An ecological perspective to cognitive limits: Modeling environment-mind interactions with ACT-R

Wolfgang Gaissmaier* and Lael J. Schooler
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development, Berlin

Rui Mata
Department of Psychology
University of Michigan

Abstract

Contrary to the common belief that more information is always better, Gigerenzer et al. (1999) showed that simple decision strategies which rely on little information can be quite successful. The success of simple strategies depends both on bets about the structure of the environment and on the core capacities of the human mind, such as recognition memory (Gigerenzer, 2004). However, the interplay between the environment and the mind's core capacities has rarely been precisely modeled. We illustrate how these environment-mind interactions could be formally modeled within the cognitive architecture ACT-R (J. R. Anderson et al., 2004). ACT-R is an integrated theory of mind that is tuned to the statistical structure of the environment, and it can account for a variety of phenomena such as learning, problem solving, and decision making. Here, we focus on studying decision strategies and show how the success of theses strategies in particular environments depends on characteristics of core cognitive capacities, such as recognition and short term memory.

Keywords: Simple heuristics, ecological rationality, cognitive limits, ACT-R.

1 Introduction

The six million dollar man was among the most popular television shows in the 1970's, at least among eight to twelve year old American boys. The credits open with a spectacular crash of a rocket jet tumbling into a fireball. A team of surgeons hovers over Steve Austin, test pilot, working feverishly to replace his injured legs, right arm and eye with superbly engineered "bionic" substitutes. In the voice over we hear: "Gentlemen, we can rebuild him. We have the technology. We have the capability to make the world's first bionic man. Steve Austin will be that man. Better than he was before. Better, stronger, faster." We cut next to Steve Austin, six million dollar man, racing across a field and seeing objects at a distance with his better-than-a-telescope eye.

Like Steve Austin's doctors, most of us believe we would be better off if we were stronger and faster. Similarly, we would be better off, or at least, would have had better college transcripts if we had been blessed with bionic cognitive abilities, such as unfailing memories and the ability to hold complex equations in mind. Another view on the humble cognitive capacities of the human mind is that those limitations — such as forgetting —

may serve important functions. Arguably, the most important function of memory is not simply to store all information we encounter, but to provide us with important information in specific situations. In this view, the human memory system is organized in a way which facilitates the retrieval of information which is recent or frequent (J. R. Anderson & Schooler, 1991) and sensitive to the context (Schooler & Anderson, 1997). In this way, the system retrieves the memories, that is, the information we are most likely to need.

Many word processors incorporate a timesaving feature that illustrates this view of forgetting. When a user goes to open a document file, the program presents a "file buffer," a list of recently opened files from which the user can select. Whenever the desired file is included on the list, the user is spared the effort of searching through the file hierarchy. For this device to work efficiently, however, the word processor must provide users with the files they actually want. It does so by "forgetting" files that are considered unlikely to be needed on the basis of the assumption that the time since a file was last opened is negatively correlated with its likelihood of being needed now. Similarly, if you want to remember where you have parked your car, it is quite useful to forget where you have parked before. There is growing evidence also from other domains (such as language acquisition) that cognitive limits can be beneficial (for an overview, see Hertwig & Todd, 2003) while too much thinking can even hurt

*Correspondence concerning this article should be addressed to Wolfgang Gaissmaier, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. Email: gaissmaier@mpib-berlin.mpg.de.

performance, for example for sports experts (Beilock, Bertenthal, McCoy, & Carr, 2004) and in implicit category learning (DeCaro, Thomas, & Beilock, in press).

In line with this view of the mind as an adaptation to the environment, the program on Fast and Frugal heuristics takes the position that humans possess a repertoire of cognitive strategies, or heuristics, which can solve specific problems (e.g., Gigerenzer, Todd, and the ABC research group, 1999). Gigerenzer et al. called this collection of cognitive strategies the *adaptive toolbox*. The rationality of these heuristics is not logical but ecological: Success is anchored both in the structure of the environment and in the core capacities of the human mind (Gigerenzer, 2004). A cognitive strategy can be simple by exploiting the core capacities (such as recognition or recall memory) of the human mind that through evolution or learning are highly automatized, requiring little or no effort.

Goldstein and Gigerenzer (2002) "consider heuristics to be adaptive strategies that evolved in tandem with fundamental psychological mechanisms" (p. 75). Within the fast and frugal program, this interplay between the environment and the mind has rarely been explored with detailed models of core capacities. The models of heuristics presupposed specific core capacities of the mind (such as recognition memory) without embedding the core capacity directly into the model. In other words, models of heuristics (and many other models of decision making) are underspecified with regard to how decision making will be affected by the interaction between the mind's core capacities and the structure of the environment.

The goal of this paper is to illustrate how cognitive modeling can capture environment-mind interactions and thereby inform decision making research. In particular, we employ a formal cognitive architecture such as ACT-R (J. R. Anderson, Bothell, Byrne, Douglass, Lebiere & Qin, 2004) to explore this issue. In principle, other cognitive models of memory could be used for the purpose of analyzing the relation between environment, memory, and the performance of inference strategies. In fact, we believe the main findings regarding memory function and the ecological rationality of inference strategies would remain largely the same as long as a reasonable model of memory was used. One such model, REM (Shiffrin & Steyvers, 1997), descends from SAM (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981), but is enriched by a Bayesian analysis in the spirit of J. R. Anderson's rational analysis (e.g., J. R. Anderson & Schooler, 1991). Another good candidate would be MINERVA-DM (Dougherty, Gettys, & Ogden, 1999), which has been shown to account for a wide variety of judgment phenomena, including for example, the availability and the representativeness heuristics. The relation between ACT-R and other models will be discussed in more detail

in the General Discussion, where we will also illustrate how ACT-R could be extended to topics and tasks beyond those examined here.

2 ACT-R as an integrative framework

ACT-R is an integrated theory of mind which is able to account for a variety of phenomena including, for example, practice and retention (J. R. Anderson, Fincham, & Douglass, 1999), decision making (Gonzalez, Lerch, & Lebiere, 2003), language learning (Taatgen & Anderson, 2002), and probability learning (Lovett, 1998). The core of ACT-R is constituted by the declarative memory system for facts (*knowing that*) and the procedural system for rules (*knowing how*). The declarative memory system consists of chunks that represent information (e.g., about the outside world, about oneself, about possible actions, etc.). These chunks take on activations that determine their accessibility. That is, whether they can be retrieved. As a consequence of following ACT-R's standard rule for reinforcing chunks, the history of how often and when chunks have been used in the past determines their activation. The activation of a chunk is higher the more frequently and the more recently it has been used. Because activation reflects frequency and recency, different histories can lead to the same level of activation at any given moment of time.

The procedural system consists of if-then rules that model the course of action an individual could perform to solve a specific task. Given that all the conditions specified on their if-side are met, the productions execute all the actions specified on the then-side. The if-side can specify conditions in the outside world that need to be fulfilled, for example that a new object has appeared on the screen, but also internal conditions, such as that a specific chunk has been retrieved. Similarly, the actions specified on the then-side include internal actions such as trying to retrieve a chunk from the declarative system as well as actions to interact with the environment, such as looking for a new object on the computer screen or pressing a key on the keyboard.

