From recognition to decisions: Extending and testing recognition-based models for multialternative inference

JULIAN N. MAREWSKI, WOLFGANG GAISSMAIER, AND LAEL J. SCHOOLER
Max Planck Institute for Human Development, Berlin, Germany

DANIEL G. GOLDSTEIN
London Business School, London, England

AND

GERD GIGERENZER
Max Planck Institute for Human Development, Berlin, Germany

The recognition heuristic is a noncompensatory strategy for inferring which of two alternatives, one recognized and the other not, scores higher on a criterion. According to it, such inferences are based solely on recognition. We generalize this heuristic to tasks with multiple alternatives, proposing a model of how people identify the consideration sets from which they make their final decisions. In doing so, we address concerns about the heuristic's adequacy as a model of behavior: Past experiments have led several authors to conclude that there is no evidence for a noncompensatory use of recognition but clear evidence that recognition is integrated with other information. Surprisingly, however, in no study was this competing hypothesis—the compensatory integration of recognition—formally specified as a computational model. In four studies, we specify five competing models, conducting eight model comparisons. In these model comparisons, the recognition heuristic emerges as the best predictor of people's inferences.

In the autumn of 2008, eight Democratic and nine Republican candidates in the U.S. primaries had invested millions of dollars in the hopes of embedding their names in the recognition memory of American voters. It is likely that elevated name recognition increased their chances of being included in voters' consideration sets of promising candidates. At the same time, not only in the United States but all over the globe, people were wondering who would emerge from the election as a winner.

In this article, we examine a number of strategies that can be relied on to make inferences about unknown quantities or uncertain future events, such as forecasting which candidates voters are most likely to favor in an election. One such strategy is the *recognition heuristic* (Goldstein & Gigerenzer, 2002). It can be applied to infer which of two alternatives, one recognized and the other not, has a larger value on a given criterion. The heuristic reads as follows: "If only one of two alternatives is recognized, infer the recognized one to be larger."

The recognition heuristic operates on an ecological sense of recognition—namely, on our ability to discriminate between stimuli that we have encountered in our environment before (e.g., famous names, such as "Clinton"), and those that we have neither seen nor heard before (e.g.,

unknown names, such as "Schauk"). As we will explain below, in doing so the heuristic can help us make accurate judgments in many domains. For instance, our recognition of universities allows us to predict their quality (Hertwig & Todd, 2003); our recognition of soccer teams and tennis players can be used to forecast their future success in competitions (Herzog & Hertwig, 2010; Pachur & Biele, 2007; Serwe & Frings, 2006); and our recognition of billionaires and musicians reflects their fortunes and record sales, respectively (Hertwig, Herzog, Schooler, & Reimer, 2008). Besides being useful, recognition is also easily accessible and remarkably lasting (Pachur & Hertwig, 2006; Shepard, 1967; Standing, 1973). For example, when scanning long supermarket shelves in an attempt to infer which of 20 shampoos are the better products, we are likely to quickly detect which brands we have heard of before. As pointed out by Goldstein and Gigerenzer (2002) and Pachur and Hertwig, these remarkable characteristics make it likely that recognition plays an important role in a multitude of tasks; and, in fact, there is evidence that reasoning by recognition is a common strategy not only in humans (Galef, 1987). Yet the recognition heuristic, as formulated for making inferences about two alternatives, is of little help when sets of three or more are evaluated. The first

J. N. Marewski, marewski@mpib-berlin.mpg.de

contribution of this article is to examine how well this heuristic can be generalized to *N* alternatives—a generalization that helps explain how people form the consideration sets from which they make their final choice. This is a key question in marketing, politics, and beyond.

Elimination by Recognition

Tversky's (1972) classic *elimination by aspects* model pares down a large set of alternatives by eliminating them on the basis of probabilistically selected criteria. Similarly, in the marketing literature, many theories of choice assume a two-stage process: When we are evaluating multiple alternatives, such as which of eight candidates to vote for, or which of 20 shampoos to buy, a smaller set of relevant alternatives is formed first, then a choice is made after more detailed examinations of the alternatives in this *consideration set* (Alba & Chattopadhyay, 1985; Hauser & Wernerfelt, 1990; Howard & Sheth, 1969). When recognition correlates with the criteria on which alternatives are evaluated, an elimination-by-recognition model could form "consideration sets" consisting of recognized alternatives. Specifically, if generalized to the multialternative case, the recognition heuristic reads as follows: "If there are *N* alternatives, then rank all *n* recognized alternatives higher on the criterion than the *N*−*n* unrecognized ones." Once they are identified, in a second stage recognized alternatives can be ranked with heuristics that use alternatives' attributes as *cues*—say, knowledge about a candidate's party affiliation, or a shampoo's ingredients.

Consideration sets facilitate decisions by reducing the number of alternatives. To illustrate, a voter may want to forecast the final rank order of eight Democrats in the primaries. But there are a total of 8 factorial (40,320) possible rank orders. In contrast, if the recognition heuristic is used, and, say, three candidates are recognized and five unrecognized, then only 3 factorial (6) ranks need to be considered. Unrecognized alternatives can be put aside (or ranked at random), because they are likely to score low on the criterion, and people typically know nothing about them in the first place. By ignoring the unheard-of and unrecognized, the recognition heuristic reduces complexity without necessarily harming accuracy.

From Recognition to Decisions:
A Competition Among Models

In contrast to the recognition heuristic, many models of consideration-set identification posit that people evaluate alternatives by weighting and adding their values on a range of cues (e.g., Hauser & Wernerfelt, 1990; Roberts & Lattin, 1991). The assumption is that an alternative's low value on one cue can be *compensated for* by a high value on another cue.

However, there is evidence that people do not always make trade-offs (e.g., Einhorn, 1970; Fishburn, 1974; Hogarth, 1987; Payne, Bettman, & Johnson, 1993; Yee, Dahan, Hauser, & Orlin, 2007). For instance, in a review of 45 process-tracing studies, Ford, Schmitt, Schechtman, Hults, and Doherty (1989) concluded that noncompensatory processes are the rule and compensatory processes are nearly always observed in situations with few alternatives

and cues. Surprisingly, by ignoring information rather than integrating it all, noncompensatory heuristics can yield more accurate judgments than compensatory ones can (e.g., Gigerenzer & Goldstein, 1996) while at the same time simplifying tasks (e.g., Einhorn, 1970; Simon, 1955).

The recognition heuristic is a noncompensatory model. Even when other cues are retrieved, when the heuristic is used these cues are ignored. In initial experiments and follow-up studies, a majority of people made inferences in accordance with the heuristic's predictions (e.g., Goldstein & Gigerenzer, 2002; Pachur & Biele, 2007; Pachur & Hertwig, 2006; Volz et al., 2006). However, at the same time, findings that people systematically make inferences inconsistent with the heuristic have raised doubts about its adequacy as a model of behavior (e.g., Bröder & Eichler, 2006; Dougherty, Franco-Watkins, & Thomas, 2008; Hilbig & Pohl, 2008, 2009; Newell & Fernandez, 2006; Newell & Shanks, 2004; Oppenheimer, 2003; Pohl, 2006). For instance, Richter and Späth (2006) ran a series of studies and—observing that fewer decisions were consistent with the heuristic when cues that contradicted recognition were available—concluded that there was no evidence of a noncompensatory use of recognition. According to them, there was clear evidence that recognition is integrated with knowledge.

However, such conclusions require further tests, for three reasons (see also Marewski, Schooler, & Gigerenzer, 2010). First, most studies that provided evidence against the noncompensatory processing of recognition reported group means, rather than individual participants' data; yet a number of experiments have indicated that there may be strong individual differences in people's treatment of recognition that would be hidden by group means. For instance, in Pachur, Bröder, and Marewski's (2008) studies (see also Hilbig, 2008), about half of the participants picked recognized alternatives regardless of the number of conflicting cues, consistent with the hypothesized noncompensatory processing of recognition; the remaining ones appeared to be influenced by the additional cues. Moreover, reanalyses of Richter and Späth's (2006) and Newell and Fernandez's (2006) data have also shown that the large majority of their participants made decisions in agreement with the recognition heuristic, even in the presence of contradictory cues (see Pachur et al., 2008).

Second, as discussed in detail by Pachur et al. (2008; see their Table 1 and pp. 187–190), most studies on the recognition heuristic differed in critical ways from the situations for which Goldstein and Gigerenzer (2002) formulated the heuristic. To give just one example: The recognition heuristic is a model for situations where people make inferences solely on the basis of information retrieved from memory (*inference from memory*, as opposed to *inference from givens*; Gigerenzer & Goldstein, 1996), and where recognition is acquired in the world prior to participation in a study. In some studies (e.g., Bröder & Eichler, 2006), recognition was experimentally induced shortly before a two-alternative forced choice (2AFC) task in such a way that the study could be easily tracked as the recognition source (for evidence that people take the source of recognition into account when making inferences from rec-

Table 1
List of Competing Models

Model Input	Decision Rule
Recognition-based Competitor 1: Recognition heuristic (Studies 1–4)	
1 recognized alternative R	
1 unrecognized alternative U	Choose R.
n recognized alternatives R_i ($i = 1, \ldots, n$)	
$N-n$ unrecognized alternatives U_j ($j = 1, \ldots, N-n$)	Rank R_i higher than U_j.
Cue-based Competitor 2: Take-one-cue (Studies 1–4)	
1 recognized alternative R with up to 1 cue with value v	If $v \leq C_1$, choose R.
1 unrecognized alternative U	If $v > C_1$, choose U.
1 cutoff criterion C_1	
n recognized alternatives R_i ($i = 1, \ldots, n$), each with up to 1 cue with value v_i	If $v_i \leq C_2$, rank R_i higher than U_j.
$N-n$ unrecognized alternatives U_j ($j = 1, \ldots, N-n$)	If $v_i > C_2$, rank U_j higher than R_i.
1 cutoff criterion C_2	
Cue-based Competitor 3: Tallying-of-negative-cues (Study 3)	
1 recognized alternative R with sum Σ of cues with negative values $v_n = -1$, and unknown values $v_u = 0$	If $\Sigma \geq C_3$, choose R.
1 unrecognized alternative U	If $\Sigma < C_3$, choose U.
1 cutoff criterion C_3	
Cue-based Competitor 4: Tallying-of-positive-and-negative-cues (Study 3)	
1 recognized alternative R with sum Σ of cues with negative values $v_n = -1$, unknown values $v_u = 0$,	
and positive values $v_p = +1$	If $\Sigma \geq C_4$, choose R.
1 unrecognized alternative U	If $\Sigma < C_4$, choose U.
1 cutoff criterion C_4	
Cue-based Competitor 5: Weighted-best-cues (Study 4)	
n recognized alternatives R_i ($i = 1, \ldots, n$), each with up to 1 cue drawn from a set of i cues with	
weights $-100 \leq w_i \leq +100$	If $w_i \geq C_5$, rank R_i higher than U_j.
$N-n$ unrecognized alternatives U_j ($j = 1, \ldots, N-n$)	If $w_i < C_5$, rank U_j higher than R_i.
1 cutoff criterion C_5	
Fluency-based Competitor 6: Weighted-fluency (Study 3)	
1 recognized alternative R with retrieval time r	If $r \leq C_6$, choose R.
1 unrecognized alternative U	If $r > C_6$, choose U.
1 cutoff criterion C_6	

Note—All competitors can also be thought of as weighted linear additive models with two classes of predictors, cues and recognition, or retrieval time and recognition, respectively. Different cutoff criteria, C_{1-6} (free parameters), measure the weight of these predictors relative to each other. The recognition heuristic can be formally represented as a special case of these models with C_{1-6} set such that the models always choose recognized alternatives, which is equivalent to assuming a noncompensatory recognition weight (see study descriptions for details; see also the General Discussion for the underlying modeling stratagem).

ognition, see Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer, 2009). Other studies focused on inference from givens (e.g., Newell & Shanks, 2004). However, even if we do not agree with some of the criticism of the recognition heuristic, we do not wish to devalue these studies; on the contrary, we consider them to be important tests of the generalizability of this heuristic across different conditions and tasks. Since the heuristic did not generalize to some of these tasks, these tests challenge researchers to identify a model capable of doing so—that is, a model that outpredicts the recognition heuristic. This leads us directly to our most important point.

Third and most critical, it is important to conduct formal model comparisons when evaluating the descriptive adequacy of theories (on some of the merits and complications of formal modeling, see Fum, Del Missier, & Stocco, 2007; Hintzman, 1991; Lewandowsky, 1993; Marewski & Olson, 2009); yet in all previous work that raised doubts about the adequacy of the heuristic as a model of behavior, the corresponding alternative hypotheses—for example, a compensatory integration of recognition and other cues—were not formally specified as testable, computational models of psychological processes. Much the same can be said of our own work: We have never reported correspond-

ing comparative model tests of the recognition heuristic. In short, except for Pachur and Biele (2007),[1] no study has ever tested a single compensatory strategy (or any other formal model of psychological processes, for that matter) against the recognition heuristic, which is needed to evaluate *both* the recognition heuristic *and* the corresponding alternative hypotheses that have been proposed. In fact, one might ask how useful it is to reject a model that can explain some data (but not all), if no other model has been shown to be better.

In this article, we open up the search for a new model. We conduct the first formal competition between the recognition heuristic and alternative compensatory models for two-alternative and multialternative inference. These models include various formalizations of the alternative hypotheses to the recognition heuristic that are discussed in the literature. All models are listed in Table 1 and are explained in the text below. In carrying out a total of eight model comparisons, we will contribute to both (1) evaluating the recognition heuristic as a model of behavior and (2) assessing the descriptive adequacy of models that have been proposed as alternatives to this heuristic. In doing so, we will also (3) explore whether there is yet another possibility besides rejecting the original recognition heu-

ristic model and/or embracing a new model. As stressed by Newell and Fernandez (2006), the findings reported by them and by many others can be interpreted in two ways. One is that they challenge the recognition heuristic's plausibility; yet another is that they point to the mechanisms that determine when people rely on the recognition heuristic and when they adopt other strategies.

Strategy Selection by Default?

No cognitive strategy is always relied upon. Rather, in keeping with many other frameworks (e.g., Beach & Mitchell, 1978; Hogarth & Karelaia, 2007; Payne et al., 1993), we assume that the mind comes equipped with a repertoire of strategies. This repertoire forms an "adaptive toolbox" of heuristics, each of which exploits how basic cognitive capacities, such as memory, represent regularities in the structure of our environment. In doing so, heuristics can yield accurate judgments by operating on little information—say, a sense of recognition (for a recent overview, see Marewski, Gaissmaier, & Gigerenzer, 2010).

Often, the study of the mechanisms determining the use of a heuristic can be informed by an analysis of the heuristic's *ecological rationality*—that is, of the environmental structures it exploits. Figure 1 illustrates the ecological rationality of the recognition heuristic in terms of three correlations. There is a criterion, an *environmental mediator*, and a person who infers the criterion. Using the heuristic is ecologically rational and helps a person make accurate inferences when there is both a substantial *ecological correlation* between the mediator and the criterion and a substantial *surrogate correlation* between the mediator and recognition. This combination can yield a substantial *recognition correlation*; that is, recognized alternatives tend to have higher criterion values than unrecognized ones do.

