g., the person's name). However, research suggests that salient information influences judgments, possibly because memories of past encounters with similar information are integrated into the judgment. We studied eye movements to trace the link between the retrieval of past instances and their influence on judgments. Participants were more likely to look at screen locations where exemplars matching items on a name attribute had appeared, suggesting the retrieval of exemplars. Eye movements to exemplar locations predicted judgments, explaining why names influenced judgments. The results provide insights into how exemplars are integrated into the judgment process when assessing memory retrieval online.

Keywords: multi-cue judgments, retrieval, bias, similarity, eye movements, process tracing

1. Introduction

When making judgments, people often rely on information retrieved from memory (Dougherty, Gettys, & Odgen, 1999; Hastie & Park, 1986). If salient, but irrelevant or unrepresentative information is retrieved, it can bias judgments. For instance, people evaluate others who resemble their significant others more positively (Bailenson, Iyengar, Yee, & Collins, 2008), possibly because memories of the similar, but for the judgment irrelevant, person were activated (see also Brooks, Norman, & Allen, 1991, Hahn, Prat-Sala, Pothos, & Brumby, 2010). A link between memory retrieval and judgments seems likely. However, so far it remained on a theoretical level, because it was not possible to directly observe what information was retrieved. The goal of this study is to provide this direct link between the retrieval of exemplars and judgment biases by making exemplar retrieval observable. Establishing this link is important because it allows measuring retrieval processes independently from judgments thereby quantifying the influence of memory on judgments.

To this goal, we focused on a multiple-cue judgment task, in which a biasing effect of retrieved exemplars was postulated (von Helversen, Herzog & Rieskamp, 2014). In multiple-cue judgments, people form a quantitative judgment by integrating information from several attributes (i.e., cues), for instance, evaluating the suitability of a job candidate based on several attributes. In general, people rely on two cognitive processes when making multiple-cue judgments: cue abstraction and exemplar memory (Hoffmann, von Helversen, & Rieskamp, 2014; Juslin, Olsson, & Olsson, 2003; Juslin, Karlsson & Olsson, 2008). Cue abstraction theories assume that people (1) abstract the relationship between cues and criterion, i.e., the relative importance of language skills for a specific job (2) make judgments by a linear additive integration of the weighted cues (e.g., Brehmer, 1994).

In contrast, exemplar-based judgment processes assume that exemplars are stored as long-term memory representations that combine past instances with criterion values (e.g., Nosofsky, 1986). Similar exemplars are retrieved when a new instance is evaluated. Often, it is assumed that people rely either on exemplar-based or cue abstraction processes, but exemplar similarity may influence judgments even when people are following rules. For instance, von Helversen and colleagues (2014) found that judgments of new candidates were biased toward the exemplar they resembled most—even though participants integrated attributes following a cue abstraction process.

To test the link between exemplar retrieval and judgments, we studied eye movements using the so-called looking-at-nothing behavior (Ferreira, Apel, & Henderson, 2008; Richardson, Altmann, Spivey, & Hoover, 2009). That is, when retrieving information from memory (e.g., cue values of job candidates), people look back to associated but emptied spatial locations on a computer screen, a behavior that has been widely applied to study memory retrieval during judgment and decision-making (e.g., Jahn & Braatz, 2014; Martarelli, Chiquet, Laeng, & Mast, 2017; Pärnamets, Johansson, Gidlöf, & Wallin, 2015; Platzer, Bröder, & Heck, 2014; Renkewitz & Jahn, 2012; Scholz, Krems, & Jahn, 2017; Scholz, von Helversen, & Rieskamp, 2015). For instance, in the study by Scholz et al. (2015), the more similar an exemplar was to a new job candidate, the longer participants gazed toward the empty spatial location associated with that exemplar, when using an exemplar-based strategy. In comparison, when using cue abstraction, people did not look at exemplar locations, demonstrating that looking-at-nothing allows studying exemplar retrieval from memory. Here, we aim to go one step further and show that eye movements are directly linked to judgment(-biases) through the retrieval of exemplars.

