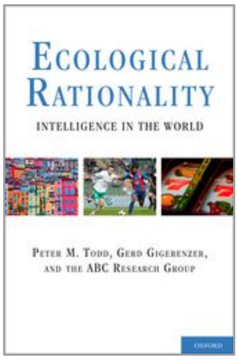

Ecological Rationality: Intelligence in the World
Peter M. Todd and Gerd Gigerenzer

Print publication date: 2012
Print ISBN-13: 9780195315448
Published to Oxford Scholarship Online: May 2012
DOI: 10.1093/acprof:oso/9780195315448.001.0001

How Smart Forgetting Helps Heuristic Inference

Lael J. Schooler
Ralph Hertwig
Stefan M. Herzog

DOI:10.1093/acprof:oso/9780195315448.003.0039

Abstract and Keywords

Theorists ranging from William James (1890) to some contemporary psychologists have argued that forgetting is the key to proper functioning of memory. The authors elaborate on the notion of beneficial forgetting by proposing that loss of information aids inference heuristics that exploit mnemonic information. They demonstrate this by implementing the recognition and fluency heuristics for two-alternative choice within the ACT-R cognitive architecture. For the recognition heuristic, forgetting can boost accuracy by increasing the chances that only a single alternative is recognized. Simulations of the fluency heuristic, choosing based on the speed with which the alternatives are recognized, indicate that forgetting aids the discrimination between recognition speeds. The authors show that retrieval fluency can be a proxy for real-world quantities, that people can discriminate between two objects' retrieval fluencies, and that people's inferences are in line with the fluency heuristic.

Keywords: fluency, fluency heuristic, recognition heuristic, ecological rationality, ACT-R, memory, forgetting, recognition

> "You see," he [Sherlock Holmes] explained, "I consider that a man's brain originally is like a little empty attic, and you have to stock it with such furniture as you choose. A fool takes in all the lumber of every sort that he comes across, so that the knowledge which might be useful to him gets crowded out, or at best is jumbled up with a lot of other things so that he has a difficulty in laying his hands upon it. Now the skilful workman is very careful indeed as to what he takes into his brain-attic. He will have nothing but the tools, which may help him in doing his work, but of these he has a large assortment, and all in the most perfect order. It is a mistake to think that that little room has elastic walls and can distend to any extent. Depend upon it—there comes a time when for every addition of knowledge you forget something that you knew before. It is of the highest importance, therefore, not to have useless facts elbowing out the useful ones."
>
> *Arthur Conan Doyle*

In *The Mind of a Mnemonist*, Luria (1968) examined one of the most virtuoso memories ever documented. The possessor of this memory—S. V. Shereshevskii, to whom Luria referred as S.—reacted to the discovery of his extraordinary powers by quitting his job as a reporter and becoming a professional mnemonist. S.'s nearly perfect memory appeared to have "no distinct limits" (p. 11). Once, for (p.145) instance, he memorized a long series of nonsense syllables that began "ma, va, na, sa, na, va, na, sa, na, ma, va" (Luria, 1968, p. 51). Eight years later, he recalled the whole series without making a single error or omission. This apparently infallible memory did not come without costs. S. complained, for example, that he had a poor memory for faces: "People's faces are constantly changing; it is the different shades of expression that confuse me and make it so hard to remember faces" (p. 64). "Unlike others, who tend to single out certain features by which to remember faces," Luria wrote, "S. saw faces as changing patterns…, much the same kind of impression a person would get, if he were sitting by a window watching the ebb and flow of the sea's waves" (p. 64). One way to interpret these observations is that cognitive processes such as generalizing, abstracting, and classifying different images of, for example, the same face require forgetting the differences between them. In other words, crossing the "'accursed' threshold to a higher level of thought" (Luria, 1968, p. 133), which in Luria's view S. never did, may require the ability to forget.

Is forgetting a nuisance and a handicap or is it essential to the proper functioning of memory and higher cognition? Much of the experimental research on memory has been dominated by questions of quantity, such as how much information is remembered and for how long (see Koriat, Goldsmith, & Pansky, 2000). From this perspective, forgetting is usually viewed as a regrettable loss of information. Some researchers have suggested, however, that forgetting may be functional. One of the first to explore this possibility was James (1890), who wrote, "In the practical use of our intellect, forgetting is as important a function as recollecting" (p. 679). In his view, forgetting is the mental mechanism behind the selectivity of information processing, which in turn is "the very keel on which our mental ship is built" (p. 680).

A century later, Bjork and Bjork (1988) argued that forgetting prevents out-of-date information—say, old phone numbers or where one parked the car yesterday—from interfering with the recall of currently relevant information. Altmann and Gray (2002) make a similar point for the short-term goals that govern our behavior; forgetting helps us to keep from retrieving the speed limit that was appropriate in town when we return to the freeway. From this perspective, forgetting prevents the retrieval of information that is likely obsolete. In fact, this is a function of forgetting that S. paradoxically had to do consciously. As a professional mnemonist, he committed thousands of words to memory. Learning to erase the images he associated with those words that he no longer needed to recall was an effortful, difficult process (Luria, 1968).

(p.146) How and why forgetting might be functional has also been the focus of an extensive analysis conducted by Anderson and colleagues (Anderson & Milson, 1989; Anderson & Schooler, 1991, 2000; Schooler & Anderson, 1997). On the basis of their rational analysis of memory, they argued that much of memory performance, including forgetting, might be understood in terms of adaptation to the structure of the environment. The rational analysis of memory assumes that the memory system acts on the expectation that environmental stimuli tend to reoccur in predictable ways. For instance, the more recently a stimulus has been encountered, the higher the expectation that it will be encountered again and information about that stimulus will be needed. Conversely, the longer it has been since the stimulus was encountered, the less likely it is to be needed soon, and so it can be forgotten.