In what follows, we suggest that the recognition heuristic may be used by default, but this default can be overruled by information indicating that it is not ecologically rational to use the heuristic because recognition is not predictive of the criterion. There are two kinds of evidence, behavioral and neural, indicating that the heuristic could be used as a default, as opposed to being just another strategy. Pachur and Hertwig (2006) and Volz et al. (2006) reported response time (RT) data suggesting that recognition of an alternative often arises before other knowledge can be recalled. This finding is consistent with the recognition memory literature, indicating that a sense of recognition (often called *familiarity*) arrives on the mental stage earlier than does recollection (Gronlund & Ratcliff, 1989; Hintzman & Curran, 1994; McElree, Dolan, & Jacoby, 1999; Ratcliff & McKoon, 1989). An fMRI study (Volz et al., 2006) suggested that judgments in disagreement with the recognition heuristic required more cognitive effort (indicated by a reduction in activation in the anterior frontomedian cortex) than did judgments in line with it. This study provides evidence for two processes: recognition and evaluation. The first identifies an alternative as recognized or not, and the second evaluates whether a default reliance on recognition should be suspended because recognition is not predictive of the criterion.

The literature offers different hypotheses about how this suspension process may operate (Pachur, Todd, Gigerenzer, Schooler, & Goldstein, in press). For instance, according to Pachur and Hertwig (2006), knowledge about alternatives' criterion values can lead people to overrule the default (but see Hilbig, Pohl, & Bröder, 2009). There is also evidence that people directly take into account the strength of the recognition correlation. When this correlation is substantial, people's inferences tend to accord with the recognition heuristic (e.g., Hertwig et al., 2008). In contrast, when it is less pronounced, they tend not to
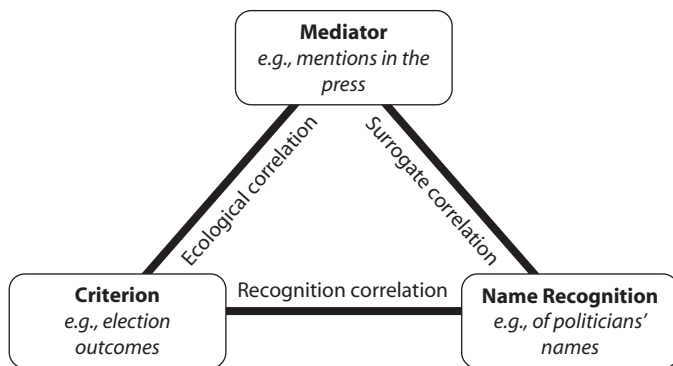

Figure 1. Ecological rationality of the recognition heuristic. An unknown criterion (e.g., the number of votes candidates win in an election) is reflected by a mediator (e.g., the press). The mediator makes it more likely for a person (e.g., a voter) to encounter alternatives with larger criterion values than those with smaller ones (e.g., the press mentions more successful candidates more frequently). As a result, the person will be more likely to recognize alternatives with larger criterion values than those with smaller ones, and, ultimately, recognition judgments can be relied upon to infer the criterion (e.g., the success of candidates in elections). The relations between the criterion, the mediator, and recognition can be measured in terms of correlations, or, as is explained in the text, in terms of validities (see Study 2).

do so (Pohl, 2006); and in fact, participants who *always* judge recognized alternatives to be larger than unrecognized ones—as opposed to people who *sometimes* infer unrecognized alternatives to be larger—estimate the recognition correlation to be larger (Pachur et al., 2008).

In our studies, we explore yet another hypothesis. As we explain in detail below, the recognition heuristic operates on a binary judgment of recognition. According to Schooler and Hertwig (2005), the strength of the underlying recognition signal, or *memory activation*, contributes to how quickly this binary judgment is carried out; activation influences the speed with which an alternative comes to mind—that is, its retrieval time, or *retrieval fluency*. Suspensions of the recognition heuristic could depend on the duration of the retrieval times, and, in fact, there is evidence that an alternative's retrieval time can reflect whether recognition is likely to help people make accurate inferences about that alternative (Hertwig et al., 2008; Marewski & Schooler, 2009).

Short Overview of the Studies

We ran our studies in a domain where multiple alternatives are common: political elections. In Study 1, we investigated how German voters use recognition when they forecast elections, establishing that the recognition heuristic is ecologically rational in this domain, and examining individual differences. In Studies 2 and 3, we focused on alternatives' retrieval fluency as a possible determinant of people's reliance on recognition, testing the recognition heuristic in elections as well as in the two-alternative task that Goldstein and Gigerenzer (1999) had originally used. In Studies 3 and 4, we returned to individual differences.

Importantly, in all studies we pitted the recognition heuristic against other models. The level of detail of these model comparisons varies across studies. For instance, Study 1 was not conducted with the samples of psychology students that so many psychological experiments make use of, but with eligible voters from various demographic backgrounds interviewed in the field. This limited how much data we could collect from each participant. Studies 3 and 4, which were run in our laboratories, allowed for more extensive data collection and, correspondingly, more extensive, and arguably stronger, model comparisons. In these studies, we tailored the models not only to individual participants' behavior, but also to their cue knowledge and retrieval fluency, assessing for each participant the optimal weights for this information and testing the models' ability to fit existing data as well as to generalize to new data.

STUDY 1
Recognition and Voters in Political Elections

Our first study took place prior to the 2004 parliamentary elections in the German federal state of Brandenburg. Here, voters had the opportunity to cast two ballots: one for a candidate who will represent the person's voting district, and the other for a party. Several weeks before the election, ads with candidates' and parties' names were placed, and they started to figure prominently in the news.

If the use of the recognition heuristic is ecologically rational for forecasts of these candidates' and parties' electoral success, we would expect the frequency of mentions of them in the press and election advertisements *before* the election to reflect the number of votes they win. Likewise, we would expect these frequencies of mentions to correlate with the number of voters who have heard of the parties and candidates *before* the election. Last, it should be possible to predict election outcomes on the basis of voters' recognition of parties and candidates alone (Figure 1). Below, we examine whether this pattern emerges and test how well the recognition heuristic predicts eligible voters' election forecasts, examining individual differences.

Method

To vote for a candidate from a given voting district, a citizen has to be a resident of that district. On two dates, 14 days and 1 day before the election, we invited residents of two voting districts in the downtown areas of the cities of Potsdam and Werder to fill out a questionnaire. Participants were recruited on the street and had to be at least 18 years old (voting age in Germany). They were paid €5 ($7). Of 246 recruited eligible voters (henceforth: *voters*), 172 completed the questionnaire (55% female; mean age = 38 years, $SD = 14.7$).

The questionnaire consisted of three sections. The first included a recognition task and a 2AFC task on the 11 candidates running in the two voting districts. The second section comprised a recognition task and a ranking task on the 15 parties running in the election. The third section included a cue knowledge task on the 11 candidates. In the recognition tasks, voters were given lists of candidates' and parties' names, respectively. The voters indicated whether they recognized each name—that is, whether they had heard or seen it before participating in the study. The two-alternative task consisted of 25 paired comparisons of candidates' names (complete pairings of the candidates from each voting district, respectively).[2] Voters indicated for each pair which candidate would win more votes in the election. In the ranking task, voters were given a list of the 15 party names and they assigned 1 of 15 ranks to each party (each rank could be assigned once), according to their forecasts of the number of votes the parties would win. In the cue knowledge task, voters indicated candidates' party affiliation, which is commonly a highly valid cue for electoral success (not unique to Germany—for the U.S., see Bartels, 2000). If an affiliation was unknown, they were instructed not to guess but to leave the answer space blank. Within the first two questionnaire sections, we counterbalanced the recognition and two-alternative tasks, as well as the recognition and ranking tasks, respectively.[3] The order of appearance of names was randomized in all tasks. Completing the tasks took about 10 min.

In addition, over a period of 54 days before the first data collection point, we counted the number of articles in which a candidate's name appeared in the *Potsdamer Neueste Nachrichten* (*PNN*) and the *Märkische Allgemeine Zeitung* (*MAZ*). These daily newspapers focus on Brandenburg and were the most frequently read in our sample. For parties, we made corresponding counts in the *MAZ*. Candidates also provided us with counts of the number of campaign materials (e.g., election posters, brochures, letters; henceforth: *flyers*) in which their names had appeared and that were distributed in the two voting districts prior to the election.

Results and Discussion

Ecological rationale for using recognition to forecast elections. Eighty-one voters (47%) reported subscribing to a newspaper, 65 of them to either the *MAZ* or the *PNN*. We found substantial correlations between the frequency with which parties and candidates were mentioned in the newspapers, the number of voters who recognized them, and the election results (Figure 2). In short, it is ecologically rational to use the recognition heuristic to forecast electoral success.
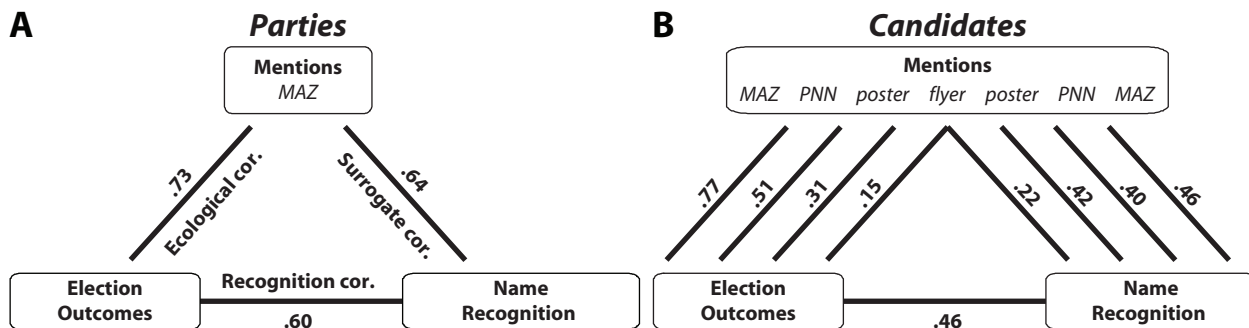
Figure 2. (A) Goodman and Kruskal's (1954) gamma computed between the number of mentions of 15 parties in the newspaper *Märkische Allgemeine Zeitung* (*MAZ*), the number of votes won by 15 parties in the election, and the number of participants who recognized the name of a party. (B) Gamma computed between the number of mentions of 11 candidates in the *MAZ* and in the *Potsdamer Neueste Nachrichten* (*PNN*), the number of election posters with a candidate's name, the number of flyers with a candidate's name, percentages of votes won by each candidate in his or her voting district, and the number of participants who recognized a candidate's name. Interestingly, newspaper mentions are the best predictors of election outcomes and name recognition (cor., correlation; *N* = 172; Study 1).

Measures. To assess how well the recognition heuristic predicts voters' forecasts in the two-alternative task, for each voter we selected the paired comparisons, in which one candidate was recognized but not the other. Across these comparisons, we counted how often (*A*) the voter had inferred that a recognized candidate would win more votes than an unrecognized one and the number of times (*D*) the opposite was inferred. The *accordance rate*, *k*, is the proportion of inferences consistent with the recognition heuristic:

$$k = A/(A + D). \tag{1}$$

To evaluate how well the recognition heuristic predicts voters' forecasts in the ranking task, we needed a measure that would reach its maximum value if, of 15 parties, the *n* recognized parties are assigned the *n* higher ranks and the (15 − *n*) unrecognized parties the (15 − *n*) lower ranks. The more often the recognized parties are ranked lower than the unrecognized ones, the more the behavioral data will deviate from the model's prediction, something we also needed our measure to take into account. As it turns out, *k* fulfills this requirement. For each voter, we used the ranking to simulate that voter's complete (i.e., exhaustive) set of *virtual* paired comparisons, consisting of one recognized and one unrecognized party, and computed the accordance rate using Equation 1.

Individual differences in two-alternative inference. For the 164 participants who had at least one pair consisting of a recognized and an unrecognized candidate, the mean recognition heuristic accordance was .80 (*SE* = .02) on these pairs (mean number of pairs = 9.90, *SE* = 0.26). Yet there may be individual differences in people's use of recognition, and some people may have integrated cues at odds with recognition rather than relying exclusively on the recognition heuristic. In German elections, knowledge about a candidate's party affiliation is such a cue, one that people could use. Party affiliations are commonly known to be highly predictive of electoral success, since candidates from smaller, less influential parties typically attract fewer votes than do those from the larger, dominating parties. In fact, according to all major opinion polls published prior to the Brandenburg election, the three largest parties

were expected to win the most votes; in short, a candidate's party was possibly the best cue voters could rely on. Did voters rely on it?

We had asked voters to indicate candidates' party affiliations, and used their responses to examine how well the recognition heuristic predicts their forecasts in two situations. First, in comparisons between a recognized and an unrecognized candidate, a voter might believe that the recognized candidate was from one of the three largest, usually most successful parties. Thus, recognition and the party cue would make the same prediction (*nonconflicting pairs*). Second, the voter might believe the recognized candidate to be from one of the smaller parties, which typically attract fewer votes in elections. In this situation, the party cue would suggest that the recognized candidate would win fewer votes, whereas recognition would suggest that this candidate would win more (*conflicting pairs*).

For 81 voters, we identified at least one conflicting (*M* = 3.15, *SE* = 0.23) *and* one nonconflicting pair (*M* = 5.60, *SE* = 0.31). The recognition heuristic predicted forecasts better on nonconflicting pairs (mean accordance computed across participants, M_k = .87, *SE* = .03) than it did on conflicting pairs (M_k = .73, *SE* = .04; 95% confidence interval [CI] on the mean difference [.07, .22]). However, with the caveat that the number of pairs per participant is very small, a substantial proportion of voters *always* behaved in accordance with the recognition heuristic on both types of pairs (Figure 3). In line with Pachur et al.'s (2008) reanalyses of Richter and Späth's (2006) and Newell and Fernandez's (2006) data, a strong contradictory cue seemed to have no impact on these voters' inferences, leaving open the possibility that these people relied on the recognition heuristic.