In the following, we test whether salient, but in this context usually irrelevant cues lead to similar biases in multi-attribute judgments from memory to those observed in previous studies (e.g., von Helversen et al., 2014). In this case, judgments for test items containing a match on this cue should systematically deviate from items with the same cue-criterion relationship but without a match on this cue. Second, we test whether eye movements reflect exemplar retrieval from memory. When judging test items, we expect longer fixation durations on more similar exemplars (see Scholz et al., 2015). Third, we test whether retrieval of exemplars can explain judgments, assuming that the longer people gaze toward the empty spaces where exemplars had been presented previously, the more their judgments should be influenced toward the criterion of the exemplar that occupied the space toward which they are looking.

2. Method

Participants' task was to judge the suitability of job candidates applying for an open position in a fictional company. Therefore, they first learned information about previous job candidates (training exemplars) who had applied for a similar position and their ratings (criterions). During test trials, new test candidates had to be judged according to their suitability. Test candidates differed in their similarity in cue values to the exemplars and in their names[1].

2.1 Participants

Fifty students (43 female, 7 male, M_{age} = 22.83 years, range 18–28 years) at the University of Zurich participated in the experiment for course credit or financial compensation [15 Swiss francs (CHF) per hour]. In addition, they could earn a bonus depending on their performance (M = 8.31 CHF). All participants had normal or corrected-to-normal vision. Mean track accuracy was, at M = 0.63° of visual angle, very high. All participants signed informed consent forms.

We chose a sample size of 50 to have appropriate power to find an effect of medium size for similarity on eye movements (η_p^2 =. 06 requires N= 48 to reach a power of 95%, see Faul, Erdfelder, & Buchner, 2007).

2.2 Apparatus and Materials

[1] All experimental materials and data are available at https://osf.io/g9fqz/ Details on subjects' performance during the task, methods, and statistical analyses can be found in the supplemental materials.

Participants were seated in front of a 22-inch computer screen (1,680 × 1,050 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 500Hz 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. We analyzed data with mixed models (using R, R Core Team, 2017, and packages: *lme4*, Bates et al., 2017; *afex*, Singmann, Bolker, Westfall, & Aust, 2018; *emmeans*, Lenth, 2018).

Study materials consisted of four training exemplars and 36 additional test items, see Table S1 in the supplemental materials. All items contained information on three cues: work experience, language skills, and professional training. Cue values ranged from very poor, poor, average, good, to very good. Item names were popular male names in the German-speaking region of Switzerland between 1980 and 2015 (www.bfs.admin.ch). Criterion values followed a linear function of the three cues assuming an equal weighting of the cues. Names were not necessary for accurately judging the suitability. The judgment scale ranged from 0 (very poor) to 60 (very good). Two exemplars had a low criterion (15) and two a high criterion (45). A small error was introduced to the exemplar criterion values of plus/minus two points and balanced between participants. Items could either have the same *name* as a low- or high-performing exemplar (*-Name*: 9 items, *+Name*: 9 items) or they could have a different name (*No-Match*: 22 items). Test items varied in their *similarity* to the exemplars, ranging from zero to three matches in cue values with one exemplar (*0:* 6, *1:* 16, *2:* 10, and *3:* 8 items).

2.3 Procedure

2.3.1 Exemplar learning. Participants learned cue information about the four exemplars presented in rectangles on the screen (see Figure 1A). During learning trials, only the emptied rectangles were visible and a name or cue value was presented auditorily.

Participants had to indicate the rectangle to which the name or cue value belonged and received visual and auditory feedback for their response. Learning continued until participants remembered all names and cue information (4 names + 12 cue values = 16 pieces of information) correctly in two consecutive blocks or after 20 blocks. One block consisted of a test of all four exemplars, that is, 16 trials.