In the next section we illustrate how a specific heuristic — the recognition heuristic — can be implemented into the ACT-R cognitive architecture. Not only has this implementation permitted investigation into environment-mind interactions (with a special focus on the impact of the mind's limitations), but it has also lead to a specification of a relative of the recognition heuristic, the fluency heuristic (cf. Jacoby & Dallas, 1981).

3 More is not always better: The recognition and the fluency heuristics

The recognition heuristic illustrates the interplay between the structure of the environment and core capacities of the human mind (Goldstein & Gigerenzer, 2002). In short, the recognition heuristic uses the information about whether an object is recognized or not to make inferences about some criterion value of this object. More specifically, the recognition heuristic can be used for paired comparisons between two objects, one recognized, the other not. It is defined as follows:

> Recognition heuristic: If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion (Goldstein & Gigerenzer, 2002, p. 76).

The recognition heuristic is simple because it can rely on the human core capacity of recognition memory. Note that this does not mean that the process of recognition is simple per se, but rather that the recognition heuristic is simple given recognition memory.

The recognition heuristic will be successful in environments in which the probability of recognizing objects is correlated with the criterion to be inferred. This is, for example, the case in many geographical domains such as city or mountain size (Goldstein & Gigerenzer, 2002), and in many competitive domains such as predicting the success of tennis players (Serwe & Frings, 2006; Scheibehenne & Bröder, 2007), or of political parties (Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer, 2008). One reason why objects with larger criterion values are more often recognized is that they are more often mentioned in the environment. There is evidence that the recognition heuristic is often in accordance with how people actually make inferences (e.g., Goldstein & Gigerenzer, 2002; Pachur & Hertwig, 2006; Pachur, Bröder & Marewski, in press; Reimer & Katsikopoulos, 2004). However, it has been heavily debated whether recognition is indeed the *only* cue that is considered in probabilistic inference (when applicable), as was originally proposed by Goldstein and Gigerenzer, or whether it is simply one cue among others, albeit a very important one (e.g., Bröder & Eichler, 2006; Newell & Fernandez, 2006; Newell & Shanks, 2004; Pohl, 2006; Richter & Späth, 2006).

To be successful, the recognition heuristic requires that a person does not recognize too much nor too little, because to be applied, only one of the alternatives needs to be recognized but not the other. If too few or too many objects are recognized, then recognition will be uninformative because it will rarely discriminate between objects. By implementing the recognition heuristic in ACT-R, Schooler and Hertwig (2005) showed that some forgetting could fuel the success of the recognition heuristic because it helps maintain the essential level of partial knowledge. The idea behind this was the following: Without forgetting, the person would, over time, recognize all of the objects. Thus, recognition is no longer a useful piece of information because it does not discriminate between objects. If, on the other hand, there were too much forgetting, a person would recognize so few objects that recognition would no longer be a useful cue. The key to success lies in recognizing some — but not all — of the objects, and forgetting helps to keep it that way, which will be demonstrated in more detail in the following.

3.1 Modeling the Recognition Heuristic within ACT-R

According to Goldstein and Gigerenzer (2002), the recognition heuristic works because there is a chain of correlations linking the criterion (e.g., city population), via environmental frequencies (e.g., how often a city is mentioned), to recognition. ACT-R's activation tracks just such environmental regularities, so that activation differences reflect, in part, these frequency differences. Thus, it appears that inferences—such as deciding which of two cities is larger—could be based directly on the activation of associated chunks (e.g., city representations). However, drawing directly on activation is prohibited in the ACT-R modeling framework for reasons of psychological plausibility: subsymbolic quantities, such as activation, are held not to be directly accessible, just as people presumably cannot make decisions on the basis of differences in the long-term potentiation of neurons in their hippocampus. Yet, the system could still capitalize on activation differences associated with various objects by gauging how it responds to them. The simplest measure of the system's response is whether a chunk associated to a specific object can be retrieved at all, and this is what Schooler and Hertwig (2005) used to implement the recognition heuristic in ACT-R.

To create their model, Schooler and Hertwig (2005) first determined the activations of the chunks associated to various German cities. Following Goldstein and Gigerenzer's (2002) original assumption that the frequency with which a city is mentioned in newspapers mirrors its overall environmental frequency, they constructed environments consisting of German cities such that the probability of encountering a city name on any given simulated day was proportional to the overall frequency with which the city was mentioned in the Chicago Tribune. The model learned about these simulated environments

by strengthening memory chunks associated with each city according to ACT-R's activation equation. In ACT-R, the activation of a chunk increases with each encounter of the item, and decays as a function of time.

Second, the model's recognition rates for the German cities were determined. Following Anderson, Bothell, Lebiere and Matessa (1998), recognizing a city was considered to be equivalent to retrieving the chunk associated with it. The model's recognition rate for a particular city was obtained by fitting the ACT-R equation that yields the probability that a chunk will be retrieved (given its activation learned in step 1) to the empirical recognition rates that Goldstein and Gigerenzer (2002) observed. These empirical recognition rates were the proportion of University of Chicago participants who recognized the city.

Third, the model was tested on pairs of German cities. To do this, the model's recognition rates were used to determine the probability that it would successfully retrieve a memory chunk associated with a city when it was presented with the city name as a retrieval cue. The successful retrieval of the chunk was taken to be equivalent to recognizing the associated city. This means that if a chunk could not be successfully retrieved (because its activation was too low), it was taken to be equivalent to not recognizing the city. Finally, the production rules for the recognition heuristic dictated that whenever one city was recognized and the other was not, the recognized one was selected as being larger, and in all other cases (both cities recognized or unrecognized) a guess was made. These decisions closely matched the observed human responses.

This implementation showed that the recognition heuristic could easily be modeled within the broader ACT-R framework with the appropriate assumptions about how recognition could be determined in the system. Once this model was in place, Schooler and Hertwig (2005) proceeded to ask a much more interesting question: Can forgetting help memory-based inferences, such as those made by the recognition heuristic, to be more accurate? The notion that forgetting serves an adaptive function has repeatedly been put forth in the history of the analysis of human memory (in line with the idea that cognitive limits may carry benefits—see Todd, Hertwig, & Hoffrage, 2005, and Hertwig & Todd, 2003). Bjork and Bjork (1988), for instance, have argued that forgetting prevents obsolete information from interfering with the recall of more current information. Altmann and Gray (2002) make a similar point for the short-term goals that govern our behavior. From this perspective, forgetting prevents the retrieval of information that is likely to be obsolete.