Model Comparison 1. To evaluate how well the recognition heuristic predicts behavior, we compared it with a compensatory model that integrates the party cue and recognition: According to *take-one-cue* (Table 1), if a voter believed that a recognized candidate was affiliated with one of the smaller parties that typically attract fewer votes in elections, he or she would infer that this candidate would win fewer votes than an unrecognized one would. Conversely, if a recognized candidate were from one of the
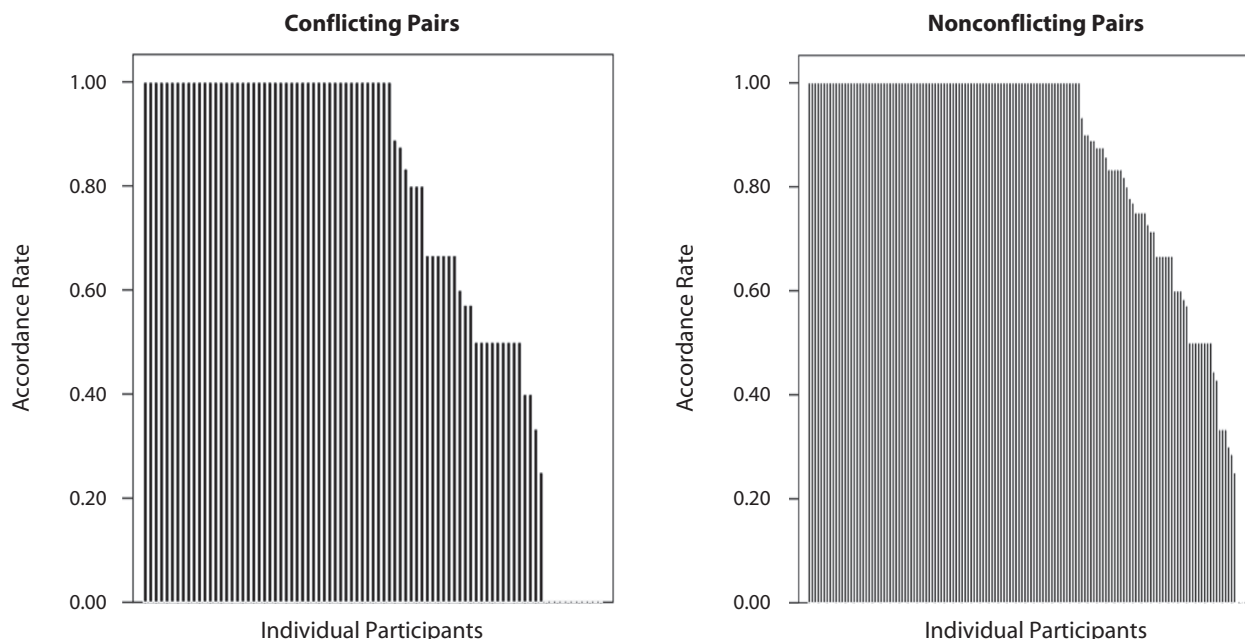
Conflicting Pairs

Nonconflicting Pairs

Figure 3. Recognition heuristic accordance rates for conflicting and nonconflicting pairs. Bars represent individual participants' accordance rates. On nonconflicting pairs, the accordance rate for 3 participants was 0; on conflicting pairs the accordance rate of 11 participants was 0. For these participants no bars are shown; dots on the figure border indicate their data ($n_{\text{nonconflicting}} = 146$; $n_{\text{conflicting}} = 85$; Study 1).

larger parties that tend to be more successful in elections, or if a recognized candidate's affiliation were unknown, the voter would infer that this candidate would win more votes than the unrecognized one would. The intuition is that if a candidate is from a sufficiently small and unsuccessful party, this then compensates for being recognized; that is, it detracts from being recognized.

Take-one-cue is more flexible than the recognition heuristic: It can also infer larger criterion values for unrecognized alternatives. The model pays for this flexibility with an increase in complexity. That is, it assumes a free parameter, the *cutoff criterion*, C_1, which measures the weight of one important cue—party affiliation—against the weight of recognition. For each participant, we classified the parties that he or she believed a candidate was affiliated with as "large, successful in elections" or as "small, unsuccessful," according to the number of votes the party had gained in the previous Brandenburg election in 1999. That is, to cover all possible classifications (i.e., all values of the parameter C_1), in a first round for $C_1 = 1$, we classified only the party that actually won the 1999 election as large and successful; all other parties were small and unsuccessful. In a second round, we then classified the two parties that won the most votes as large and successful ($C_1 = 2$), and so on, until finally we classified all parties as large and successful ($C_1 = 15$). For each participant, we computed the take-one-cue accordance across all pairs of a recognized and an unrecognized candidate for all values for C_1 using Equation 1 in the same way as we did for the recognition heuristic. Formally, the recognition heuristic is a special case of take-one-cue with C_1 set to be noncompensatory, such that recognized candidates are always forecast to win over unrecognized ones;

that is, the recognition heuristic's election forecasts only differ from take-one-cue's forecasts if take-one-cue's free parameter C_1 is set to be compensatory. Does this increase in model complexity pay off?

It does not. As Figure 4 shows, only when take-one-cue forecasts recognized candidates to win over unrecognized ones in 99% to 100% of the pairs (from $C_1 = 6$ [99%] to $C_1 = 15$ [100%]) does it fit voters' forecasts, on average, as well as the recognition heuristic does; the recognition heuristic, however, predicts voters' forecasts from scratch, without assuming an additional parameter. From $C_1 = 1$ to $C_1 = 5$, take-one-cue makes the same predictions as the recognition heuristic does in 47% to 91% of the pairs. Here, the recognition heuristic predicts voters' forecasts better. In short, a more complex compensatory model does not outperform the recognition heuristic.

Individual differences in multialternative inference. On average, there were 46.63 ($SE = 0.85$, $n = 163$) virtual paired comparisons of recognized and unrecognized parties; and indeed, voters tended to rank recognized parties higher than unrecognized ones ($M_k = .77$, $SE = .02$), consistent with the prediction that people identify consideration sets of recognized alternatives. We did not assess individual voters' knowledge about the parties, so we could not examine their use of recognition in multialternative inference in the same way as we did in forecasts concerning two candidates. However, there were two ways in which we could examine the *accuracy of their election forecasts*— that is, the degree to which their rankings reflected the rank order of the number of votes the parties won in the election. Exhaustively pairing all parties, we counted how often the parties a voter ranked higher had indeed won more votes in the election than the parties that the voter had ranked
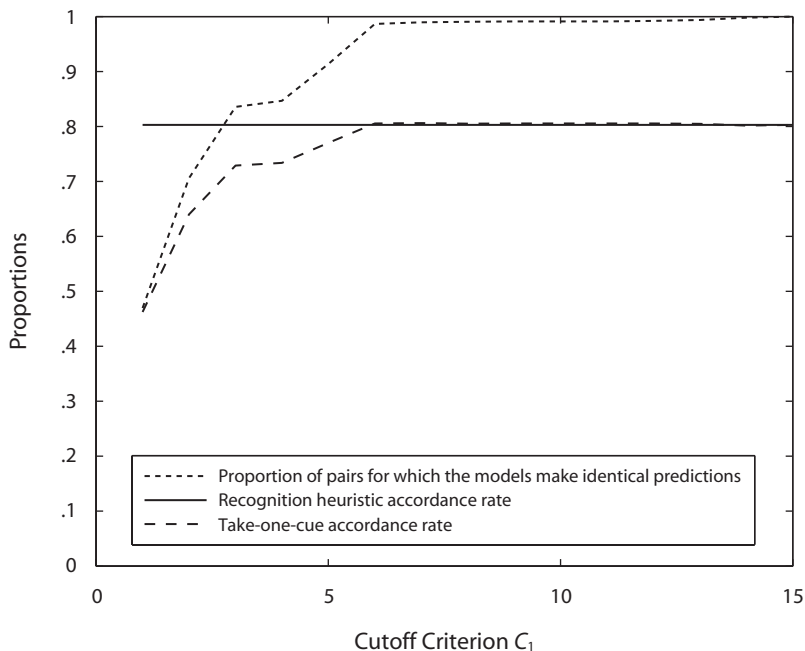
Figure 4. Recognition heuristic versus take-one-cue in a two-alternative forced choice task, in which people inferred which of two candidates would win more votes in the 2004 Brandenburg election. Lines show mean accordance rates of the two models computed for each of the possible values of the cutoff criterion C_1, as well as the mean proportion of paired comparisons in which the two models made the same predictions. The cutoff criterion C_1 was the parties' success in the previous Brandenburg election in 1999. If a party did not run in 1999, this number was set to zero. Note that the same pattern of results emerged when we used other measures for C_1, such as opinion polls conducted prior to the 2004 Brandenburg election, or the outcomes of this 2004 election ($n = 164$; Study 1).

lower. The accuracy is the proportion of correct rankings, computed across these virtual comparisons of parties.

First, does deviating from the recognition heuristic improve the accuracy of a voter's forecasts more than consistently following it does? We compared voters who had a subscription to a newspaper with those who had none. Virtually all newspapers reported on the elections, and many mentioned cues useful when forecasting elections, such as the results of opinion polls. Were newspaper subscribers more successful than nonsubscribers when *not* adhering to the recognition heuristic? For nonsubscribers, there were 48.25 ($SE = 1.03$, $n = 88$) virtual pairs of recognized and unrecognized parties on average; for subscribers there were 44.73 ($SE = 1.39$, $n = 74$). Subscribers' accordance rates ($M_k = .79$, $SE = .02$) were similar to those of non-subscribers ($M_k = .75$, $SE = .03$). The accuracy of the election forecasts was .79 ($SE = .01$) for subscribers on average. If they had assigned all ranks as prescribed by the recognition heuristic, their rankings would have been equally accurate (achievable proportion correct in this case, $M = .80$, $SE = .01$; 95% CI on the mean difference between achievable and effectively achieved proportion correct [−.01, .03]). For nonsubscribers, however, the proportion of correct forecasts was low ($M = .69$, $SE = .02$). Had they always relied on the recognition heuristic, the mean achievable proportion of correct forecasts would have been as large as .75 ($SE = .01$; 95% CI on the mean

difference between achievable and achieved proportion correct [.04, .09]). That is, even the more knowledgeable newspaper-subscribing voters were unable to outwit the recognition heuristic. At the same time, presumably less knowledgeable voters who did not subscribe to a paper would have done better by blindly using it.

Second, consistent with two-stage theories of consumer choice, the recognition heuristic allows a person to split all alternatives into two sets: the consideration set of recognized alternatives and a set of unrecognized alternatives. Within the set of unrecognized alternatives, all alternatives are treated equally, to be ranked by mere guessing. The recognized alternatives in the consideration set, in contrast, can be ranked on the basis of more detailed examinations—for instance, by using alternatives' attributes as cues. In fact, within the set of unrecognized parties (consisting of 13.20 virtual paired comparisons on average; $SE = .78$, $n = 157$), the proportion of correct election forecasts was at chance level ($M = .51$, $SE = .02$), consistent with the prediction that people ranked these alternatives on the basis of mere guessing. At the same time, voters' forecasts were accurate within the consideration set of recognized parties (proportion of correct rankings: $M = .76$, $SE = .01$; mean number of virtual comparisons = 48.76, $SE = 1.74$, $n = 172$), consistent with the hypothesis that people ranked these alternatives on the basis of more detailed examinations.

To summarize, when deciding which of two candidates will win, many voters made forecasts consistent with the recognition heuristic, even when a strong conflicting cue was available; and in fact, a more complex compensatory model did not fit voters' forecasts better than the recognition heuristic. When voters forecasted the election for 15 parties, in line with the recognition heuristic, they ranked recognized parties higher than unrecognized ones, appearing to form consideration sets of recognized parties for which their forecasts were accurate. More knowledgeable voters would have done as well by always adhering to the recognition heuristic than by going against it as much as they did; at the same time, less knowledgeable ones even impaired their accuracy by not always adhering to the heuristic.

STUDY 2
When Is It Ecologically Rational to Rely on Recognition?

In a study series, Pohl (2006) asked people to categorize alternatives into those they recognized without knowing anything else about them (R^-) and those they recognized and had knowledge about (R^+). The recognition heuristic predicted people's inferences better on R^+U pairs (i.e., where one alternative was recognized and there was knowledge about it and the other was unrecognized [U]) than on R^-U pairs (i.e., where one alternative was recognized but there was no knowledge about it and the other was unrecognized). Does this finding indicate that people integrate cues into their inferences, rather than rely on the recognition heuristic?

The recognition heuristic operates on judgments of recognition (Gigerenzer, Hoffrage, & Goldstein, 2008; Goldstein & Gigerenzer, 2002). By implementing this heuristic in the *ACT–R cognitive architecture* (e.g., Anderson et al., 2004), Schooler and Hertwig (2005) provided a formal model of these recognition judgments (see also Pachur, in press; Pleskac, 2007). As we will explain in more detail in Study 3, the judgments depend on the strength of an alternative's recognition signal—that is, on its activation in memory. Activation is a function of the frequency and recency of a person's past encounters with an alternative and determines the alternative's retrieval fluency and recognition speed. The more strongly an alternative is activated, the more quickly it will be retrieved and recognized. In a series of studies and computer simulations with the ACT–R memory model, Marewski and Schooler (2009) showed that alternatives about which people are likely to recall knowledge (R^+) tend to be more strongly activated than alternatives about which knowledge is unlikely to be available (R^-). Consequently, people are slower to recognize and retrieve R^- alternatives. Since the recognized alternative's retrieval fluency tends thus to be larger in R^+U pairs than in R^-U pairs, it may often be harder to rely on recognition on R^-U than on R^+U pairs, resulting in lower recognition heuristic accordance rates (see Equation 1) on R^-U pairs.

In fact, Marewski and Schooler (2009) demonstrated that the way in which memory works can make it easier for a person to use a given heuristic when using it is also likely to result in accurate judgments. This may hold true for the recognition heuristic: When the recognition correlation (see Figure 2) is substantial, the probability of retrieving knowledge about an alternative correlates with the criterion. As a result, alternatives with knowledge (R^+) score on average higher on the criterion than alternatives without knowledge (R^-). Since both tend to have larger criterion values than do unrecognized alternatives on average, R^+U pairs tend to reflect larger differences on the criterion than do R^-U pairs. This, in turn, may result in a stronger recognition correlation on R^+U than on R^-U pairs such that it may actually be ecologically rational to use the recognition heuristic more on R^+U than on R^-U pairs.

In short, one can formulate two competing hypotheses: (1) If Pohl's (2006) finding implies that people use compensatory strategies rather than the recognition heuristic, such models should predict inferences better than the recognition heuristic; alternatively, (2) if no model predicts people's inferences better than the recognition heuristic does, Pohl's finding would leave open the possibility that people rely on the recognition heuristic less often on R^-U than on R^+U pairs, because recognition is harder to assess and less predictive on R^-U than on R^+U pairs. We will continue to provide model comparisons below; next, however, we test whether the recognition correlation is smaller on R^-U than on R^+U pairs.

The setting of our study is the 2005 parliamentary election in North Rhine-Westphalia, the most populous German state. In contrast to the previously discussed election, here voters did not cast a ballot directly for a candidate, but chose among 24 parties.

Method

Sixty-one participants (44% female; mean age = 26 years, SD = 3.6) filled out a questionnaire 3 to 11 days before the North Rhine-Westphalia election. About half of them completed the questionnaire in the labs of the Max Planck Institute for Human Development in Berlin and received €5 ($7) for their participation; the other half worked on it in a class at the Free University in Berlin. All participants had to be at least 18 years of age.

The questionnaire consisted of a ranking and a recognition task for the 24 parties. These tasks were identical to the ones used in Study 1. In a detailed recognition task, we additionally presented participants a list of all parties and asked them how much they knew about each party. There were three possible answers: (1) they had *never* heard of the party and had *never* seen it before participating in the study (U); (2) they had heard of the party or had seen it before but *did not know anything else* about it beyond recognizing its name (R^-); (3) they had heard of it or had seen it before and knew something about the party beyond simply recognizing it (R^+). We counterbalanced the ranking and recognition tasks; the detailed recognition task was at the end.[4] The order of appearance of parties was randomized. Filling out the questionnaire took about 10 min.