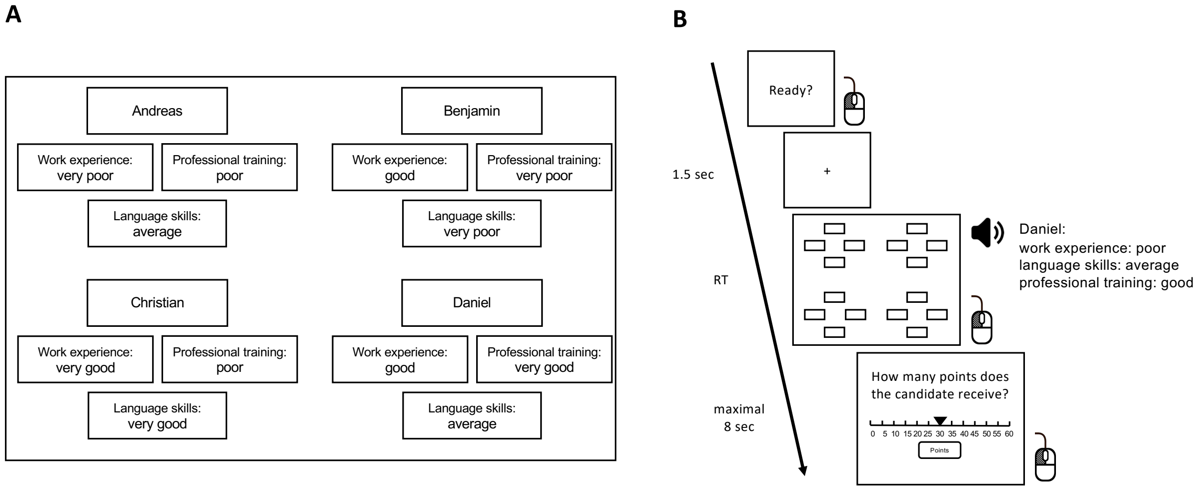

Figure 1. A) Learning materials consisting of information about four training candidates (two high- and two low-performing exemplars). Each exemplar was presented in one of the four screen quadrants. The distance from the center of the screen to the center of each of the four exemplars was 7.04° of visual angle (414 pixels). Names and cue values were presented as black text in rectangles with black borders and a light grey background. Each rectangle had a size of 2.55° × 1.28° of visual angle (150 × 75 pixels). The center of each of the four rectangles containing the information describing one exemplar had a distance of 2.72° of visual angle (160 pixels) from the center of each quadrant. Visual materials were presented in four balanced orders, varying the positions of the exemplars on the screen and the order of cue values between participants. Note that the size of the rectangles containing cue information in the Figure was increased to increase readability and the cue information translated into English. B) Experimental procedure during test trials. Auditory stimuli consisted of names, cue dimensions, and cue values. Items were always presented with the names first, followed by the cue values from left to right as visually presented on the screen during exemplar

learning, thereby varying the presentation order between participants. For the gaze analyses, we drew four rectangular areas of interest (AOIs) around each of the four exemplar locations. Each exemplar AOI had a size of 8.93° × 6.38° of visual angle (525 × 375 pixels). The size of the exemplar AOI exceeded the outer borders of each of the four rectangles describing one exemplar by 0.64° of visual angle, that is, half the height of one rectangle.

2.3.2 Criterion learning. Participants learned how each exemplar had been evaluated in the past. In each trial, participants were auditorily presented with the name and cue values of one of the four exemplars while looking at the empty rectangles (see Figure 1B) and could think about their judgments. Once they decided on a judgment they proceeded to a separate screen where they entered their judgment on a rating scale and received visual and auditory feedback on the correctness of their judgments. We limited the time to enter the ratings to eight seconds, to ensure that participants would only proceed when they had determined their judgment. This was to avoid that decision-making continued while participants were looking at the rating scale, and thus to capture as much of the judgment process in participants' looking-at-nothing behavior as possible. Criterion learning ended after participants correctly judged all four exemplars in two consecutive blocks or after ten blocks. One block consisted of a test of the four exemplar criterion values, that is, four trials.

2.3.3 Test phase. The test phase followed the same procedure as the criterion learning phase, but without feedback. Overall, participants judged the 40 test items (four exemplars and 36 test items) twice in a random sequence in two blocks, resulting in 80 trials per person.

3. Results

Before testing whether exemplar retrieval can be directly linked to judgment biases, we assessed two presumptions: (1) the names manipulation influences judgments, (2) participants gaze longer at more similar exemplars, indicating exemplar retrieval.

3.1 Preparatory data analyses

For all participants and test trials, 190 trials were removed (4.8 % of all trials): In 22 trials (0.6 % of trials), no gaze data was recorded (e.g., due to eyes closed, looking off the screen), and in 168 trials (4.2 % of all trials), no rating was recorded because participants took longer than eight seconds to provide their judgment.