Schooler and Hertwig (2005) were interested in whether forgetting could enhance decision making by strengthening the usefulness of recognition. To find out, they varied forgetting rates in terms of how quickly chunk
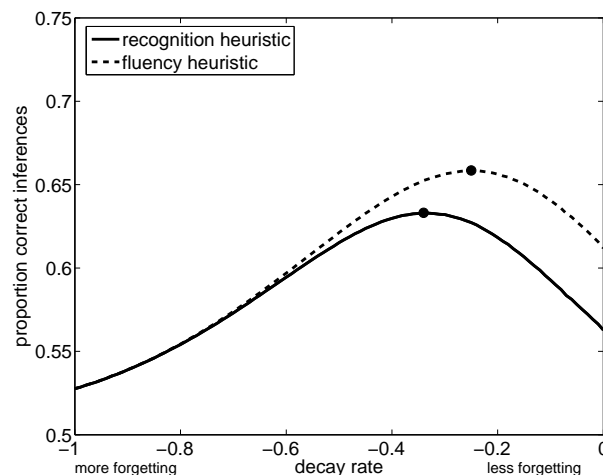

Figure 1: Performance of the recognition and fluency heuristics vary with decay rate. (Reprinted with permission from Schooler & Hertwig, 2005.)

activation decays in memory (i.e., ACT-R's parameter d), and looked at how this affects the accuracy of the recognition heuristic's inferences. The results are plotted in Figure 1, showing that the performance of the recognition heuristic peaks at intermediate decay rates. In other words, the recognition heuristic does best when the individual forgets some of what she knows—with too little forgetting, performance actually declines (as it does with too much forgetting as well, though this is what one would normally expect). This happens because intermediate levels of forgetting maintain a distribution of recognition rates that are highly correlated with the criterion, and as stated earlier, it is just these correlations on which the recognition heuristic relies.

3.2 Using continuous recognition values: The fluency heuristic

The recognition heuristic (and accordingly its ACT-R implementation) relies on a binary representation of recognition: an object is simply either recognized (and retrieved by ACT-R) or it is unrecognized (and not retrieved). But this heuristic essentially discards information when two objects are both recognized but one is recognized more strongly than the other—a difference that could be used by some other mechanism to decide between the two objects, but which the recognition heuristic ignores. Considering this situation, Schooler and Hertwig (2005) noted that recognition could also be assessed within ACT-R in a continuous fashion in terms of how quickly an object's chunk can be retrieved. This information can then be used to make inferences with a related simple mechanism, the fluency heuristic. Such a heuristic for using the fluency of reprocessing as a cue in inferen-

tial judgment has been suggested earlier (e.g., Jacoby & Dallas, 1981; Kelley & Jacoby, 1998; Whittlesea, 1993; Whittlesea & Leboe, 2003), but Schooler and Hertwig define it more precisely for the same context as the recognition heuristic, that is, selecting one of two alternatives based on some criterion on which the two can be compared. Following this version of the fluency heuristic, if one of two objects is more fluently reprocessed, then infer that this object has the higher value with respect to the criterion.

For such a heuristic to be psychologically plausible, individual decision makers must be sensitive to differences in recognition times, for instance able to tell the difference between recognizing "Berlin" instantaneously and taking a moment to recognize "Stuttgart." Schooler and Hertwig (2005) then propose that these differences in recognition time partly reflect retrieval time differences, which, in turn, reflect the base-level activations of the corresponding memory chunks, which correlate with environmental frequency, and finally with city size. Further, rather than assuming that the system can discriminate between minute differences in any two retrieval times, they allow for limits on the system's ability to do this: if the retrieval times of the two alternatives are within a just-noticeable-difference of 100 ms, then the system cannot distinguish its fluency for the alternatives and must guess between them.

Schooler and Hertwig's (2005) model of the fluency heuristic is related to the notion of availability (Tversky & Kahneman, 1973). In fact, we believe that Schooler and Hertwig's implementation of the fluency heuristic offers a definition of availability that interprets the heuristic as an ecologically rational strategy by rooting fluency in the informational structure of the environment. This precise formulation transcends the criticism that availability has been only vaguely sketched (e.g., Gigerenzer & Goldstein, 1996). Furthermore, one could argue that the notion of availability and the fluency heuristic incorporate the recognition heuristic as a special case. Namely, if one object cannot be retrieved at all (is unrecognized), this would represent an extreme case of influent reprocessing. However, we believe it to be useful to keep these two heuristics separate because the productions that implement them are different in psychologically important ways. In the case of the recognition heuristic, one can immediately decide for the recognized object, without any further thinking: The recognition heuristic is rarely in competition with knowledge-based strategies because knowledge is usually not available for the unrecognized object (but see Oppenheimer, 2003). In contrast, if both objects are recognized one cannot immediately decide between the two objects without taking another step. In this case, one could bet on fluency. However, fluency is most often in competition with knowledge-based strate-
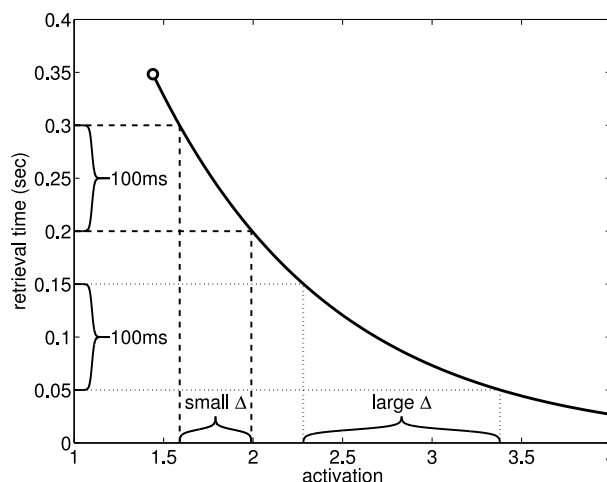

Figure 2: A chunk's activation determines its retrieval time. (Reprinted with permission from Schooler & Hertwig, 2005.)

gies which retrieve further information about the objects and may be more successful in cases where fluency is not predictive of the criterion. In what they call the cognitive niche of the fluency heuristic, Marewski and Schooler (2008) demonstrated that the fluency heuristic should and is relied upon most when knowledge about the objects (besides recognition) cannot be retrieved from memory. When, however, additional knowledge about the objects is available, then knowledge-based strategies are favored over the fluency heuristic.

The performance of the fluency heuristic turns out to be influenced by forgetting in much the same way as the recognition heuristic, as shown by the upper line in Figure 1, which shows the combined performance of the fluency and recognition heuristics. In the case of the fluency heuristic, intermediate amounts of forgetting increase the chances that differences in the retrieval times of two chunks will be detected. The explanation for this is illustrated in Figure 2, which shows the exponential function that relates a chunk's activation to its retrieval time. Forgetting lowers the range of activations to levels that correspond to retrieval times that can be more easily discriminated. In other words, a difference in activation at a lower range results in a larger, more easily detected difference in retrieval time than an activation difference of the same magnitude at a higher range.