Results and Discussion

Strength of the recognition correlation. The recognition correlation can be expressed in terms of the *recognition validity*, α—that is, the probability of a recognized alternative scoring higher on the criterion than an unrecognized one (Goldstein & Gigerenzer, 2002). Across virtual paired comparisons of recognized and unrecognized parties, we counted for each participant the number of times T a recognized party had gained more votes than an

unrecognized one and the number of instances W in which the reverse had happened:

$$\alpha = T/(T + W). \qquad (2)$$

As expected, the recognition validity was larger on R^+U pairs ($M_\alpha = .92$, $SE = .01$) than on R^-U pairs ($M_\alpha = .68$, $SE = .03$; 95% CI on the mean difference [.29, .19]) ($n = 54$; mean number of R^+U pairs = 100.19, $SE = 3.25$; mean number of R^-U pairs = 35.92, $SE = 2.87$). At the same time, the recognition heuristic accordance rate (computed across virtual comparisons between parties) was larger on R^+U pairs ($M_k = .89$, $SE = .01$) than on R^-U pairs ($M_k = .62$, $SE = .03$; 95% CI on the mean difference [.34, .21]) ($n = 54$). In short, people acted more strongly in accordance with the recognition heuristic when recognition was more predictive.

Model Comparison 2. Overall, people were unlikely to rank recognized parties lower than unrecognized ones: When computed across all virtual pairs of recognized and unrecognized parties (mean number of pairs = 136.02, $SE = 1.54$, $n = 59$), the average accordance rate was .82 ($SE = .01$), consistent with the hypothesis that people identify consideration sets of recognized alternatives. To evaluate how well the recognition heuristic predicts behavior, we compared it with a generalization of take-one-cue to multiple alternatives (Table 1). Assuming that people had some knowledge about the characteristics of the parties we classified as R^+, we implemented take-one-cue as follows: If an R^+ party belonged to the smaller parties that typically attract fewer votes in elections, a participant would rank this party lower than all unrecognized ones; if the R^+ party represented a larger, commonly more successful party, the participant would rank it higher than all unrecognized ones. The participant would rank recognized parties without knowledge (R^-) higher than unrecognized ones. As in Study 1, the intuition is that if a party is sufficiently small and unsuccessful, this then compensates for being recognized; that is, it detracts from being recognized.

Also in the multialternative case, take-one-cue assumes a free parameter, the cutoff criterion C_2, which measures the weight of the party cue against recognition, determining which parties are small, unsuccessful ones. For each participant, we classified all R^+ parties as "small, unsuccessful" or "large, successful," according to the results of the previous North Rhine-Westphalia election in 2000. To cover all possible classifications (i.e., all values of C_2), in a first round for $C_2 = 1$, we classified only the party that won the election as large and successful; all others were small and unsuccessful. In a second round, we classified the two parties that won the most votes as large and successful ($C_2 = 2$), and so on, until finally we classified all parties as large and successful ($C_2 = 24$). For each participant, we computed the take-one-cue accordance (Equation 1) across virtual pairs of recognized and unrecognized parties (mean number of pairs = 136.02, $SE = 1.54$) for all values of C_2. Does this increase in model complexity pay off?

It does not. Only when take-one-cue ranked recognized parties higher than unrecognized ones in 100% of the virtual party pairs (for $C_2 = 24$; Figure 5) did it fit people's forecasts, on average, as well as the recognition heuristic

does. The recognition heuristic, however, predicts people's forecasts without assuming an additional parameter. From $C_2 = 1$ to $C_2 = 24$, take-one-cue makes the same predictions as the recognition heuristic in 36% to 99% of the pairs. Here, the recognition heuristic predicts voters' forecasts better.

In short, consistent with the assumption that people rely on the recognition heuristic to identify consideration sets of recognized alternatives, participants ranked recognized parties higher than unrecognized ones. When recognition was more predictive, people acted more strongly in accordance with the recognition heuristic. A more complex compensatory model did not fit people's rankings better than the recognition heuristic.

STUDY 3
A Competition of Strategies for Two-Alternative Inference: Modeling Individual Differences

There might be individual differences in the weighting of cues, which would have been ignored in the previous model comparisons. In Studies 3 and 4, we take such individual differences into account and test the recognition heuristic against five compensatory strategies. These strategies include two variants of take-one-cue as well as models that integrate several cues to make decisions. For instance, *tallying-of-negative-cues* and *tallying-of-positive-and-negative-cues* compute sums of various positive and/or negative cues, weighting them against recognition (Table 1). Except for the weighting, they are equivalent to tallying (Gigerenzer & Goldstein, 1996) and unit-weight linear models (e.g., Dawes, 1979; Einhorn & Hogarth, 1975).

What could be a compensatory alternative to the recognition heuristic when cues are not available (i.e., on R^-U pairs)? Above, we argued that the speed with which the recognized alternative is retrieved—that is, its retrieval fluency—tends to be lower in R^-U than in R^+U pairs, making it harder to use the recognition heuristic on R^-U than on R^+U pairs, which may result in occasional suspensions of this heuristic on R^-U pairs, and as a result, lower recognition heuristic accordance rates. Yet an alternative hypothesis is that the recognition heuristic is an implausible model, and as such *never* used. Rather, people always systematically integrate retrieval fluency into their inferences (see Dougherty et al., 2008; Newell & Fernandez, 2006)—for instance, by weighting and adding it. Below, we pit the recognition heuristic for the first time against a corresponding model: *weighted-fluency* (Table 1). The intuition is that if an alternative's fluency falls below a threshold, this compensates for being recognized, just as in take-one-cue.

Fluency-based heuristics have been defined in many different ways (e.g., Jacoby & Brooks, 1984; Whittlesea, 1993). Weighted-fluency is a model that builds on Schooler and Hertwig's (2005) *fluency heuristic*, which in turn is grounded in these earlier definitions and in a long research tradition on fluency (e.g., Jacoby & Dallas, 1981), as well as on related notions such as accessibility (e.g., Bruner, 1957) or familiarity (e.g., Hintzman, 1988; Mandler, 1980). Schooler and Hertwig implemented their fluency heuristic side by side with the recognition heuristic in the ACT–R cognitive architecture. In ACT–R, the same
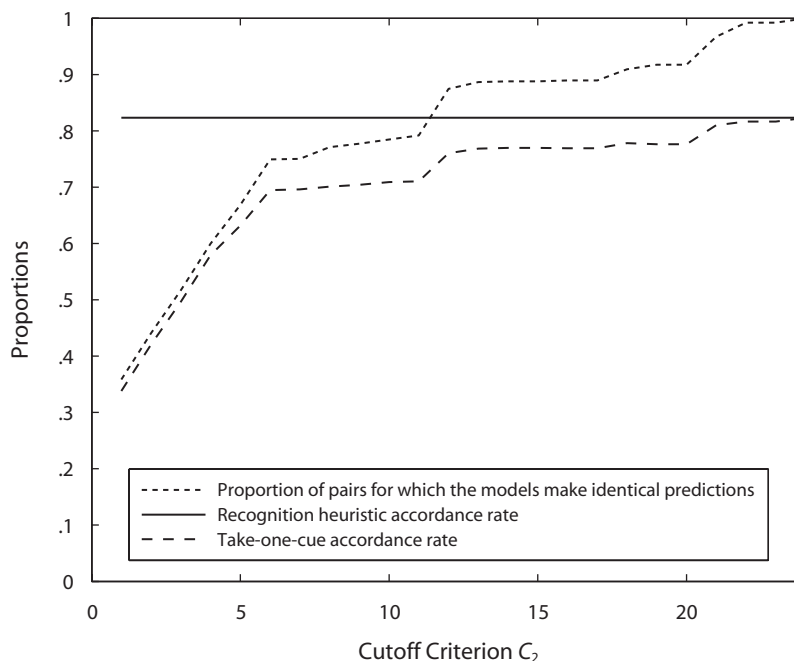
Figure 5. Recognition heuristic versus take-one-cue in a ranking task, in which people ranked parties according to their inferences about how many votes they would win in the 2005 North Rhine-Westphalia election. Lines show mean accordance rates for the two models computed for each of the possible values of the cutoff criterion C_2 as well as the mean proportion of virtual paired comparisons in which the two models made the same predictions. The cutoff criterion C_2 was the parties' success in the previous election in North Rhine-Westphalia in 2000. If a party had not run in 2000, this number was set to zero. Note that as in Study 1, the same pattern of results emerged when we used other measures for C_2, such as opinion polls conducted prior to the 2005 North Rhine-Westphalia election, or the outcomes of this 2005 election ($n = 59$; Study 2).

memory currency—a continuous activation trace—determines (1) whether an alternative will be retrieved or not, and (2) the time it takes to retrieve it. Schooler and Hertwig adopted Anderson, Bothell, Lebiere, and Matessa's (1998) assumption that an alternative's retrieval implies recognizing it, adding the assumption that the more quickly the alternative is retrieved, the greater the sense of recognition. A person using the recognition heuristic can base inferences on the binary outcome of this memory process (retrieved or not). A person using the fluency heuristic and/or weighted-fluency, in turn, can base inferences on the more graded outcome of the same process: the speed with which the alternatives come to mind—that is, their retrieval time or retrieval fluency. By this token, both the fluency heuristic and weighted-fluency are computational instantiations of the version of Tversky and Kahneman's (1973) *availability heuristic* that bases judgments on ease of retrieval.

To pit weighted-fluency, the two tallying models, and take-one-cue against the recognition heuristic, we reanalyze data from Marewski and Schooler (2009). This data allows us to examine inferences of city size—the task first used to test the recognition heuristic (Goldstein & Gigerenzer, 1999) and which most subsequent studies have used.

Method

Forty-nine right-handed participants (43% female; mean age = 24 years, $SD = 3.1$) completed a computerized experiment. They received a guaranteed payment of €13 ($19) supplemented by a performance bonus. Stimuli were 240 cities (i.e., of the 70 largest Austrian, British, French, German, Italian, Spanish, and U.S. cities, essentially those consisting of five to eight letters, excluding the countries' capital cities; participants were told that they would be shown real cities, but participants were not told that these cities were drawn from among the 70 largest of each country).

First, in a 2AFC task, we presented two cities on a computer screen (one on the left and the other on the right) and asked participants to infer which city had more inhabitants. Pairs of cities were randomly drawn for each country without replacement, such that each city could appear only once throughout the task (yielding 120 pairs of cities). On top of the guaranteed payment of €13, participants received €0.04 ($0.06) for each correct inference. Four cents was subtracted from this additional gain for each incorrect inference. (No feedback on the correctness of the responses was given until after the experiment.)

Next, in a recognition task, we gave participants the name of one city at a time and asked them to judge whether they had seen or heard of the city prior to participating in the study. Thereafter, in a detailed recognition task, we again presented one city at a time and asked them how much they knew about each city. As in Study 2, there were three possible answers: (1) never heard of it and never saw it before participating in the study (U); (2) heard of it or saw it before but do not know anything else about it beyond recognizing its name (R^-); (3) heard of it or saw it before and know something about the city beyond recognizing it (R^+).

Last, in a cue knowledge task, we asked participants to indicate for each city whether it (1) had an international airport, (2) had a university, (3) was a significant industry site, and (4) was a world-famous tourist site. Responses could be "yes," "no," or "I do not know," yielding positive, negative, and unknown cue values, respec-

tively. (Previous participants from the same subject pool of our lab considered these cues as the most useful ones for inferring city size; Pachur et al., 2008.) For each correct response, participants received €0.04 on top of the guaranteed payment. For incorrect responses, €0.04 was subtracted from this additional gain. (No feedback on the correctness of the responses was given until after the experiment.) Participants did not receive payment nor did they lose money for "don't know" responses.

Participants took a 20-sec break every other 12 trials (two-alternative choice and recognition tasks) and every other 12 question blocks (cue knowledge task), respectively. In all tasks, each trial was preceded by a fixation cross for 1,000 msec, and participants were instructed to always fixate this cross when it appeared. In all tasks, participants were instructed to respond as quickly and accurately as possible. Positive responses were always made with the index finger of the right hand. In all tasks, the order of presentation of cities was always completely randomized. Completing the experiment took about 70 min.[5]

Results and Discussion

Model Comparisons 3–6. Each person's individual cutoff criteria—C_1, C_3, C_4, and C_6—are free parameters that determine when each of four compensatory models infers unrecognized cities to be larger than recognized ones (Table 1). As in Studies 1 and 2, take-one-cue infers unrecognized cities to be larger than recognized ones when knowledge about the recognized city can be recalled (R^+) and that city's size falls below C_1. For instance, a person may know that a recognized city is small. In a comparison of that city and an unrecognized one, he may therefore decide for the unrecognized city. For take-one-cue we thus made the assumption that people have some knowledge about recognized (but not unrecognized) alternatives' criterion values.

The remaining three compensatory models dispense with this assumption. Coding participants' cue-knowledge as +1, −1, and 0 (positive, negative, and unknown cue value, respectively), tallying-of-negative-cues decides against recognized cities when the sum of negative and unknown cue values for the recognized cities falls below C_3. (Depending on how many cues have negative values, C_3 can thus take values ranging from −4 to +1.) In tallying-of-positive-and-negative-cues this happens when the sum of negative, positive, and unknown cues is smaller than C_4. (C_4 could thus take values ranging from −4 to +5.) Weighted-fluency decides against recognized cities when the retrieval time for the recognized city falls above C_6—that is, when it takes too much time to judge a city as recognized.

Table 1 shows when the four compensatory models decide in favor of recognized cities. Take-one-cue picks recognized cities when the recognized city's size falls above, or is equal to, C_1. In tallying-of-negative-cues, this is the case when the sum of negative and unknown cue values exceeds or is equal to C_3. In tallying-of-positive-and-negative-cues, this happens when the sum of positive, negative, and unknown cues falls above, or is equal to, C_4. Weighted-fluency decides in favor of recognized cities when the retrieval time for the recognized city falls below or equals C_6—that is, when it takes little time to judge a city as recognized.

When comparing nested models that differ in the number of free parameters, as is the case for the recognition heuristic and these four competitors, a good model evaluation criterion is the models' generalizability to new data (see also Pitt, Myung, & Zhang, 2002). We evaluated the models' ability to generalize to new data in a cross-validation.

To pit the three knowledge-based models (i.e., tallying-of-positive-and-negative-cues, tallying-of-negative-cues, take-one-cue) against the recognition heuristic, we selected for each participant those pairs of cities in which, according to the responses in the detailed recognition task, the participant had knowledge about the recognized city (R^+U pairs; mean number of R^+U pairs = 20.41, SE = 1.11). For the R^+ cities, we used participants' responses in the cue knowledge task to compute the sums of cue values used by the two tallying models to make predictions. To derive the predictions made by take-one-cue, we looked up the size of the R^+ cities. To assess which strategy describes people's inferences best when no knowledge is available, we tested weighted-fluency against the recognition heuristic on the R^-U pairs (mean number of R^-U pairs = 17.88, SE = 1.16).[6] In doing so, we operationalized retrieval time as recognition time, using each voter's response times (RTs) in the recognition task to generate the predictions made by weighted-fluency. (This procedure has been successfully applied to model people's inferences with the fluency heuristic; see Hertwig et al., 2008.)