3.2 Names bias judgments

Names biased judgments similarly to faces (see von Helversen et al., 2014). Linear mixed model analyses[2] revealed a significant main effect for the factor name (-Name, No-Match, +Name): $F(2,115.3) = 5.80$, $p = .004$. Judgments were the lowest for items with the same name as low-performing exemplars ($EMM_{-Name} = 27.81$) and highest for items with the name of high-performing exemplars ($EMM_{+Name} = 28.84$). Judgments for items with a different name lay in the middle ($EMM_{No\text{-}Match} = 28.66$). Pairwise comparisons revealed significant differences between the -Name and No-Match ($p = .001$) and between the -Name and +Name conditions ($p = .008$), but not between the No-Match and +Name conditions ($p = .47$). Judgments were also significantly predicted by the cue values of the three cues (all $ps < .001$).

3.3 Eye movements reflect exemplar retrieval

Fixation proportions per trial were calculated based on the number of fixations on four exemplar AOIs (see Figure 1). For each test item, we determined fixations on the most similar exemplar, defined as the exemplar with the highest number of matching cues (the relevant exemplar). If two exemplars were equally similar to the test item, fixation proportions for one of them was chosen randomly[3]. Similarity varied between 0 and 3 matches.

[2] Details on the fixed and random effects structure, the parameter fitting procedure and calculation of degrees of freedom for this and all following analyses are described in the supplemental materials.

[3] This was the case for six items sharing one cue value with two exemplars. See Table S1 in the supplemental materials, see also https://osf.io/g9fqz/

If test items matched the name of an exemplar, fixation proportions were increased, main effect name: $\chi^2(2) = 80.09$, $p < .001$ (Figure 2). Additionally, the more a test item resembled the relevant exemplar, the higher fixation proportions toward the relevant exemplar, main effect similarity: $\chi^2(3) = 15.38$, $p < .002$, replicating the results of Scholz et al. (2015). The factors name and similarity interacted, $\chi^2(6) = 170.25$, $p < .001$, with a significant linear increase of fixation proportions as a function of similarity only in the No-Match condition, *Estimate* = 2.74, *SE* = 0.47, $z = 5.80$, $p < .001$. There was no difference between the –Name and +Name conditions, *Estimate* = -0.19, *SE* = 0.18, $z = -1.06$, $p = 0.29$. These results show that (1) when the name matched an exemplar, fixation proportions toward that exemplar were relatively high over all levels of similarity, and (2) it did not matter if the name matched a high- or low-performing training candidate.