Both the recognition and fluency heuristics can be understood as means to indirectly tap the environmental frequency information locked in the activations of chunks in ACT-R. These heuristics will be effective to the extent that the chain of correlations—linking the criterion values, environmental frequencies, activations and responses—is strong. By modifying the rate of memory decay within ACT-R, Schooler and Hertwig (2005)

demonstrated the surprising finding that forgetting actually serves to improve the performance of these heuristics by strengthening the chain of correlations on which they rely. Future research will have to tell whether these surprising benefits of forgetting also hold for other heuristics, such as Take-The-Best (Gigerenzer & Goldstein, 1996), which relies on complexes of declarative knowledge.

4 How too much thinking can hurt

The recognition and fluency heuristics are effective because recognition implicitly detects correlations in the world. More generally, detecting correlations is fundamental to making predictions. Congruent with the assumption that cognitive limits can serve important functions, Kareev and colleagues have introduced the idea that cognitive limits may actually be beneficial in the detection of correlations (Kareev, 1995a, 1995b, 2000, 2004; Kareev, Lieberman, & Lev, 1997). The idea behind this is as follows. Kareev (1995b) argued that people rely on samples from the environment to assess correlations between, for example, two dimensions of a set of objects. The size of these samples is supposed to be bounded by short-term memory capacity. In a theoretical analysis, Kareev concluded that the use of small sample sizes facilitates the early detection of correlations by amplifying them. Specifically, both the median and the mode of the sampling distribution of the Pearson correlation exceed the population correlation, and the smaller the sample, the more it does so. Building on the assumption that people's perception of correlation is the result of calculating the correlation on the basis of a sample, Kareev assumed that consideration of a small sample is more likely to result in a more extreme perception of correlation. Because people with a lower short-term memory capacity (*low spans*) consider smaller samples than those with a higher short-term memory capacity (*high spans*), the argument goes, low spans should be more likely to perceive the correlation as more extreme, and thereby detect it earlier.

Kareev and his colleagues provided experimental support for this theoretical argument by showing that low spans indeed performed better on a correlation detection task (Kareev et al., 1997). The task consisted of predicting, trial-by-trial, which of two possible symbols (X or O) an envelope (which could be either red or green) contained. The number of Xs and Os within the envelopes was varied to yield correlations ranging from $\Phi = -.60$ to $\Phi = .60$. A correlation here means that, for example, there are more Xs in red envelopes and more Os in green envelopes. Detecting this correlation helps people to increase their predictive performance. We will refer to this task as the *envelope task*. Based on the find-

ing that low spans outperformed high spans on this task, Kareev et al. concluded that people with a lower short-term memory capacity, and hence a smaller sample size to consider, "perceived the correlation as more extreme and were more accurate in their predictions" (p. 278). We will call this Kareev's *small sample hypothesis* of correlation detection.

However, the small sample hypothesis has been criticized because the advantage of small samples in correlation detection does not seem to be as general as Kareev and colleagues implied. Juslin and Olsson (2005) pointed out that the adaptive value of different sample sizes in detecting correlations is determined by the posterior probability of a hit (i.e., correctly inferring that there is a non-trivial population correlation based on a sample correlation), and not by the hit rate (i.e., detecting a non-trivial sample correlation given that there is a non-trivial population correlation). Applying this method also takes into account false alarms (i.e., believing that there is a positive correlation when it is in fact zero or negative), and demonstrates that the alleged benefits of small samples do not occur. At least the benefits are manifest only when one makes the additional assumption that people only decide that a correlation is present in the population when the correlation they observe in the sample exceeds a decision threshold and otherwise neglect it (R. B. Anderson, Doherty, Berg & Friedrich, 2005). In response to these criticisms, Kareev (2005) restricted the benefits of small samples to the detection of large correlations. However, there also existed a low capacity advantage for small correlations in Kareev et al. (1997), which then cannot be explained by Kareev's small sample hypothesis. Furthermore, research on the estimation of correlations has shown that estimates of correlations increase with sample size (e.g., Clément, Mercier, & Pasto, 2002; Shanks, 1985, 1987), which counters what would be expected by the small sample hypothesis.

Thus, an account of the low capacity advantage in Kareev et al.'s (1997) correlation detection task that follows from the small sample hypothesis is not wholly satisfying, and so it may be profitable to consider alternatives. Gaissmaier, Schooler and Rieskamp (2006) developed an alternative explanation which was drawn from the probability learning literature. The probability learning literature is concerned with tasks that are basically identical to the task used by Kareev et al., despite being simpler. In those tasks, people have to predict one of two events that occur with different probabilities. For example, event E1 could occur with a probability of $p(E1) = .75$, while event E2 only occurs with $p(E2) = 1 — p(E1) = .25$. Given that the successive events are conditionally independent, the best that people could do is to always predict the occurrence of the more more frequent event E. This strategy, called maximizing, would yield an aver-

age accuracy of 75%. However, a strategy which is very often observed is probability matching, that is, predicting events in proportion to their probability of occurrence, with an expected accuracy in this case of only 62.5% on average ($.75 \cdot .75 + .25 \cdot .25$). Probability matching is typically considered a choice anomaly in that it is not the best strategy, at least with respect to maximizing payoff. Although it is possible to make probability matching largely disappear, for example with high monetary incentives or extensive training (Shanks, Tunney, & McCarthy, 2002), it is altogether a rather robust phenomenon (for reviews, see Myers, 1976; Vulkan, 2000). Even if overmatching (i.e., predicting the more common event with a relative frequency slightly higher than the actual event probability) is often observed with monetary incentives and large numbers of trials, it seems fair to say that humans are rather slow to settle on a pure maximizing strategy.

Why do people fail to find the optimal solution in such a simple task? It is often assumed that people are not smart enough to understand the its structure. Support for this view comes from tasks in which a hypothetical probability learning task was described to participants and they had to specify, in advance, what they would do. In this situation, people with higher SAT scores (West & Stanovich, 2003) and older students (Gal & Baron, 1996) were more likely to deliberately opt for a maximizing strategy. Another common explanation of probability matching is that people would be bored by making the same prediction over and over again, as the maximizing strategy requires (e.g., Gal & Baron, 1996; Siegel & Goldstein, 1959), although it is a positive surprise to guess the infrequent event correctly (e.g., Brackbill & Bravos, 1962).

Although we find truth in all of these accounts, there is also another, very different reason why people could end up matching probabilities (which does not exclude the other accounts): Probability matching could be the result of a more complex strategy, such as exploring the hypothesis space that has the goal of improving the long run performance at the expense of short term gains. One hypothesis people typically hold in those tasks is that there are patterns in the sequence, and any reasonable pattern tends to match the probabilities (Wolford, Miller, & Gazzaniga, 2000). That people indeed search for patterns in those experiments has been nicely demonstrated by Yellott (1969). In the last block of his experiment, participants always received feedback indicating that their predictions were correct, irrespective of what they predicted. They continued to match probabilities as they did before, and when they were asked for their impressions afterwards, most responded that they finally found the pattern in the sequence. Congruently, Unturbe and Corominas (2007) showed that participants who reported to have found complex rules in a (random) sequence of binary

events were closer to probability matching behavior than those who did not report such rules.