We divided all R^+U pairs and all R^-U pairs, respectively, 10,000 times randomly into two halves. The first half represented the calibration set in which we calculated for each of 49 participants that person's *optimal* values for C_1, C_3, C_4, and C_6 for the four compensatory models. Optimal values are those at which a model's accordance rate is greatest; we derived them by exhaustively searching each participant's entire parameter space of values observed for C_1, C_3, C_4, and C_6 in that participant's calibration set. (If multiple values were optimal, we picked one at random.) In doing so, we estimated a total of 196 free parameter values (49 participants × 4 models). We used these values to compute the proportion of inferences consistent with each model in the other half, the validation set, where the models' generalizability is evaluated. For each partition, we also computed the recognition heuristic accordance rates. For almost all participants, the recognition heuristic predicted inferences better than each of the four alternative models (Figures 6–9).

We additionally assessed the models' ability to fit existing data. To this end, we recalculated the optimal values for C_1, C_3, C_4, and C_6 for each participant's *complete* set of R^+U pairs and R^-U pairs, respectively, and computed the associated accordance rates on the complete set of pairs. In doing so, we exhaustively searched each participant's entire parameter space of values observed for C_1, C_3, C_4, and C_6 in that participant's complete set of pairs, and reestimated the (49 participants × 4 models =) 196 free parameter values. The recognition heuristic and its competitors are nested models; however, in contrast to the latter, the recognition heuristic has no free parameter. This is why when the values for C_1, C_3, C_4, and C_6 are optimal, the recognition heuristic's accordance rate can *never* exceed its competitors' accordance rates, but its competitors' rates can exceed the recognition heuristic's accordance rate. Did they?

Corroborating our previous results, for most participants recognition had a noncompensatory weight; that is,
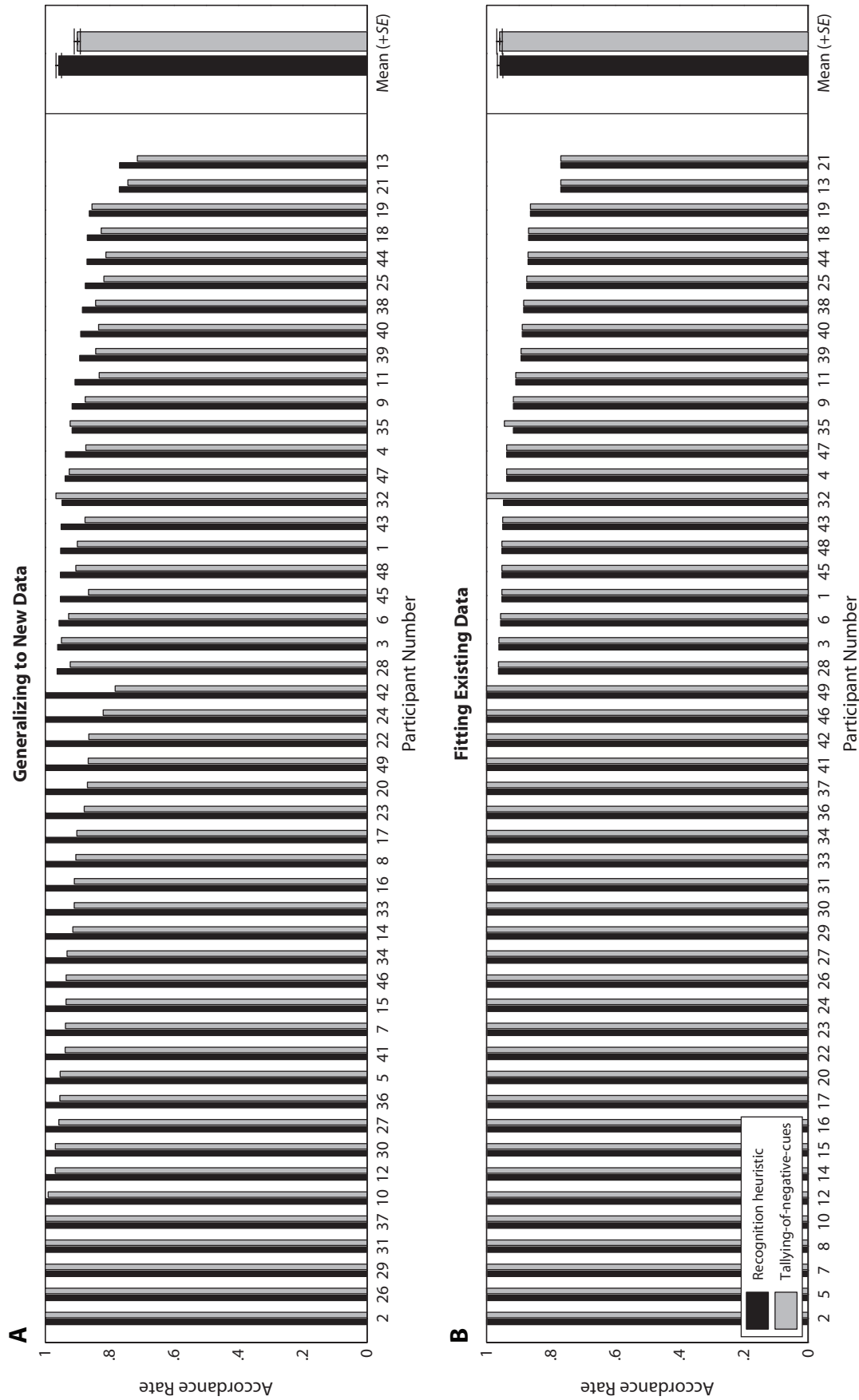
Figure 6. Recognition heuristic versus tallying-of-negative-cues in a two-alternative forced choice task. (*A*) Generalizing to new data: Bars show means of participants' accordance rates computed in the validation set across 10,000 random partitions of each participant's data. For Participants 32 and 35, the competitor predicted decisions better than the recognition heuristic; for all others the recognition heuristic predicted decisions best. For 44 participants, recognition had a noncompensatory weight in the competitor of the recognition heuristic, resulting in the competitor always making the same decisions as the recognition heuristic does. For 2 participants (32, 35), compensatory weights yielded the same fit as the recognition heuristic's competitor. For 3 other participants (1, 11, 25), compensatory weights resulted in the better fit for the recognition heuristic's competitor. (*B*) Fitting existing data: Bars show participants' accordance rates for the two models. In both panels, the two rightmost bars show mean accordance rates, computed across participants, and *SE* (*N* = 49; Study 3).
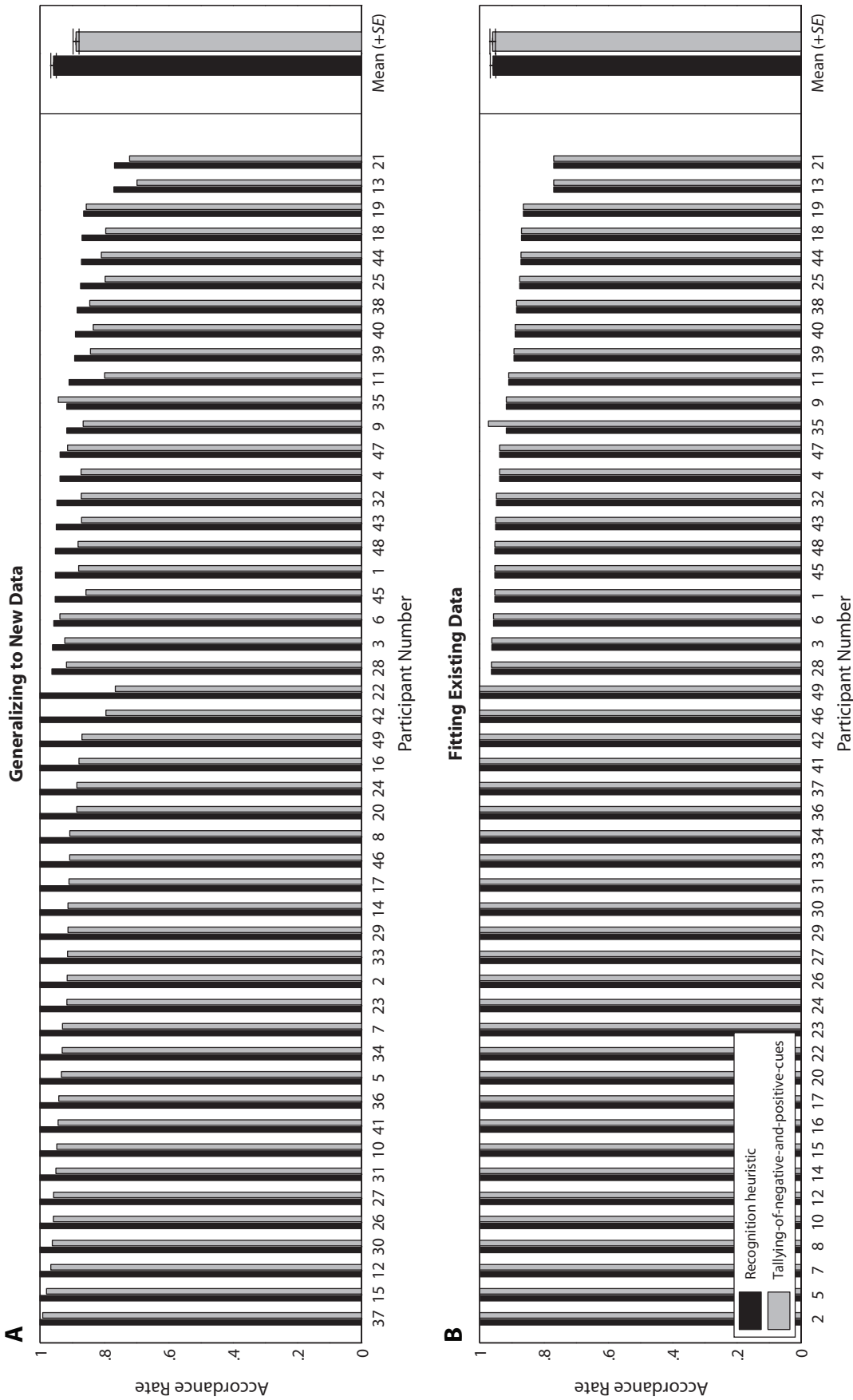
Figure 7. Recognition heuristic versus tallying-of-negative-and-positive-cues in a two-alternative forced choice task. (A) Generalizing to new data: Bars show means of participants' accordance rates computed in the validation set across 10,000 random partitions of each participant's data. For Participant 35, the competitor predicted decisions better than the recognition heuristic; for all others the recognition heuristic predicted decisions best. (B) Fitting existing data: Bars show participants' accordance rates for the two models. For 44 participants, recognition had a noncompensatory weight in the competitor of the recognition heuristic, resulting in the competitor always making the same decisions as the recognition heuristic does. For Participant 35, compensatory weights resulted in the better fit for the recognition heuristic's competitor. For 4 other participants (1, 11, 25, 32), compensatory weights yielded the same fit as noncompensatory weights. In both panels, the two rightmost bars show mean accordance rates, computed across participants, and *SE* (*N* = 49; Study 3).
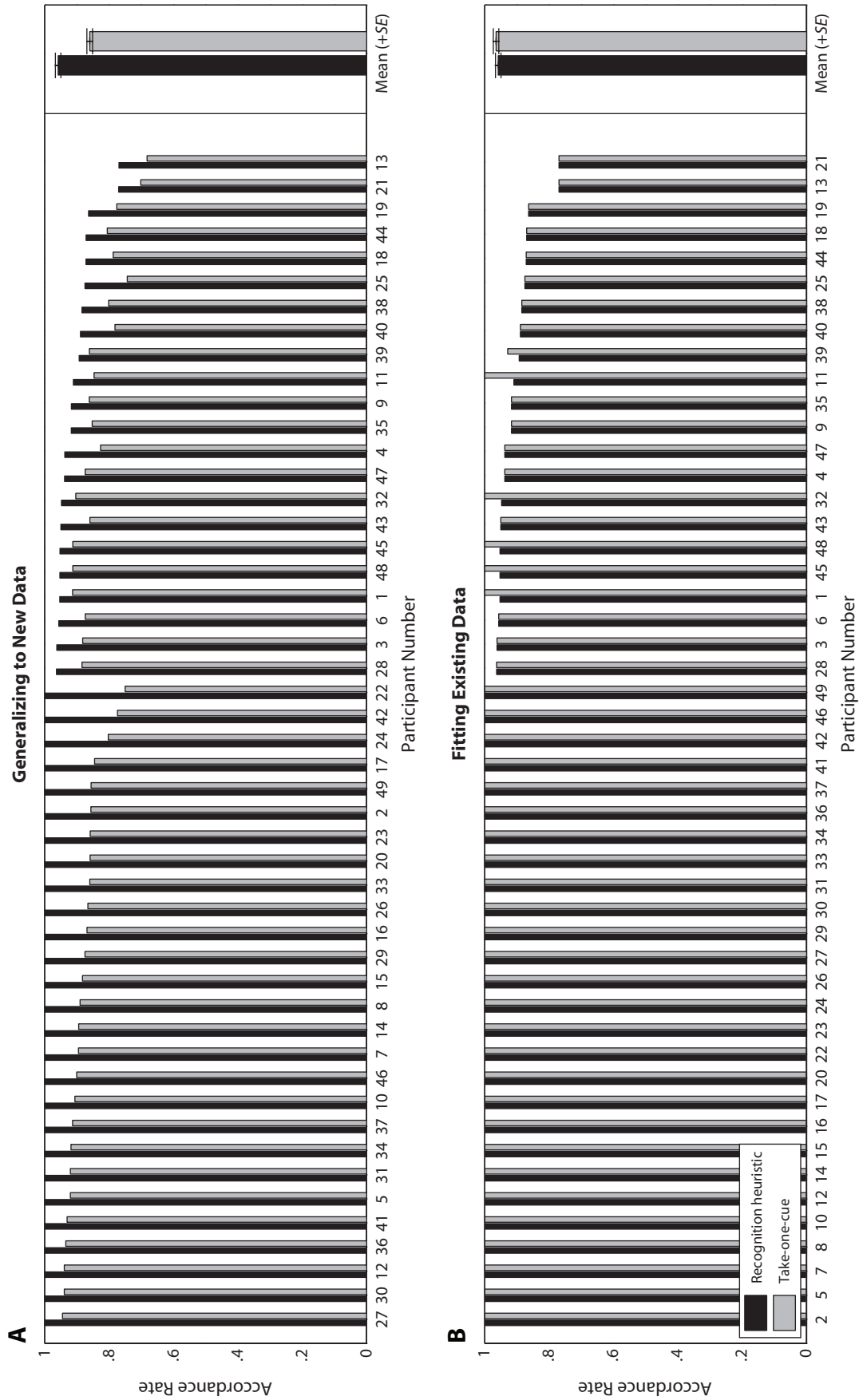
Figure 8. Recognition heuristic versus take-one-cue in a two-alternative forced choice task. (A) Generalizing to new data: Bars show means of participants' accordance rates computed in the validation set across 10,000 random partitions of each participant's data. For 0 participants, take-one-cue predicted decisions best; that is, the recognition heuristic predicted best for all. (B) Fitting existing data: Bars show participants' accordance rates. For 40 participants, recognition had a noncompensatory weight, resulting in take-one-cue always making the same inferences as the recognition heuristic does. For 6 participants (1, 11, 32, 39, 45, 48), compensatory weights resulted in a better fit for take-one-cue. For 3 participants (10, 14, 22), compensatory weights yielded the same fit as the recognition heuristic ($N = 49$; Study 3).
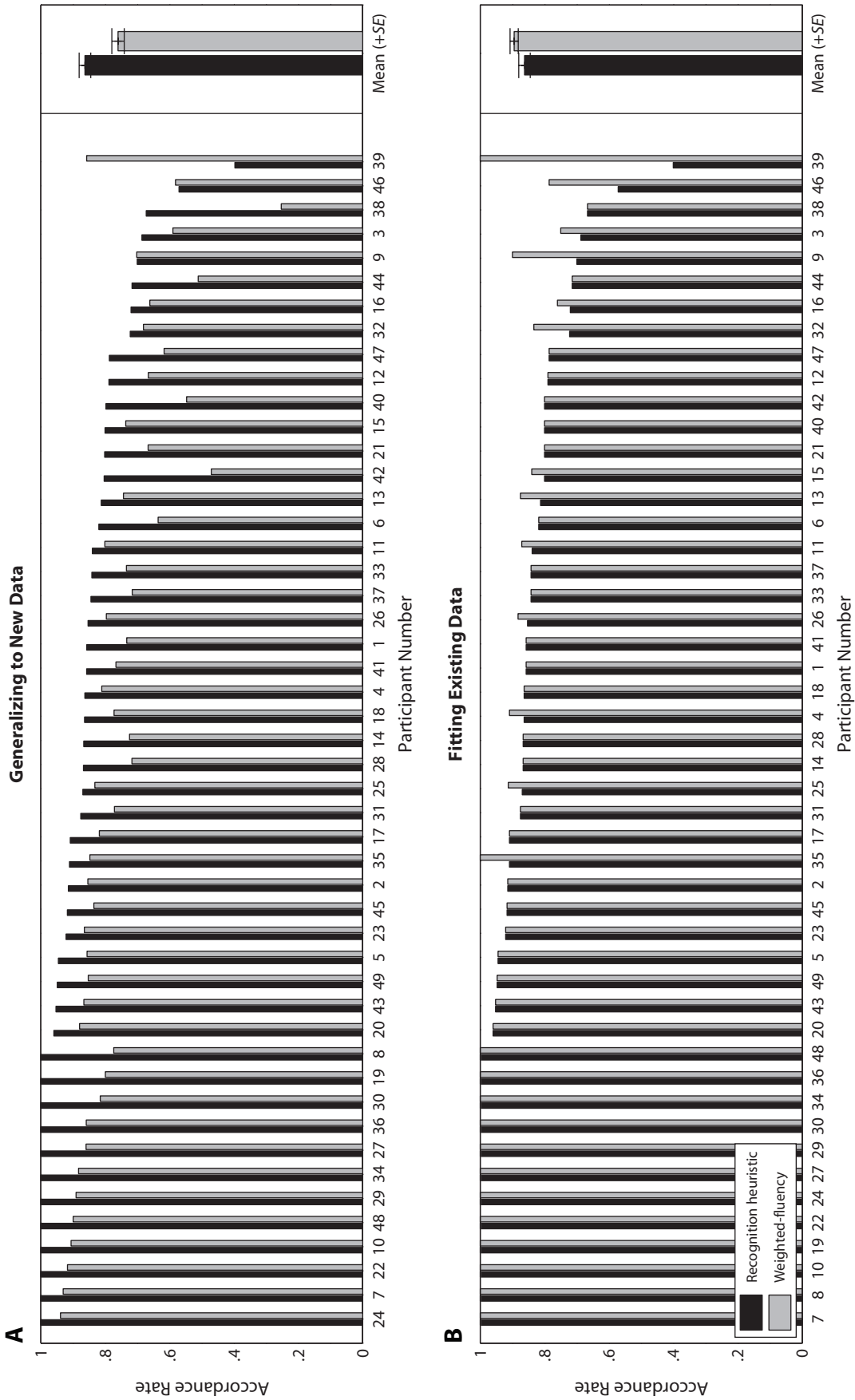
Figure 9. Recognition heuristic versus weighted-fluency in a two-alternative forced choice task. (A) Generalizing to new data: Bars show means of participants' accordance rates computed in the validation set across 10,000 random partitions of each participant's data. For 3 participants (9, 39, 46), weighted-fluency predicted decisions best; for all others the recognition heuristic was best. (B) Fitting existing data: Bars show participants' accordance rates. For 31 participants, recognition had a noncompensatory weight, resulting in weighted-fluency always making the same inferences as the recognition heuristic does. For 13 participants, compensatory weights resulted in a better fit for weighted-fluency. For 5 participants (5, 21, 40, 44, 45), compensatory weights yielded the same fit as noncompensatory ones ($N = 49$; Study 3).