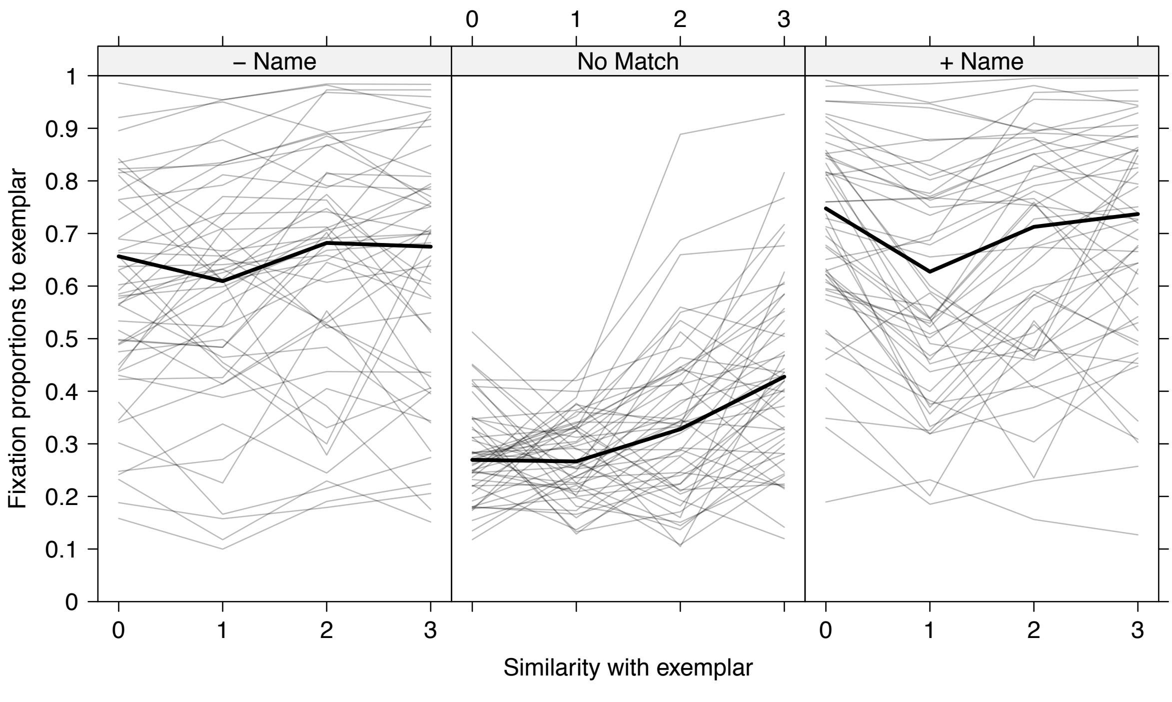

Figure 2. Two-way interaction of similarity × name match on fixation proportions toward the relevant exemplar (exemplar with maximal no. of shared cues). The black lines show predictions based on fixed effects. Gray lines in the background show predictions of the individual random effects from all participants.

3.4 Exemplar retrieval predicts judgments

Fixation proportions to high-performing exemplars (summed fixation durations for the two locations associated with high-performing exemplars divided by the summed fixation durations to all locations) significantly predicted judgments, main effect fixation proportions: $F(1,66.6) = 18.43$, $p < .001$. The longer participants looked at the empty rectangles containing the cue information of the high-performing exemplars, the higher their judgments (Figure 3). Like in the first analysis, names as well as cue values determined participant's judgments, Name: $F(2,182.7) = 3.99$, $p = .02$, Cues: all *ps* < .001. There was no interaction between name and fixation proportions: $F(2,1791.1) = 2.54$, $p = .08$, indicating that there was no significant difference in fixation proportions to high-performing exemplars and judgments between the name conditions.

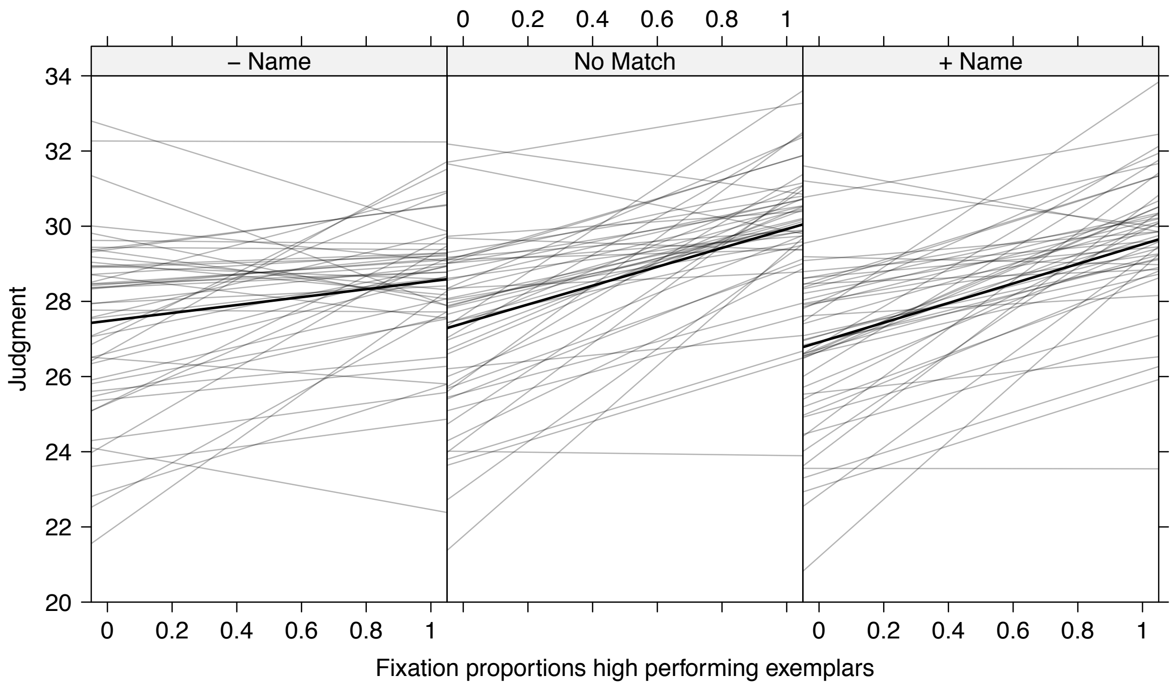

Figure 3. Relation of fixation proportions to the two high-performing exemplars and judgments in the three name conditions. The black lines show predictions based on fixed effects. Gray lines in the background show predictions of the individual random effects from all participants.