Because there are no patterns, searching for them is of course counterproductive. One major reason why people search for patterns seems to be that they do not accept that the sequence is random, even if they are told so. Fostering the belief in randomness increases the prevalence of maximizing. This is, for example, the case if the task resembled a 'gambling' task, compared to a structurally identical task that appeared to be a 'problem solving' task (Goodnow, 1955), or if the alternation rate was slightly higher than expected by chance, which people perceive to be more random although it is actually less random (Wolford, Newman, Miller, & Wig, 2004).

This also leads to the seemingly counterintuitive finding that distracting people, and thereby preventing the search for patterns, can result in more maximizing behavior, and thus in behavior that is considered more rational. For example, Bauer (1972) reported that people who simultaneously were asked to estimate the relative frequencies explicitly while making predictions maximized more strongly (see also Neimark & Shuford, 1959, who obtained similar results). Bauer speculated that this may be due to the simplicity of the maximizing strategy which "puts less cognitive strain on the subject" (p. 206), and that this could be important when the task gets more complicated (as may be the case with the simultaneous estimation task). More direct and thereby more convincing evidence comes from Wolford et al. (2004), who found that a distracting secondary verbal working memory task resulted in more maximizing behavior.

Thus, the low capacity advantage described by Kareev et al. (1997) could be the same kind of phenomenon as the less-is-more effect in probability learning. People with lower cognitive capacities make simpler predictions, which are more successful in this task, while people with higher cognitive capacities are more likely trying to search for patterns resulting in probability matching. Given the slow learning curves in the probability learning literature, people could still well be searching for patterns (and thus matching probabilities) after several hundred trials, which is the range of the number of trials in the experiments by Kareev et al. and Gaissmaier et al. (2006). Gaissmaier et al. proposed an alternative to the small sample hypothesis, the *predictive behavior hypothesis*, which states that people with lower capacities make simpler predictions. They implemented both hypotheses in ACT-R to test these hypotheses.

4.1 Modeling simple predictions vs. exaggerated perception in ACT-R

Gaissmaier et al.'s (2006) ACT-R model is based on Logan's (1988) idea that people make predictions by retriev-
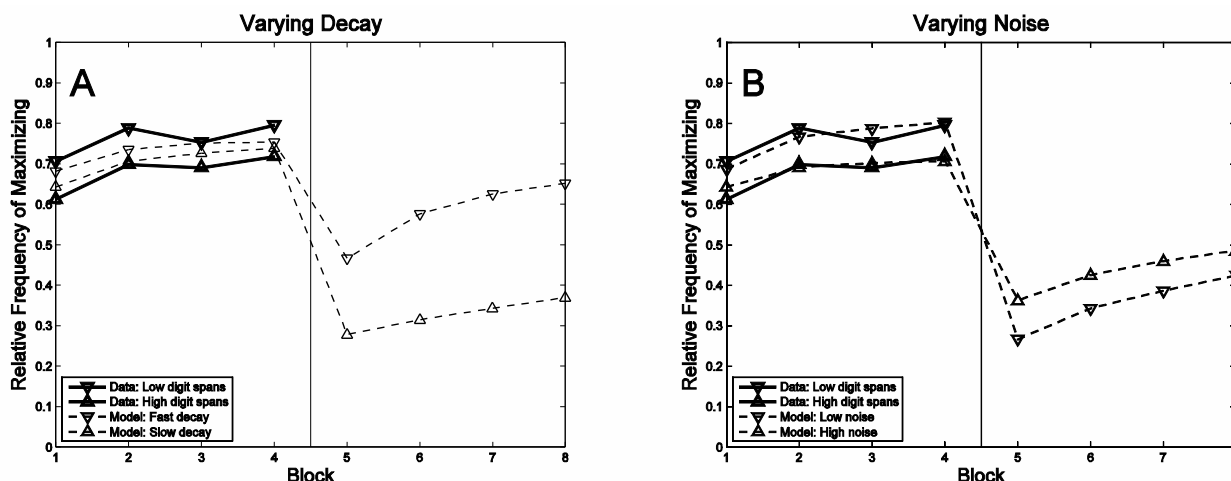
Figure 3: Model predictions of (A) the decay and (B) the noise variant. The models were fitted to data on 4 blocks of 32 trials each, and then predictions were made for behavior after a shift in the environment (indicated by the vertical line). (Reprinted with permission from Gaissmaier, Schooler, & Rieskamp, 2006.)

ing predictions from previous trials. Congruently, each time an envelope is presented, the model attempts to retrieve one of the two responses associated with the envelope's color. For example, if there is a red envelope, the model attempts to retrieve the chunks "red X" and "red O." These two chunks enter a retrieval competition since only one of them can be retrieved at a time. The likelihood of retrieving a chunk depends on its activation relative to other competing chunks. The activation of a chunk is higher the more frequently and the more recently it has been used. Depending on its activation level, a chunk is probabilistically selected and determines the model's response. After the response, the model receives feedback whether it was right or wrong, reinforcing the chunk representing the correct answer.

Gaissmaier et al. (2006) focused on two parameters because they can be related to the two hypotheses (small sample hypothesis vs. predictive behavior hypothesis). One parameter, a decay parameter, affects the impact of recency on the activation of chunks. Without decay, each outcome would be weighed equally, irrespective of how long ago it has been observed. A model with high decay puts more weight on recent information and tends to disregard old information. Thus, the decay parameter offers a precise way to implement the small sample hypothesis proposed by Kareev (1995b; Kareev et al., 1997) in ACT-R. The higher the decay, the greater the impact of more recent trials, which amounts to paying attention to a small sample.

Another parameter, a noise parameter, affects how likely it is that the more activated chunk will actually be retrieved in competition with other chunks. Without noise, the most activated chunk will always be retrieved

(given that it is above a retrieval threshold). Given that the model assumes that the retrieval of a chunk determines the choice of a person (i.e., to choose X or O given a red or a green envelope), zero noise would result in perfect maximizing in the limit. A higher noise level allows less activated chunks to be retrieved from time to time. While such noise results in suboptimal behavior under some conditions, it is also used to model exploration (Taatgen, Lebiere, & Anderson, 2006). Thus, the noise parameter provides a simple way to model facets of predictive behavior, without developing a precise model of how people go about searching for patterns. In this regard, it is important not to interpret noise solely as error. Rather, higher levels of noise capture a proliferation of hypotheses that a participant may entertain, yielding behavior that looks like the model is searching for patterns in the data. This searching results in probability matching, whereas low levels of noise result in deterministic maximizing behavior. Gaissmaier et al. (2006) argued that the higher complexity of this behavior makes the relation to short-term memory plausible, supporting the interpretation that variation in this parameter nicely captures the predictive behavior hypothesis.