the optimal values for C_1, C_3, C_4, and C_6 resulted in the respective compensatory model inferring recognized cities to be larger than unrecognized ones, just as the recognition heuristic does. Therefore, the accordance rates for the recognition heuristic and its competitors are identical for most participants (Figures 6–9). To compare, the figures additionally show mean accordance rates computed across participants, for both fitting existing and generalizing to new data. As can be seen, also on this aggregate level the 49×4 parameters estimated for the four compensatory models do not yield a substantially improved fit.

Retrieval fluency on R^+U and R^-U pairs. As in Study 2, the recognition validity was larger on R^+U pairs ($M_\alpha = .82$, $SE = .01$) than on R^-U pairs ($M_\alpha = .74$, $SE = .02$), whereas the recognition heuristic accordance rate was also larger on R^+U pairs ($M_k = .96$, $SE = .01$) than on R^-U pairs ($M_k = .86$, $SE = .02$).[7] Consistent with the hypothesis that it is more difficult to apply the recognition heuristic when a person using it is less likely to make accurate inferences with it (see Study 2), inferences on R^-U pairs took on average 208 msec longer (mean of median RT, $M_{Mdn} = 1,872$ msec, $SE = 79$) than did inferences on R^+U pairs ($M_{Mdn} = 1,664$, $SE = 69$; 95% CI on the mean difference [126, 290]) ($N = 49$). Similarly, as reported by Marewski and Schooler (2009), in the recognition task, judging an R^- alternative as recognized took on average 164 msec longer ($M_{Mdn} = 856$ msec, $SE = 28$) than judging an R^+ alternative as recognized ($M_{Mdn} = 692$, $SE = 13$; 95% CI on the mean difference [124, 204]) ($N = 49$).

STUDY 4
A Competition of Strategies for Multialternative Inference: Modeling Individual Differences

In situations with many alternatives, people might integrate only a few cues rather than many (see Ford et al., 1989; Payne et al., 1993). Next, we pit the recognition heuristic against corresponding models for consideration-set identification: weighted-best-cues and take-one-cue (Table 1). In contrast to the recognition heuristic, these two compensatory strategies trade off recognition against cues.

Method

Twenty-seven participants (44% female; mean age = 25 years, $SD = 3.7$) filled out a questionnaire in the labs of the Max Planck Institute for Human Development approximately 2 years after the 2005 German national election. We modified the design of Study 2 slightly: In a ranking task, participants ranked the 25 parties that ran in this election according to the election outcome they would expect if the next national election were to take place on the subsequent day. They also completed a recognition task and a detailed recognition task, which served to classify parties as R^+, R^-, or U (identical to those in Study 2). Besides individual differences in the weighting of recognition (as modeled with individual values for the cutoff criterion C), the cues people use to forecast elections may differ from person to person *and* from party to party. For instance, a person may consider the fact that a certain party has lost in past elections to be most informative, but in the case of another party the same person may take the fact that it lacks a charismatic leader as more indicative of electoral success. Another person may consider other cues more relevant. In a cue-knowledge task, we presented participants a list of the party names. For each party they had some knowledge about, participants identified the cue they considered to be most rel-

evant for inferring the election outcomes for that particular party. They described this cue on a blank line beside the party name and assigned a weight (w) to it (scale ranging from -100 to 100, with -100 signifying that the cue strongly indicated that this party would win few votes, 0 signifying that the cue implied neither many nor few votes, and 100 signifying that the cue strongly indicated that the party would win many votes). Completing all tasks took about 30 min.

Results and Discussion

Model Comparisons 7 and 8. We implemented weighted-best-cues (Table 1) as follows: If a participant had knowledge about a recognized party (R^+), *and* if the most relevant cue for this particular party was assigned a weight (w) below this participant's individual cutoff criterion C_5, that participant would rank the recognized party lower than all unrecognized ones (U). Conversely, if the cue had a weight above or equal to C_5, that participant would rank the recognized party higher than all unrecognized ones. The intuition is that a strongly weighted cue compensates for being recognized, similar to take-one-cue—which, however, always considers the same cue for all alternatives.

To implement take-one-cue, we assumed that people had some knowledge about the characteristics of R^+ parties: If—according to a participant's individual cutoff criterion C_2—an R^+ party belonged to the smaller parties that commonly attract fewer votes in elections, the participant would rank this party lower than all unrecognized parties. If—according to C_2—the R^+ party was large and successful, the rank assigned would be higher than the ranks given to the unrecognized parties. We classified each participant's R^+ parties as "large, successful" or "small, unsuccessful," according to the share of votes the parties had won in the 2005 national election. (Other classifications, based on polls conducted prior to this election, or the results of the 2002 national election, yielded similar results.)

As in Study 3, we tested the models on R^+U pairs where—depending on how C_2 and C_5 are set—the models' predictions could diverge from the recognition heuristic. First, we assessed the models' generalizability in a cross-validation. To generate the R^+U pairs, for each participant, we split the parties 10,000 times randomly into two halves, one representing the calibration set to compute the participant's optimal values for C_2 and C_5, respectively, and the other representing the validation set, with which the generalizability of the models could be assessed. As in Study 3, roughly the same number of recognized parties with knowledge (R^+), as well as the same number of unrecognized parties (U), was included in each set. For weighted-best-cues, we additionally made sure that roughly the same number of parties with weights (w) was included in each set, excluding those without weights. This resulted in an average of 25.63 ($SE = 1.21$) virtual paired comparisons per set for weighted-best-cues, and 26.26 ($SE = 1.14$) for take-one-cue. In the calibration set, for each participant, we calculated the optimal values for C_2 and C_5. Optimal values were those at which a model's accordance rate was greatest. We derived them by exhaustively searching each participant's entire parameter space of values observed for C_2 and C_5 in that participant's calibration set. (If multiple values were optimal, we picked

one at random.) In this way, we estimated a total of 52 parameter values (26 participants $\times$ 2 models), which we used to compute the take-one-cue and weighted-best-cues accordance rates, respectively, in the validation set. For each participant, we also computed the recognition heuristic accordance rate in each partition of parties.

The recognition heuristic is not only the simpler model; it also predicts people's inferences better. Computed as the average over 10,000 partitions, its accordance rate was larger than that of weighted-best-cues for 25 of 26 participants, and larger than that of take-one-cue for all 26 participants (Figures 10 and 11).

Next we assessed the models' ability to fit existing data. To this end, we recalculated the optimal value for C_2 and C_5 for each participant's complete set of virtual paired comparisons (i.e., for $M = 102.50$ R^+U pairs with weights [$SE = 4.83$] for weighted-best-cues; $M = 105.08$ R^+U pairs [$SE = 4.56$] for take-one-cue) and computed the associated take-one-cue and weighted-best-cues accordance rates on the complete set of pairs (exhaustively searching each participant's entire parameter space of values observed for C_2 and C_5 in that participant's complete set of pairs and reestimating the 52 free parameter values). Since the recognition heuristic and its two competitors are nested, when C_2 and C_5 are optimal the competitors' accordance rates can *never* be smaller than the recognition heuristic accordance. However, the competitors' accordance rates can exceed the recognition heuristic accordance rate. Did they?

For weighted-best-cues, this was the case for 9 of 26 participants (Figure 10). For 14 participants, the recognition weight was noncompensatory; that is, the optimal value for C_5 resulted in weighted-best-cues always ranking recognized parties higher than unrecognized ones, just as the recognition heuristic does. Since the models' predictions are thus identical for these participants, in Figure 10 the accordance rates for the two models are the same. Similar results emerged for take-one-cue, which yielded the best fit in 13 participants (Figure 11).

Consideration-set identification. Recall that, by ranking all recognized parties higher than unrecognized ones, the recognition heuristic identifies consideration sets of recognized alternatives. In doing so, the heuristic ignores cues that might be available beyond recognition. However, once the consideration set is generated, the alternatives *within* the consideration set can be ranked on the basis of cues; and, in fact, people appear to have used cues to rank recognized parties within the consideration set. Parties with a larger cue weight (w) were ranked higher than those with a lower one in, on average, 84% ($SE = 2$) of all virtual comparisons (mean number of pairs = 46.92, $SE = 5.28$, $n = 26$) between two R^+ parties that differed in the weight.

To summarize Studies 3 and 4, first, in generalizing to new data, the recognition heuristic predicted most participants' behavior better than five compensatory heuristics did. Second, in fitting existing data, the recognition heuristic accounted for a majority of participants' behavior as well as the competing heuristics did. Third, the recognition heuristic was the simpler model, making use of no free parameters to predict individual behavior. Judged on these model selection criteria (see Jacobs & Grainger, 1994;

Marewski & Olsson, 2009; Pitt et al., 2002), it is the winner of the model competition for a majority of participants.

GENERAL DISCUSSION

Much research has investigated how people make decisions on the basis of a sense of recognition, fluency, availability, familiarity, or accessibility (e.g., Bruner, 1957; Jacoby & Dallas, 1981; Tversky & Kahneman, 1973). In this article, we contribute to this literature. We reformulated the recognition heuristic for tasks with multiple alternatives and explored when people may employ it and when they rely on other strategies instead. Opening up the search for a better model than the recognition heuristic, we formally specified a range of competing compensatory models and tested them in eight model comparisons in four studies.

Elimination by Recognition

We showed that people make inferences consistent with a generalization of the recognition heuristic when ranking up to 25 alternatives. This generalization complements another, which suggested that the heuristic may be used when the choice is among up to four alternatives (Frosch, Beaman, & McCloy, 2007; see also McCloy, Beaman, & Smith, 2008). In particular, we proposed that the recognition heuristic identifies consideration sets of recognized alternatives with large criterion values, which can be evaluated using other heuristics. By eliminating all unrecognized alternatives, the recognition heuristic reduces complexity; since people are unlikely to have knowledge about unrecognized alternatives, putting them aside or assigning their criterion values at random does not necessarily imply a loss of accuracy. In fact, we found that the accuracy of voters' election forecasts for unrecognized parties was at chance level (Study 1). Moreover, more knowledgeable voters did not buy an increase in accuracy by sometimes going against the recognition heuristic. At the same time, less knowledgeable voters actually impaired their accuracy by not always adhering to it (for similar results in two-alternative inference, see Pachur & Hertwig, 2006).