Lastly, we tested if the effect of name on judgments can be explained by the retrieval of exemplars from memory with a mediation analysis for multilevel data (Imai, Keele, and Tingley, 2010; Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). We estimated the mediation effect for the contrast between low- and high-performing names. The effect of name on judgments could be accounted for by fixation proportions (see Figure 4) with a high overall proportion of 0.58 mediated, $p = .002$.

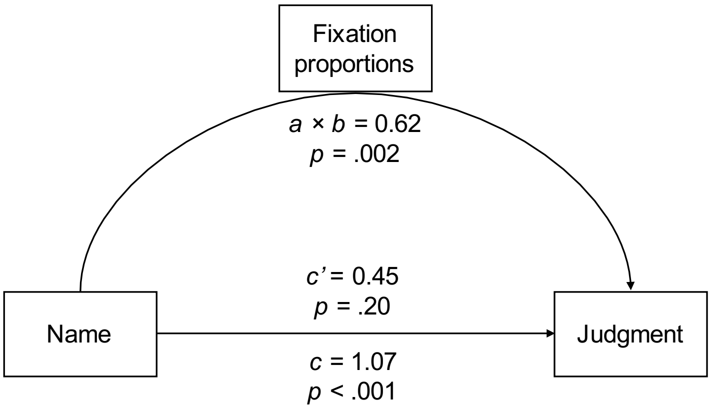

Figure 4. Mediation model with one mediating variable. Displayed is the *total effect* of name on judgments (c), the *direct effect* of name on judgments in the model also allowing for mediation by fixation proportions (c'), and the *indirect effect* of name on judgments via fixation proportions (a × b). Unstandardized coefficients are reported.

4. Discussion

Many studies assume a link between memory retrieval and judgments, but often without providing independent measures of the retrieval process. The purpose of this study was to provide direct evidence for the relation between the retrieval of exemplars and judgments. To test this, we studied eye movements to observe retrieval processes during multi-attribute judgments.

When participants judged the suitability of job candidates based on cue information such as work experience, we found that increasing similarity by additionally presenting a salient attribute like the name of the candidate, increased fixations to the name matching exemplar. Previous research proposed that eye fixations to associated, but emptied spatial locations indicate memory retrieval (e.g., Richardson & Spivey, 2000), and when applied to multi-attribute decision-making, they can reflect exemplar retrieval (Scholz et al., 2015). Accordingly, the results of this study suggest that the name increased the probability that an exemplar with the same name was retrieved and this biased judgments in the direction of the retrieved exemplar. This interpretation is in line with the result that participants rated candidates worse than similarly qualified candidates when they had the same name as a low-performing previous candidate. Although, judgments were increased for items sharing a name with a high-performing exemplar in comparison to items with a new name, this difference was not significant — possibly because participants tended to give relatively low ratings. Mean ratings in all name conditions were below 30, the midpoint of the judgments scale.

Importantly, we could establish a direct link between exemplar retrieval measured through eye movements and judgments. The more participants gazed back toward high-performing exemplars, the higher their suitability ratings. Furthermore, a mediation analysis showed that fixation proportions accounted for the effect of names on judgments. One interpretation is that the differences in likelihood with which exemplars were retrieved led to the effect of names on judgments.

The results provide first direct evidence for a tight interplay between retrieval processes and judgment biases when salient, but potentially irrelevant information is retrieved. There are two limitations to the interpretation of these results. First, some participants may have used the candidate's name as a deliberative cue, even though in a job context names are unlikely to be relevant for judging suitability and considering them was not necessary to make accurate judgments. Still, during the criterion learning phase participants could perform accurately by only considering the name. Thus, we cannot rule out that participants deliberatively used the name in the test phase. Second, this study does not allow to draw a causal conclusion between memory retrieval and judgment bias. For instance, a tendency to give a high judgment could lead participants to retrieve a high performing exemplar. Nevertheless, our findings demonstrate the importance of memory for understanding judgment processes in general (Dougherty et al., 1999) and particularly the influence of exemplars (Juslin et al., 2008) in producing the observed biases.