Two variants of the model, a decay and a noise variant, were fitted to the relative frequency of maximizing responses, that is, the average proportion choosing the maximizing answer, in Kareev et al.'s (1997) data. This was done separately for high and low spans as defined by Kareev et al. To do so, only the respective parameter (i.e., decay or noise) was varied in each of the model variants while keeping everything else constant.

To conclusively distinguish between the predictions made by the two different models, Gaissmaier et al.

(2006) used the models that were fitted to Kareev et al.'s (1997) data to make predictions about how high and low spans would adapt to a change in the correlational structure of the environment, henceforth called a *shift*. After the shift, the correlations were reversed. That is, if before the shift red was predictive of Xs and green was predictive of Os, this was reversed after the shift.

Both models were able to capture the low capacity advantage in correlation detection in a stable environment. As soon as the environment changed, however, a clear difference between the models emerged. If lower capacities result in simpler predictions (i.e., the predictive behavior hypothesis), then performance should be impaired if the environment changes. If, however, lower cognitive capacities indeed result in a more exaggerated perception of correlation (i.e., the small sample hypothesis), this should facilitate the detection of a change (Figure 3).

4.2 The low capacity advantage comes with a price in an unstable environment

Congruent with differences in the way participants make predictions, two experiments revealed a low capacity advantage before the environment changed, but a high capacity advantage afterwards. The low capacity advantage in this task comes with a price in an unstable environment. Figure 4 exemplifies this result by showing data from one of Gaissmaier et al.'s (2006) experiments[1]. This result demonstrates how important it is to consider the match between a strategy and the environment in which it operates: The presumably simpler, less explorative strategy by low spans allowed them to outperform high spans as long as the environment was stable. However, as soon as the environment changed, more explorative behavior paid off.

This also means that probability matching (or the more explorative behavior presumably underlying it) may not be as irrational as it initially appears. More explorative behavior could be a good habit to follow most of the time, because the cost of missing a non-random sequence could well be higher than the price of detecting patterns where there are none (Lopes, 1982). But explorative behavior fares poorly in stationary binary choice tasks. Choice tasks with stationary, constant probabilities, are rarely found outside of psychological laboratories and casinos (Ayton & Fischer, 2004). Gaissmaier et al. (2006) used random noise to model behavior they interpreted as systematic exploration. But even random noise can some-

---

[1]Note that another experiment revealed that this pattern of results only holds for men, while digit span capacity does not explain any variance in the behavior of women. This surprising finding could also be found in Kareev et al.'s (1997) data and is further discussed in Gaissmaier et al. (2006).
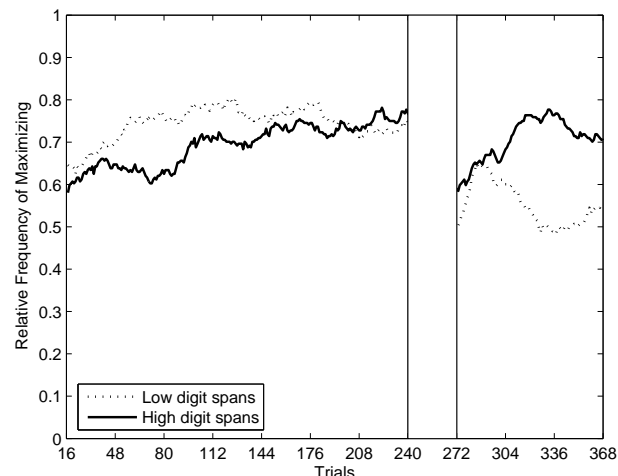

Figure 4: Maximizing on all trials, Experiment 1, late shift condition. Low and high digit spans were averaged separately across trials within a moving window of 32 trials. To prevent an overlap between trials before and after the shift in this window, I started averaging again after the shift, which is indicated by the two vertical lines at trials 240 and 272. That is, the last depicted data point before the shift consists of the last 32 trials before the shift, and the first depicted data point after the shift consists of the first 32 trials after the shift. (Reprinted with permission from Gaissmaier, Schooler, & Rieskamp, 2006.)

times be an effective way to escape local minima in optimization problems in a process called simulated annealing (Kirkpatrick, Gelatt, & Vecchi, 1983). Again, such a strategy is not good or bad per se, but only relative to a particular environmental structure. Gaissmaier et al. (2006) have shown how ACT-R can be used to make predictions about how cognitive limitations, decision behavior, and the environment interact, and how those predictions could then be used to disentangle different hypotheses experimentally.

5   General discussion

We have started with the premise that the mind is well adapted to the environment. In this regard, we are sympathetic to the idea that humans possess a repertoire of cognitive strategies, or heuristics, which can solve specific problems, captured by the metaphor of an adaptive toolbox (Gigerenzer et al., 1999). Those heuristics are called ecologically rational if they nestle into both the structure of the environment and the core capacities of the human mind.

We have dealt with issues that constitute the very core of ecological rationality: How exactly do the core capacities of the human mind on the one hand, and the structure

of the task environment on the other, shape the success of different cognitive strategies? More specifically, we have focused on making predictions about how cognitive limits affect decision making depending on the structure of the environment. To do so, we have illustrated how the adaptive toolbox approach could be combined with a unified, integrated theory of cognition, ACT-R (J. R. Anderson et al., 2004), by embedding different cognitive strategies from the adaptive toolbox within ACT-R.

5.1 Psychological plausibility

We have reviewed work by Schooler and Hertwig (2005) who showed that intermediate amounts of forgetting can be beneficial because it allows certain heuristics (such as the recognition and the fluency heuristic) to function well even after an organism has learned a lot about the environment. The forgetting parameter values they found to be particularly successful are very close to the default parameter value of forgetting which has been successfully used in a broad variety of tasks (see, e.g., J. R. Anderson & Lebiere, 1998). We have also reviewed work by Gaissmaier et al. (2006) in which different ideas of how cognitive limits could affect predictive behavior were implemented in ACT-R, leading to testable predictions that distinguish between the exploratory and small sample accounts of why those with limited short-term memory perform well in many probability learning tasks. Gaissmaier et al. found that simple predictions can be successful as long as the environment is stable, but they risk failing to detect changes in the environment. Similar to Schooler and Hertwig, Gaissmaier et al.'s model was constrained by the ACT-R architecture, and it also found parameter values comfortably in range of what is commonly used across many different tasks for the predictive behavior model that was supported by the data. In contrast, the best fitting parameters for the small sample hypothesis model settled on more extreme and atypical parameter values, signaling a problem with the model (see Gaissmaier et al., for details).

To be able to model behavior successfully within the constraints of the architecture is supportive of the psychological plausibility of the models developed by Schooler and Hertwig (2005) and by Gaissmaier et al. (2006). Psychological plausibility is an important dimension on which to evaluate cognitive models and such evaluations are facilitated by integrating cognitive strategies, such as heuristics, into a cognitive framework. If one only considered models in isolation, it would be impossible to judge the reasonableness of the parameter values or the processes and representations that the model depends on.