The hypothesis that people rely on the recognition heuristic is consistent with work in consumer choice showing that priming familiar brands increases the probability that they will be considered for purchase. Even just a single exposure can lead people to consider buying novel brands (Coates, Butler, & Berry, 2004, 2006). The hypothesis that people rely on the recognition heuristic is also consistent with theories of consideration-set identification that assume noncompensatory heuristics (Kohli & Jedidi, 2007). In fact, the recognition heuristic was initially proposed as the first step in *take-the-best* (Gigerenzer & Goldstein, 1996), a noncompensatory heuristic for two-alternative decisions. Hogarth and Karelaia (2005) generalized take-the-best to the multialternative case, proposing a model called *deterministic elimination by aspects*. In doing so, they essentially built on Tversky's (1972) elimination-by-aspects model. It is important to realize, however, that take-the-best differs from elimination by aspects. For instance, the latter is a model of preferential choice that does not include a rule for how the cue (i.e., aspect) weights are
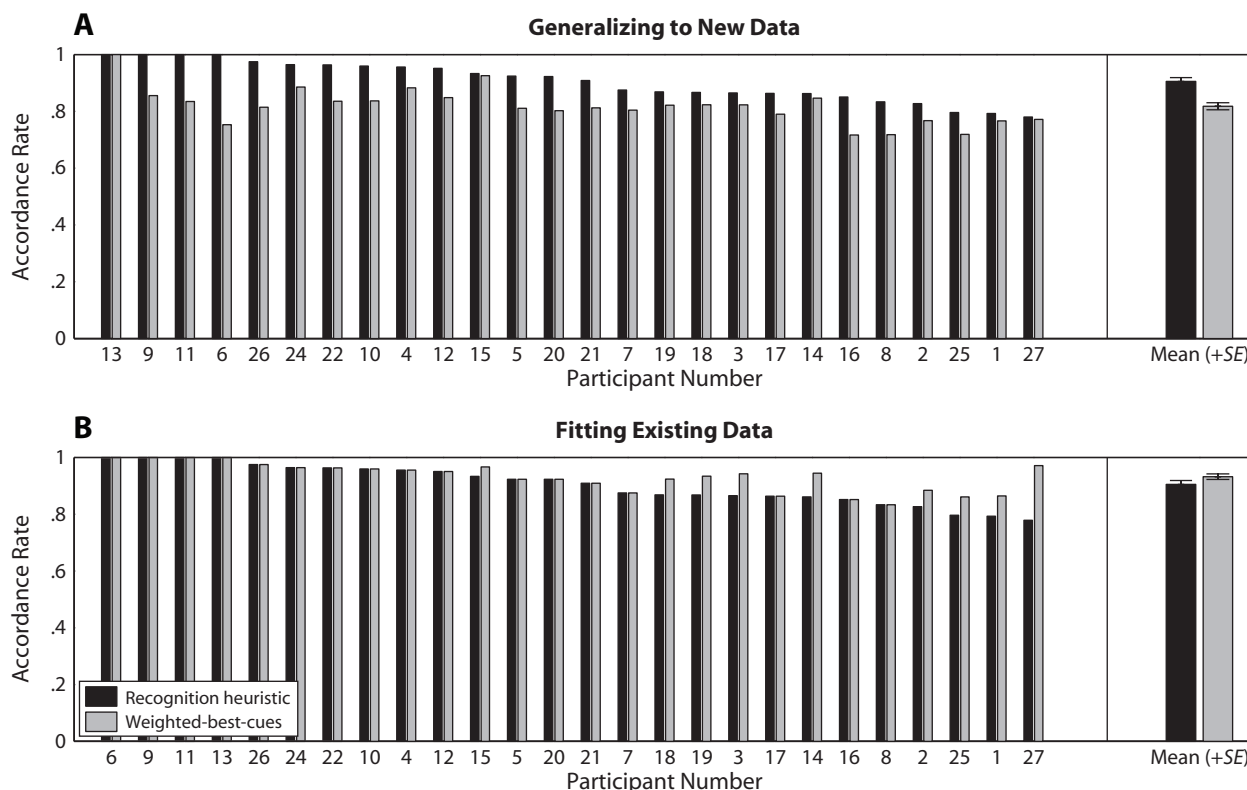
Figure 10. Recognition heuristic versus weighted-best-cues in a ranking task. (A) Generalizing to new data: Bars show means of participants' accordance rates computed in the validation set across 10,000 random partitions of each participant's data. For all participants except 1 (13), the recognition heuristic predicted participants' rankings best. In Participant 13, the weights were such that both models always predicted the same rankings. (B) Fitting existing data: Bars show participants' accordance rates for the two models. For 14 participants, the recognition weight was noncompensatory, resulting in weighted-best-cues always ranking parties in exactly the same way as the recognition heuristic does. For 9 participants, compensatory weights resulted in the better fit for weighted-best-cues. For 3 participants (7, 8, 20), compensatory weights yielded the same fit as noncompensatory weights. In both panels, no data are included for Participant 23 because, unlike all other participants, this person did not identify knowledge ($n = 26$; Study 4).

computed. Instead, it has an aspiration level for each cue. Take-the-best, in contrast, is a model of inference operating on cues with binary values and a specified order of cues. The present extension of the recognition heuristic complements these earlier models. Specifically, the recognition heuristic could be the first step in deterministic elimination by aspects in which one would eliminate all unrecognized alternatives. In a second step, one would evaluate recognized alternatives based on the best cues. In fact, although the cues people considered to be best played little role in the elimination of unrecognized parties, they predicted their rankings within the consideration set of recognized parties (Study 4).

Strategy Selection by Default?

In keeping with theories that assume memory processes as major determinants of decision behavior (e.g., Dougherty, Gettys, & Ogden, 1999; Pleskac, 2007; Schooler & Hertwig, 2005), we suggested that a possible default use of the recognition heuristic could be overruled when a weak memory activation causes recognized alternatives' retrieval fluency to be low. Using ACT–R's memory model to quantitatively predict people's recognition and retrieval time distributions, as well as their knowledge, Marewski

and Schooler (2009) found strong correlations between the probability of a person recognizing an alternative, its retrieval time, and the probability of retrieving knowledge about it; that is, alternatives about which people are likely to recall knowledge (R^+) tend also to be more strongly activated in memory and more quickly retrievable than recognized alternatives about which no knowledge is available (R^-). As a result, it could often be easier to make use of recognition in pairs that include an R^+ alternative and an unrecognized alternative (R^+U pairs) than in R^-U pairs. In fact, as we have shown, not only are R^+ alternatives recognized more quickly than R^- alternatives are, but inferences are also made faster on R^+U than on R^-U pairs, and people's inferences are more likely to agree with the recognition heuristic on R^+U than on R^-U pairs. At the same time, recognition validities tend to be larger on R^+U than on R^-U pairs; that is, it is actually ecologically rational to rely on recognition more on R^+U pairs (Studies 2 and 3). Our results are consistent with findings by others (Hertwig et al., 2008; Newell & Fernandez, 2006), suggesting that alternatives' retrieval fluency affects people's use of recognition. This thesis is also supported by fMRI data. Volz et al. (2006) found that inferences in accordance with the recognition heuristic correlated with higher acti-
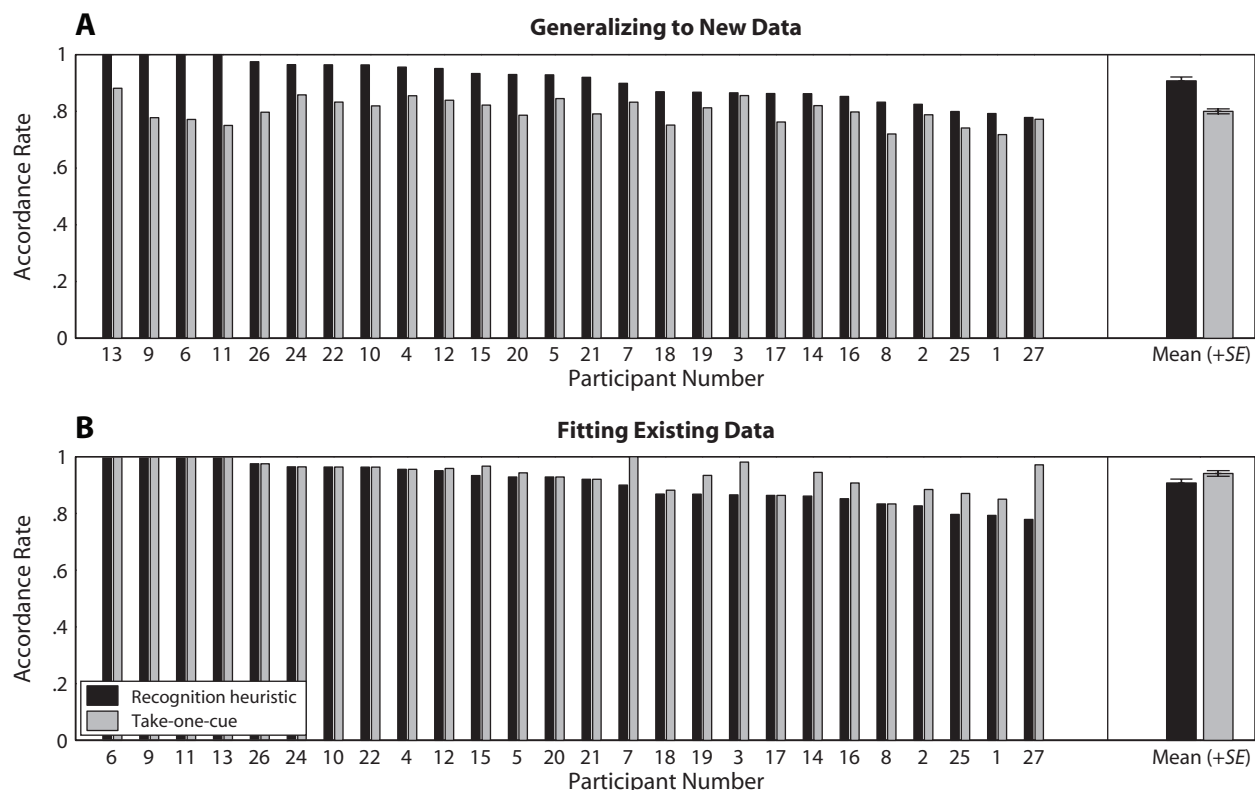
Figure 11. Recognition heuristic versus take-one-cue in a ranking task. (A) Generalizing to new data: Bars show means of partici-pants' accordance rates computed in the validation set across 10,000 random partitions of each participant's data. For all participants, the recognition heuristic predicted participants' rankings best. (B) Fitting existing data: Bars show participants' accordance rates for the two models. For 12 participants, the recognition weight was noncompensatory, resulting in take-one-cue always ranking parties in exactly the same way as the recognition heuristic does. For 13 participants, compensatory weights resulted in the better fit for take-one-cue. For Participant 8, compensatory weights yielded the same fit as noncompensatory weights. In both panels, no take-one-cue accor-dance rate is shown for Participant 23 because, unlike all other participants, this person did not identify knowledge ($n = 26$; Study 4).

vation in brain areas that had previously been associated with greater recognition confidence.

Importantly, we found that the recognition heuristic pre-dicts people's decisions better than do compensatory strate-gies on R^+U and R^-U pairs. These model tests provide some evidence against plausible alternative explanations for our and Pohl's (2006) findings—namely, that systematic dif-ferences in people's adherence to the recognition heuristic on R^+U and R^-U pairs are an artifact produced by people's *overall* use of compensatory strategies. Rather, our data are consistent with the hypothesis that these differences are a result of occasional suspensions of the recognition heuris-tic, caused by a low retrieval fluency of the R^- alternative on R^-U pairs, which in turn results from a low memory activation of the R^- alternative. However, as we will discuss shortly, our data do not rule out the possibility that people relied on another class of strategies we did not consider here. Moreover, we do not yet have a solid understanding of *how* a low retrieval fluency leads to suspensions of the recogni-tion heuristic, although there are potential explanations. For example, Marewski and Schooler (2009) assumed that raw retrieval times do not *directly* compensate for being rec-ognized (as implied by weighted-fluency); rather, retrieval times are first filtered by a psychophysical mechanism for time perception, which is modeled as an internal clock in

ACT–R (Taatgen, Van Rijn, & Anderson, 2007), and which could govern the suspension process. Another possibility consistent with ACT–R is that weak memory activations result in occasional retrieval failures for R^- alternatives, leading people to feel less confident about their recognition judgment, and in the extreme case, to mistakenly judge R^- alternatives as unrecognized, resulting in lower recognition heuristic accordance rates on R^-U pairs. Finally, the ad-herence to the recognition heuristic may not depend only on the retrieval times for recognized alternatives, but also on the confidence with which people judge alternatives as unrecognized. Although it was beyond the scope of this ar-ticle to test more detailed models that take these and other complexities into account, we hasten to stress that without them, our suspension hypothesis remains incomplete.

From Recognition to Decisions: A Competition Among Models

Addressing concerns about the adequacy of the recog-nition heuristic as a model of behavior (e.g., Dougherty et al., 2008; Newell & Shanks, 2004; Pohl, 2006; see above), and opening up the search for a better model, we examined how well five competing models predict behav-ior (Studies 1–4). We found that the recognition heuristic predicts the majority of people's inferences best.

Would other model selection criteria have yielded different results? As is well known in the model selection literature, the method employed will influence the outcomes of model comparisons. In fact, when one exclusively considers the models' ability to fit existing data, some people's inferences are best accounted for by the competing models (Figures 6–11), lending support to the thesis that there are individual differences in people's reliance on recognition (Hilbig, 2008; Pachur et al., 2008; Pachur, Mata, & Schooler, 2009), just as there may be individual differences in people's use of other decision strategies (e.g., Bergert & Nosofsky, 2007; Bröder & Gaissmaier, 2007; Cokely & Kelley, 2009; Mata, Schooler, & Rieskamp, 2007; Rieskamp & Otto, 2006).

We would like to stress that, prior to carrying out the model tests, we did not anticipate that the recognition heuristic would fare so well. For example, at least one of us would have bet on fluency-based models. Weighted-fluency seemed to be a promising model, because it takes retrieval times into account, which could have allowed it to explain why people choose recognized alternatives less often when retrieval times are longer on R^-U than on R^+U pairs. Possibly, our measure for retrieval times was too noisy in Study 3, where we assessed them in terms of RTs. On the other hand, using the same measure in a similar study, Marewski (2008) found that a substantial number of people made decisions in accordance with weighted-fluency, suggesting that RTs per se are not a bad measure (see also Hertwig et al., 2008), and leaving open the possibility that fluency-based models other than weighted-fluency may be able to explain people's decisions on both R^-U and R^+U pairs.

This is the first time that formal model comparisons have been conducted for the recognition heuristic. In carrying out these comparisons, we followed Jacobs and Grainger's (1994) research stratagem of *nested modeling* (see also Grainger & Jacobs, 1996). According to this stratagem, any new model should be related to, or include, its own direct precursors (e.g., as special cases) and be tested on data that the old model was able to account for. Here, we tested all models in both new tasks to which the recognition heuristic had not previously been applied (i.e., political elections; ranking of 15 to 25 alternatives), as well as in the task for which the recognition heuristic had originally been proposed (i.e., inferences about city size; 2AFC). Moreover, all alternative models tested here can both completely mimic the recognition heuristic *and* predict the very opposite data pattern; that is, depending on how the free parameters (i.e., the weights) in the alternative models are set, they will treat recognition, cues, and retrieval fluency in a noncompensatory or in a compensatory manner. In short, the tested alternative models not only formally include the recognition heuristic as a special case but also cover a larger hypothesis space (i.e., they are more general than the recognition heuristic is).