Not only names, but also cue values of test candidates determined judgments. This suggests that participants did not just use criterion values of retrieved exemplars, but integrated them with a cue-based judgment process. Furthermore, eye movements were also related to judgments when names were new and thus did not influence exemplar retrieval. In these cases, retrieval was a function of similarity between the test item and the exemplars suggesting that exemplar retrieval is an inherent part of the judgment processes.

In this study, cue information about exemplars was learned by heart. This could lead to an increased influence of exemplars on judgments, because exemplars were highly available in memory. Although, in the study by Scholz et al. (2015), in which exemplar information was learned by heart, people used other strategies (e.g., cue abstraction), future research should test, if the same judgment and gaze behavior can be observed, when participants are only trained on the learning criterion and not explicitly trained on cue values of exemplars. This would also make the task more similar to standard multiple-cue tasks.

Besides its merits, memory can be a source of bias during judgments when the information that is activated should be irrelevant for the judgment at hand. We showed how this effect of memory could be directly observed by studying looking-at-nothing behavior.

5. Acknowledgments

Both authors gratefully acknowledge the support of the Swiss National Science Foundation, Switzerland (grant no. 157432). Furthermore, the authors thank Henrik Singmann for helpful comments on mixed effects modeling and Ramona Obrist for collecting the data.

References

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., … Eigen, C. (2017). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–113. http://doi.org/10.18637/jss.v067.i01

Bailenson, J. N., Iyengar, S., Yee, N., & Collins, N. A. (2008). Facial similarity between voters and candidates causes influence. *Public Opinion Quarterly, 72(5),* 935-961.

Brehmer, B. (1994). The psychology of linear judgement models. *Acta Psychologica*, *87(2-3)*, 137–154.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General*, *120*(3), 278–287.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory process model for judgments of likelihood. *Psychological Review, 106,* 180–209.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191.

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

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

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review, 93(3),* 258.

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,* 2242–2267. https://dx.doi.org/10.1037/a0037989

Imai K, Keele L, Tingley D (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15(4),* 309–334.

Jahn, G., & Braatz, J. (2014). Memory indexing of sequential symptom processing in diagnostic reasoning. *Cognitive Psychology*, *68*, 59–97. http://doi.org/10.1016/j.cogpsych.2013.11.002

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*, 259–298. http://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*, 133–156. http://doi.org/10.1037/0096-3445.132.1.133

Lenth, R. (2018). emmeans: Estimated marginal means, aka least-squares means. R package version 1.1.3. https://CRAN.R-project.org/package=emmeans

Martarelli, C. S., Chiquet, S., Laeng, B., & Mast, F. W. (2017). Using space to represent categories: insights from gaze position. *Psychological Research*, *81*, 721–729. http://doi.org/10.1007/s00426-016-0781-2

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

Pärnamets, P., Johansson, R., Gidlöf, K., & Wallin, A. (2015). How information availability interacts with visual attention during judgment and decision tasks. *Journal of Behavioral Decision Making*, *29,* 218-231.http://doi.org/10.1002/bdm.1902

Platzer, C., Bröder, A., & Heck, D. W. (2014). Deciding with the eye: How the visually manipulated accessibility of information in memory influences decision behavior. *Memory & Cognition*, *42*, 595–608. http://doi.org/10.3758/s13421-013-0380-z

R Core Team (2017). R: A language and environment for statistical computing. R foundation for statistical computing Vienna, Austria. URL: https://www.R-project.org/.

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*, 1622–1639. http://doi.org/10.1037/a0028073

Richardson, D. C., Altmann, G. T. M., 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,* 235-236. http://dx.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*, 269–295. http://doi.org/10.1016/S0010-0277(00)00084-6

Scholz, A., Krems, J. F., & Jahn, G. (2017). Watching diagnoses develop: Eye movements reveal symptom processing during diagnostic reasoning. *Psychonomic Bulletin and Review*, *24*, 1398-1412. http://doi.org/10.3758/s13423-017-1294-8

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

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). afex: Analysis of Factorial Experiments. R package version 0.20-2. https://CRAN.R-project.org/package=afex

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59(5)*, 1-38.

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