5.2 Relations between ACT-R and other models

As we have pointed out in the introduction, the applications of ACT-R presented here could have similarly been handled by other memory models, such as REM (Shiffrin & Steyvers, 1997), SAM (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981), or MINERVA-DM (Dougherty, Gettys, & Ogden, 1999). For instance, Schooler, Shiffrin & Raaijmakers (2001) developed REMI, a variant of REM designed to handle implicit memory effects in perceptual identification. One of the applications was to two alternative forced choices, where the focus was on the way in which noisy perceptual information is integrated with mnemonic information. The model was specified at such an abstract level that it could just as well have been applied to the problem of integrating cue-knowledge in forced choice tasks, such as those that the recognition heuristic is applicable to. Essentially, there was some chance, p(w), that a particular word would appear in the environment and this probability was taken into account when making perceptual judgments. Schooler et al. note that p(w) serves the same function in REMI that base level activation does in ACT-R. Though Schooler et al. only speculate that ACT-R could handle the range of implicit memory effects covered by REMI, Wagenmakers, Steyvers, Raaijmakers, Shiffrin, van Rijn & Zeelenberg's (2004) REM-LD modeled a challenging pattern of lexical decision data, which was also modeled in ACT-R by van Rijn & Anderson (2003). Given the apparent isomorphism between REM and ACT-R it seems likely that REM could well have been used to explore how forgetting aids heuristic inference.

There are undoubtedly advantages in working with simple models, such as REM. For instance, Schooler et al. were able derive closed form equations that fully described the behavior of the model, facilitating investigation of how it worked. Yet, the very simplicity of these models makes it difficult to know how to proceed when the task unfolds over several seconds or even minutes. ACT-R's model of memory is every bit as detailed, worked out, and tested as that of SAM, REM, or MINERVA-DM, but ACT-R's memory module is tightly integrated with theories of perception and motor control, strategy selection, and action. With ACT-R one has the choice of ignoring these complexities, as Schooler & Hertwig (2005) chose to do, but one can readily entertain how ACT-R's individual modules could work in concert in the service of more complex decision making activities. In short, ACT-R provides a general framework to investigate the relation between environment and mind, with the potential to be extended to a broad variety of topics and tasks, such as the issues of cognitive aging and strategy selection, which we illustrate in the following section.

5.3 Understanding the aging decision maker

We have suggested that the success of different strategies depends on both the structure of the environment and core cognitive abilities. However, the mind's core capacities change across the life span. Aging is associated with losses in working memory and the speed with which cognitive operations take place (Baltes, Staudinger, & Lindenberger, 1999). What is the impact of age-related cognitive decline on the adaptive toolbox? One promising avenue of research which may contribute to answering this question involves implementing strategies in ACT-R and assessing the role of aging by systematically varying parameters potentially related to age-related cognitive decline. For example, age-related decline in fluid abilities is associated with the use of simple inference strategies and may be related to strategy execution errors (Mata, Schooler, & Rieskamp, 2007). This is congruent with findings that people rely more on simple inference strategies when working memory load is high (Bröder & Gaissmaier, 2007). ACT-R parameters previously used to model working memory abilities are therefore candidate choices to model the increased reliance on simpler strategies and strategy execution deficits of increased age.

ACT-R simulations evaluating the role of age-related cognitive decline on strategy use in different environments could provide important insights into the conditions that lead people to fail or succeed as a result of aging. These results could in turn provide support for our ecological perspective on cognitive limitations. In particular, we believe the focus on mind-environment fit will lead to the conclusion that older adults' increased reliance on less cognitively demanding strategies may not always be a drawback, as these simpler strategies may fit well in specific environments.

5.4 Strategy selection

We believe that integrating the adaptive toolbox into an overarching framework such as ACT-R provides one possible answer to Alan Newell's (1973) warning that the only way to make progress in understanding human behavior can be made by developing unified theories of cognition (cf. Todd & Schooler, 2007). Such unified theories of cognition are not inconsistent with the metaphor of the adaptive toolbox, as we have illustrated by implementing various heuristics and decision strategies in ACT-R. Yet, we appreciate that there is the risk of a proliferation of tiny tools, one for each and every problem, which brings to the fore the problem of how people select among those tools (B. R. Newell, 2005). This will become increasingly important as more tools in the toolbox are proposed.

Although the important issue of strategy selection was not part of this paper, we believe that the ACT-R architecture would be a promising way to tackle it, and there are already some examples of what the approach might look like. Nellen (2003) implemented the Take-The-Best heuristic from the adaptive toolbox (Gigerenzer & Goldstein, 1996) in ACT-R. Via production learning, the model was able to adaptively select either the Take-The-Best heuristic or a competitor model (a weighted additive model), depending on which strategy was more successful in the environment (a similar approach, though not in the ACT-R framework, has been taken by Rieskamp & Otto, 2006; see also Rieskamp, this issue). Successful ACT-R models of strategy selection in the Tower of Hanoi (Fum & Del Missier, 2001) and an isomorph of the Water Jugs task (Lovett, 1998) portend the use of ACT-R to model the selection of heuristics from the adaptive tool box more generally.

5.5 Conclusion

In sum, we believe that working on specific cognitive strategies that are designed to solve particular problems, such as the tools in the adaptive toolbox, and to simultaneously try to integrate them into a unified framework such as ACT-R is not contradictory. To the contrary, we hope this article has shown the promise of this research strategy for the study of judgment and decision making. Perhaps, had the creators of the six million dollar man read our article, the preamble would have been. "Gentlemen, we can rebuild him. We have the technology. We have the capability to make the world's first bionic man. Steve Austin will be that man. The same as he was before. Forgetful, simple, myopic. However, he would only work well in some environments but not others." But perhaps such a nuanced story would not have glued the second author to the TV in his youth.

References

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science, 13*, 27–33.

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 and Language, 38,* 341–380.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1120–1136.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2,* 396–408.

Anderson, R. B., Doherty, M. E., Berg, N. D., & Friedrich, J. C. (2005). Sample size and the detection of correlation: A signal detection account. *Psychological Review, 112,* 268–279.

Ayton, P., & Fischer, I. (2004). The hot-hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition, 32,* 1369–1378.

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology, 50,* 471–507.

Bauer, M. (1972). Relations between prediction- and estimation-responses in cue-probability learning and transfer. *Scandinavian Journal of Psychology, 13,* 198–207.

Beilock, S.L., Bertenthal, B.I., McCoy, A.M., & Carr, T.H. (2004). Haste does not always make waste: Expertise, direction of attention and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review, 11,* 373–379.

Bjork, E. L., & Bjork, R. A. (1988) On the adaptive aspects of retrieval failure in autobiographical memory. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.). *Practical aspects of memory II* (pp. 283–288). London: Wiley.

Brackbill, N., & A. Bravos (1962). Supplementary report: The utility of correctly predicting infrequent events. *Journal of Experimental Psychology, 62,* 648–649.

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 multi-attribute decisions. *Psychonomic Bulletin and Review, 14,* 895–900.