We hasten to add that we see the present modeling exercise as only a first step in what will hopefully evolve into a more extensive series of model comparisons in the literature. The models tested here represent various formalizations of the alternative hypotheses to the recognition heuristic that have been discussed in the literature. Although it

has been critical to examine how well these models actually fare in comparison to the recognition heuristic, in the future other models should be considered as well; for instance, it is possible that a person chooses recognized alternatives over unrecognized ones while still integrating knowledge into her decisions. Rather than directly impacting on the person's decisions, such an integration of knowledge could result in her being less confident in her decisions. Future model comparisons could also involve decision strategies that probabilistically (rather than deterministically) rank recognized alternatives above or below recognized ones. Moreover, besides considering other models, other methods for testing the models should be examined. For example, in the present tests, we assessed people's recognition and knowledge in separate tasks and used this information to predict their decisions in inference tasks. Although we believe it to be plausible that the information elicited in the recognition and knowledge tasks is used by people in the inference task, this may not necessarily be the case.

To conclude, until more model tests have been conducted, Newell and Fernandez's (2006) conclusion may be the best one we can draw: As they have pointed out, the past findings on the recognition heuristic can be interpreted in two ways. One is that some of these findings challenge the recognition heuristic's plausibility. Yet another, more in line with the results of the present model comparisons, is that the findings point to the mechanisms that determine when people rely on the recognition heuristic and when they adopt other strategies.

REFERENCES

Alba, J. W., & Chattopadhyay, A. (1985). Effects of context and part-category cues on recall of competing brands. *Journal of Marketing Research*, **22**, 340-349.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, **111**, 1036-1060.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory & Language*, **38**, 341-380.

Bartels, L. M. (2000). Partisanship and voting behavior, 1952–1996. *American Journal of Political Science*, **44**, 35-50.

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. In L. R. Beach (Ed.), *Image theory: Theoretical and empirical foundations* (pp. 145-158). Mahwah, NJ: Erlbaum.

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 107-129.

Bröder, A., & Eichler, A. (2006). The use of recognition information and additional cues in inferences from memory. *Acta Psychologica*, **121**, 275-284.

Bröder, A., & Gaissmaier, W. (2007). Sequential processing of cues

in memory-based multiattribute decisions. *Psychonomic Bulletin & Review*, **14**, 895-900.

BRUNER, J. S. (1957). On perceptual readiness. *Psychological Review*, **64**, 123-152.

COATES, S. L., BUTLER, L. T., & BERRY, D. C. (2004). Implicit memory: A prime example for brand consideration and choice. *Applied Cognitive Psychology*, **18**, 1195-1211.

COATES, S. L., BUTLER, L. T., & BERRY, D. C. (2006). Implicit memory and consumer choice: The mediating role of brand familiarity. *Applied Cognitive Psychology*, **20**, 1101-1116.

COKELY, E. T., & KELLEY, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment & Decision Making*, **4**, 20-33.

DAWES, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, **34**, 571-582.

DOUGHERTY, M. R. P., FRANCO-WATKINS, A. N., & THOMAS, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychological Review*, **115**, 199-213.

DOUGHERTY, M. R. P., GETTYS, C. F., & OGDEN, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, **106**, 180-209.

EINHORN, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, **73**, 221-230.

EINHORN, H. J., & HOGARTH, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior & Human Performance*, **13**, 171-192.

FISHBURN, P. C. (1974). Lexicographic orders, utilities and decision rules: A survey. *Management Science*, **20**, 1442-1471.

FORD, J. K., SCHMITT, N., SCHECHTMAN, S. L., HULTS, B. M., & DOHERTY, M. L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior & Decision Processes*, **43**, 75-117.

FROSCH, C. A., BEAMAN, C. P., & McCLOY, R. (2007). A little learning is a dangerous thing: An experimental demonstration of ignorance-driven inference. *Quarterly Journal of Experimental Psychology*, **60**, 1329-1336.

FUM, D., DEL MISSIER, F., & STOCCO, A. (2007). The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research*, **8**, 135-142.

GALEF, B. G., JR. (1987). Social influences on the identification of toxic foods by Norway rats. *Animal Learning & Behavior*, **15**, 327-332.

GIGERENZER, G., & GOLDSTEIN, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, **103**, 650-669.

GIGERENZER, G., HOFFRAGE, U., & GOLDSTEIN, D. G. (2008). Fast and frugal heuristics are plausible models of cognition: Reply to Dougherty, Franco-Watkins, and Thomas (2008). *Psychological Review*, **115**, 230-239.

GOLDSTEIN, D. G., & GIGERENZER, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 37-58). New York: Oxford University Press.

GOLDSTEIN, D. G., & GIGERENZER, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, **109**, 75-90.

GOODMAN, L. A., & KRUSKAL, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, **49**, 732-769.

GRAINGER, J., & JACOBS, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, **103**, 518-565.

GRONLUND, S. D., & RATCLIFF, R. (1989). The time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 846-858.

HAUSER, J. R., & WERNERFELT, B. (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research*, **16**, 393-408.

HERTWIG, R., HERZOG, S. M., SCHOOLER, L. J., & REIMER, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **34**, 1191-1206.

HERTWIG, R., & TODD, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. Hardman & L. Macchi (Eds.), *Thinking:*

Psychological perspectives on reasoning, judgment and decision making (pp. 213-231). Chichester, U.K.: Wiley.

HERZOG, S. M., & HERTWIG, R. (2010). *The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition*. Manuscript submitted for publication.

HILBIG, B. E. (2008). Individual differences in fast-and-frugal decision making: Neuroticism and the recognition heuristic. *Journal of Research in Personality*, **42**, 1641-1645.

HILBIG, B. E., & POHL, R. F. (2008). Recognizing users of the recognition heuristic. *Experimental Psychology*, **55**, 394-401.

HILBIG, B. E., & POHL, R. F. (2009). Ignorance- vs. evidence-based decision making: A decision time analysis of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **35**, 1296-1305.

HILBIG, B. E., POHL, R. F., & BRÖDER, A. (2009). Criterion knowledge: A moderator of using the recognition heuristic? *Journal of Behavioral Decision Making*, **22**, 510-522.

HINTZMAN, D. L. (1988). Judgment of frequency and recognition memory in a multiple trace memory model. *Psychological Review*, **95**, 528-551.

HINTZMAN, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 39-56). Hillsdale, NJ: Erlbaum.

HINTZMAN, D. L., & CURRAN, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory & Language*, **33**, 1-18.

HOGARTH, R. M. (1987). *Judgment and choice: The psychology of decision*. Chichester, U.K.: Wiley.

HOGARTH, R. M., & KARELAIA, N. (2005). Simple models for multiattribute choice with many alternatives: When it does and does not pay to face trade-offs with binary attributes. *Management Science*, **51**, 1860-1872.

HOGARTH, R. M., & KARELAIA, N. (2007). Heuristics and linear models of judgment: Matching rules and environments. *Psychological Review*, **114**, 733-758.

HOWARD, J. A., & SHETH, J. N. (1969). *The theory of buyer behavior*. New York: Wiley.

JACOBS, A. M., & GRAINGER, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 1311-1334.

JACOBY, L. L., & BROOKS, L. R. (1984). Nonanalytic cognition: Memory, perception and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 1-47). New York: Academic Press.

JACOBY, L. L., & DALLAS, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **110**, 306-340.

KOHLI, R., & JEDIDI, K. (2007). Representation and inference of lexicographic preference models and their variants. *Marketing Science*, **26**, 380-399.

LEWANDOWSKY, S. (1993). The rewards and hazards of computer simulations. *Psychological Science*, **4**, 236-243.

MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.

MAREWSKI, J. N. (2008). *Ecologically rational strategy selection*. Unpublished doctoral dissertation, Free University, Berlin.

MAREWSKI, J. N., GAISSMAIER, W., & GIGERENZER, G. (2010). Good judgments do not require complex cognition. *Cognitive Processing*, **11**, 103-121.

MAREWSKI, J. N., GAISSMAIER, W., SCHOOLER, L. J., GOLDSTEIN, D. G., & GIGERENZER, G. (2009). Do voters use episodic knowledge to rely on recognition? In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2232-2237). Austin, TX: Cognitive Science Society.

MAREWSKI, J. N., & OLSSON, H. (2009). Beyond the null ritual: Formal modeling of psychological processes. *Zeitschrift für Psychologie/ Journal of Psychology*, **217**, 49-60.

MAREWSKI, J. N., & SCHOOLER, L. J. (2009). *Cognitive niches: An ecological model of emergent strategy selection*. Manuscript submitted for publication.

MAREWSKI, J. N., SCHOOLER, L. J., & GIGERENZER, G. (2010). Five principles for studying people's use of heuristics. *Acta Psychologica Sinica*, **42**, 72-87.

MATA, R., SCHOOLER, L. J., & RIESKAMP, J. (2007). The aging decision

maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology & Aging*, **22**, 796-810.

McCloy, R., Beaman, C. P., & Smith, P. T. (2008). The relative success of recognition-based inference in multi-choice decisions. *Cognitive Science*, **32**, 1037-1048.

McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 563-582.

Newell, B. R., & Fernandez, D. (2006). On the binary quality of recognition and the inconsequentiality of further knowledge: Two critical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, **19**, 333-346.

Newell, B. R., & Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 923-935.

Oppenheimer, D. M. (2003). Not so fast! (and not so frugal!): Rethinking the recognition heuristic. *Cognition*, **90**, B1-B9.

Pachur, T. (in press). Recognition-based inference: When is less more in the real world? *Psychonomic Bulletin & Review*.

Pachur, T., & Biele, G. (2007). Forecasting from ignorance: The use and usefulness of recognition in lay predictions of sports events. *Acta Psychologica*, **125**, 99-116.

Pachur, T., Bröder, A., & Marewski, J. N. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making*, **21**, 183-210.

Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 983-1002.

Pachur, T., Mata, R., & Schooler, L. J. (2009). Cognitive aging and the use of recognition in decision making. *Psychology & Aging*, **24**, 901-915.

Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J., & Goldstein, D. G. (in press). Is ignorance an adaptive tool? A review of recognition heuristic research. In P. M. Todd, G. Gigerenzer, & the ABC Research Group (Eds.), *Ecological rationality: Intelligence in the world*. New York: Oxford University Press.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.

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

Pleskac, T. J. (2007). A signal detection analysis of the recognition heuristic. *Psychonomic Bulletin & Review*, **14**, 379-391.

Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, **19**, 251-271.

Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology*, **21**, 139-155.

Richter, T., & Späth, P. (2006). Recognition is used as one cue among others in judgment and decision making. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 150-162.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, **135**, 207-236.

Roberts, J. H., & Lattin, J. M. (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research*, **28**, 429-440.

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, **112**, 610-628.

Serwe, S., & Frings, C. (2006). Who will win Wimbledon? The recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making*, **19**, 321-332.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior*, **6**, 156-163.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, **69**, 99-118.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, **25**, 207-222.

Taatgen, N., Van Rijn, H., & Anderson, J. R. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention, and learning. *Psychological Review*, **114**, 577-598.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, **79**, 281-299.

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

Volz, K. G., Schooler, L. J., Schubotz, R. I., Raab, M., Gigerenzer, G., & von Cramon, D. Y. (2006). Why you think Milan is larger than Modena: Neural correlates of the recognition heuristic. *Journal of Cognitive Neuroscience*, **18**, 1924-1936.

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1235-1253.

Yee, M., Dahan, E., Hauser, J., & Orlin, J. (2007). Greedoid-based noncompensatory inference. *Marketing Science*, **26**, 532-549.

NOTES

1. Pachur and Biele's (2007) analyses differ from the kind of comparative model tests we have in mind. They did not assess individual participants' knowledge about the alternatives, and, as a result, all competing models made the same predictions for all participants. Especially in our Studies 3 and 4, we tailor the models to individual participants' information about the alternatives, assessing for each participant optimal weights for the information and comparing the models' ability to fit existing data as well as to generalize to new data. Moreover, for the first time we also pit the recognition heuristic against a compensatory model that operates on the speed of retrieving and recognizing an alternative.

2. Due to an error in the questionnaire production, one of the paired comparisons was randomly selected and replaced by another randomly generated paired comparison. This comparison was excluded from all analyses.

3. In other studies on the recognition heuristic, order effects were not found (e.g., Goldstein & Gigerenzer, 2002; Pachur & Hertwig, 2006). In our study, voters ($n = 87$) who worked on the recognition task before the 2AFC task recognized fewer candidates ($M = 4.8$, $n = 87$) than did voters who worked on the tasks in the reverse order ($M = 5.9$, $n = 85$; 95% CI on the mean difference [0.5, 1.9]). However, there were no reliable differences with regard to the accordance rate, and voters recognized similar numbers of parties ($M = 10.1$) regardless of task order.

4. Two participants were excluded because they did not complete the questionnaire. Participants working on the ranking task before the recognition task recognized about as many parties ($M = 11.8$, $n = 28$) as did those working on the tasks in the opposite order ($M = 13.0$, $n = 31$; 95% CI on the mean difference [-0.2, 2.5]). Also, the (paid) participants who filled out the questionnaire in the laboratory did not differ with regard to central variables from those participants who filled out the questionnaire in a university class.

5. Marewski and Schooler (2009) monitored participants' eye movements when collecting the data analyzed here. The eye-tracking system used (Tobii 1750) makes the experimental setup hardly distinguishable from a setup without an eyetracker. Only those trials were included in the analyses for which the total duration of gazes on the computer screen closely matched the response time for the trial. For these and other details about their experiment and data, please see Marewski and Schooler.

6. In contrast to the two tallying models and take-one-cue, weighted-fluency does not depend on knowledge. It can therefore favor unrecognized cities over recognized ones across both $R^{+}U$ pairs and $R^{-}U$ pairs. Since we envisioned weighted-fluency as an alternative to the recognition heuristic on $R^{-}U$ pairs, above we only report tests of the two models on these pairs. Testing weighted-fluency across $R^{-}U$ and $R^{+}U$ pairs does not change the pattern of results.

7. Why do the recognition heuristic accordance rates differ between Studies 2 and 3 on $R^{-}U$ pairs (.62 vs. .86)? We can only speculate about reasons. Here is one: In Study 3, the accordance rates reflect decisions in a 2AFC task. In Study 2, they reflect rankings of 24 alternatives, and this ranking task prevented participants from being intransitive, because they were instructed to uniquely identify each alternative with a rank and each rank could be assigned only once. Therefore, the ranking task may have been more difficult than the two-alternative task. At the same time, occasional suspensions and/or application errors of the recognition heuristic (or any other strategy) may have a stronger impact on accordance rates in a ranking task than in a two-alternative task, because one erroneous rank has the potential of affecting many other rankings.