Clément, M., Mercier, P., & Pasto, L. (2002). Sample size, confidence, and contingency judgement. *Canadian Journal of Experimental Psychology, 56,* 128–137.

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

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.

Fum, D., & Del Missier, F. (2001). Adaptive selection of problem solving strategies. In *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, 313–318.

Gaissmaier, W., Schooler, L. J., & Rieskamp, J. (2006). Simple predictions fueled by capacity limitations: When are they successful? *Journal of Experimental Psychology: Learning, Memory & Cognition, 32,* 966–982.

Gal, I., & Baron, J. (1996). Understanding repeated simple choices. *Thinking & Reasoning*, 81–98.

Gigerenzer, G. (2004). Fast and frugal heuristics: The tools of bounded rationality. In D. Koehler & N. Harvey (Eds.), *Handbook of judgement and decision making* (pp. 62–88). Oxford: Blackwell.

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

Gigerenzer, G., & Todd, P. M. and the ABC Research Group (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1–67.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109,* 75–90.

Gonzalez, C., Lerch, F. J., & Lebiere, C. (2003). Instance-based learning in real-time dynamic decision making. *Cognitive Science, 27,* 591–635.

Goodnow, J. J. (1955). Determinants of choice-distribution in two-choice situations. *American Journal of Psychology, 68*, 106–116.

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, UK: Wiley.

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

Juslin, P., & Olsson, H. (2005). Capacity limitations and the detection of correlation: Comment on Kareev (2000). *Psychological Review, 112,* 256–267.

Kareev, Y. (1995a). Positive bias in the perception of covariation. *Psychological Review, 102,* 490–502.

Kareev, Y. (1995b). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56,* 263–269.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review, 107,* 397–402.

Kareev, Y. (2004). On the perception of consistency. *Psychology of Learning and Motivation: Advances in Re-*

search and Theory, 44, 261–285.

Kareev, Y. (2005). And yet the small-sample effect does hold: Reply to Juslin and Olsson (2005) and Anderson, Doherty, Berg, and Friedrich (2005). *Psychological Review, 112,* 280–285.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General, 126,* 278–287.

Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica, 98*, 127–140.

Kirkpatrick, S., Gelatt Jr., C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science, 220,* 671–680.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Lopes, L. L. (1982). Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 626–636.

Lovett, M. C. (1998). Choice. In J. R. Anderson & C. Lebiere (Eds.), *The atomic components of thought* (pp. 255–296). Mahwah, NJ: Erlbaum.

Marewski, J. N., Gaissmaier, W., Schooler, L. J., Goldstein, D. G., & Gigerenzer, G. (2008). *Strategy selection by default: Recognition-based inference in federal and state elections.* Manuscript in preparation.

Marewski, J. N., & Schooler, L. J. (2008). *How memory aids strategy selection.* Manuscript in preparation.

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

Myers, J. L. (1976). Probability learning and sequence learning. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Approaches to human learning and motivation* (pp. 171–205). Hillsdale, NJ: Erlbaum.

Neimark, E. D., & Shuford, E. H. (1959). Comparison of predictions and estimations in a probability learning situation. *Journal of Experimental Psychology, 57,* 294–298.

Nellen, S. (2003). The use of the "take-the-best" heuristic under different conditions, modeled with ACT-R. In: F. Detje, D. Dörner, & H. Schaub (Eds.), *Proceedings of the fifth international conference on cognitive modeling* (pp. 171–176). Germany: Universitätsverlag Bamberg.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W.G. Chase (Ed.), *Visual Information Processing* (pp. 283–308). New York: Academic Press.

Newell, B. R. (2005). Re-visions of rationality. *Trends in Cognitive Sciences, 9*, 11–15.

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

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, and Cognition, 32,* 983–1002.

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

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

Reimer, T., & Katsikopoulos, K. V. (2004). The use of recognition in group decision-making. *Cognitive Science, 28,* 1009–1029.

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

Scheibehenne, B., & Bröder, A. (2007). Predicting Wimbledon tennis results 2005 by mere player name recognition. *International Journal of Forecasting, 23*, 415–426.

Schooler, L. J., & Anderson, J. R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology, 32*, 219–250.

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

Schooler, L., Shiffrin, R. M., & Raaijmakers, J. G. W. (2001) A model for implicit effects in perceptual identification. *Psychological Review, 108*, 257–272.

Serwe, S., & Frings, C. (2006). Who will win Wimbledon? The recognition heuristic in predicting sports

events. *Journal of Behavioral Decision Making , 19,* 321–332.

Shanks, D. R. (1985). Continuous monitoring of human contingency judgement across trials. *Memory & Cognition, 13,* 158–167.

Shanks, D. R. (1987). Acquisition functions in causality judgement. *Learning and Motivation, 18,* 147–166.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making, 15*, 233–250.

Shiffrin, R.M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving Effectively from Memory. *Psychonomic Bulletin & Review, 4*, 145–166.

Siegel, S., & Goldstein, D. A. J. (1959). Decision-making behavior in a two-choice uncertain outcome situation. *Journal of Experimental Psychology: General, 57*, 37–42.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say "broke"? A model of learning the past tense without feedback. *Cognition, 86,* 123–155.

Taatgen, N. A., Lebiere, C., & Anderson, J. R. (2006). Modeling paradigms in ACT-R. In R. Sun (Ed.), *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation* (pp. 29–52). Cambridge University Press.

Todd, P. M., & Schooler, L. J. (2007). From disintegrated architectures of cognition to an integrated heuristic toolbox. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 151–164). New York: Oxford University Press.

Todd, P.M., Hertwig, R., and Hoffrage, U. (2005). The evolutionary psychology of cognition. In D.M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 776–802). Hoboken, NJ: Wiley.

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

Unturbe, J., & Corominas, J. (2007). Probability matching involves rule-generating ability: A neuropsychological mechanism dealing with probabilities. *Neuropsychology, 21,* 621–630.

van Rijn, H., & Anderson, J. R. (2003). Modeling lexical decision as ordinary retrieval. In F. Detje, D. Doerner, & H. Schaub (Eds.), In *Proceedings of the Fifth International Conference on Cognitive Modeling* (pp. 207–212). Bamberg, Germany: Universitats-Verlag Bamberg.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys, 14,* 101–118.

Wagenmakers, E.-J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task. *Cognitive Psychology, 48*, 332–367.

West, R. F., & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition, 31,* 243–251.

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

Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language, 49*, 62–79.

Wolford, G., Miller, M. B., & Gazzaniga, M. (2000). The left hemisphere's role in hypothesis formation. *The Journal of Neuroscience, 20* (RC64), 1–4.

Wolford, G., Newman, S., Miller, M. B., & Wig, G. (2004). Searching for patterns in random sequences. *Canadian Journal of Experimental Psychology, 58,* 221–228.

Yellott, J. I., Jr. (1969). Probability learning with noncontingent success. *Journal of Mathematical Psychology, 6,* 541–575.