A simple time-saving feature found in many word processors can help illustrate how recency can be used to predict the need for information. When a user prepares to open a document file, some programs present 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 either remembering in which folder the file is located or searching through folder after folder. For this mechanism 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. The word processor uses the heuristic that the more recently a file has been opened, the more likely it is to be needed again now. In the rest of this chapter, we show how human memory bets on the same environmental regularity, and how this bet can enable simple heuristics, including the recognition and fluency heuristics, to operate effectively.

Forgetting: The Retention Curve

The rational analysis of memory rests on the assumption that environmental stimuli make informational demands on the cognitive system that are met by retrieving memory traces associated with the stimuli. Consequently, memory performance should reflect the patterns with which environmental stimuli appear and reappear in the environment. An implication is that statistical regularities in the environment can be used to make predictions about behavior, say, performance in memory experiments. Conversely, performance on memory tasks can provide predictions about the environment. One such prediction follows from the retention function, an iconic manifestation of the regularity behind forgetting in human memory. This function is studied by exposing people to an item and then (p.147) testing performance at various lags, known as retention intervals. Squire (1989), for example, presented people with the names of real and made-up TV shows. They had to decide whether the names were of real shows. Figure 6-1 plots people's recognition performance as a function of the number of years since the show's cancellation. The more time has passed since a TV show was cancelled, the lower the memory for that show.

From the perspective of the rational analysis of memory, performance falls as a function of retention interval because memory performance reflects the probability of encountering a particular environmental stimulus (e.g., a name), which in turn falls as a power function of how long it has been since the stimulus was last encountered. For instance, the probability that you will encounter the TV show name "The Mary Tyler Moore Show," a hit in the 1970s, should currently be much lower than the probability that you will encounter the name "Grey's Anatomy," a top-rated show as we write this chapter. Anderson and Schooler (1991) tested the link between memory performance and environmental regularities in environments that place informational demands on people (see also Anderson & Schooler, 2000; Schooler & Anderson, 1997). One such environment involves the daily distribution of people who sent electronic mail messages, capturing aspects of a social environment. Another environment, linguistic in nature, involves word usage in speech to children. A third environment is that of *New York Times* headlines. Figure 6-2 shows the probability of a word (p.148) occurring in the headlines as a function of the number of days since that word had previously occurred.[1] Just as memory performance falls as a function of retention interval, so too does the probability of a word appearing—that is, it falls as a function of the time since it was last mentioned. Consistent with Anderson and Schooler's predictions, the memory retention function reflects statistical regularities in the world, and vice versa. The rational analysis of memory framework accounts for a variety of memory phenomena (see Anderson & Schooler, 2000, for a review), including spacing effects, to which we turn now.

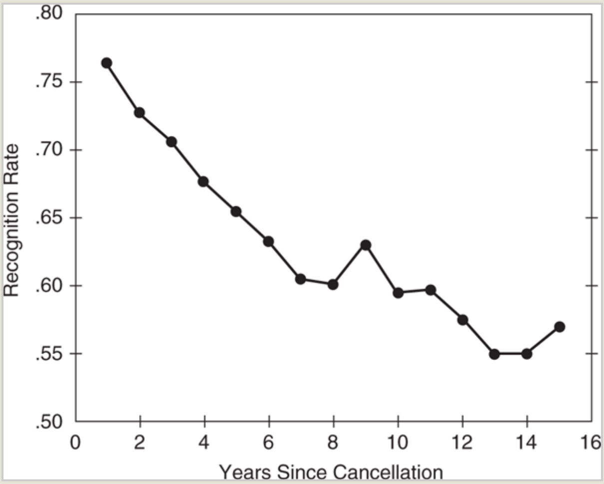

Figure 6-1: Mean recognition rates of television shows as a function of years since the show was canceled (data from Squire, 1989).

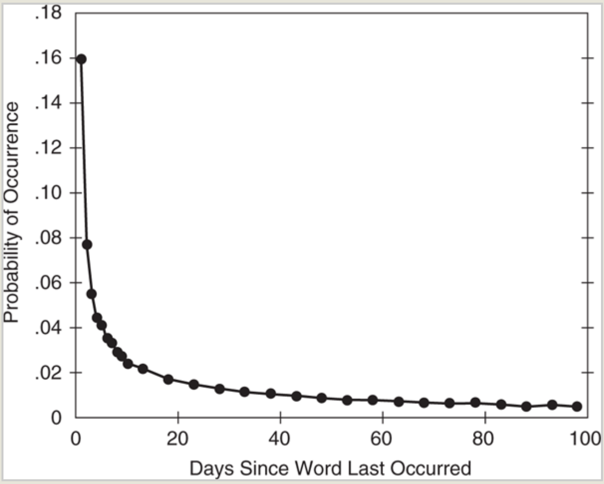

Figure 6-2: Probability of a word being used in *New York Times* headlines as a function of number of days since it was last used (data from Anderson & Schooler, 1991).

Spacing Effects in Memory

Nearly all laboratory memory experiments involve the presentation of material to participants that must be retrieved later. When material is presented multiple times, the lag between these presentations is known as spacing, and the lag between the final presentation and test is again called the retention interval. The spacing effect (p.149) involves the interaction of the spacing between presentations and the retention interval. For verbal material, one tends to observe that at short retention intervals performance is better for tightly massed presentations (i.e., separated by short intervals), but at longer retention intervals performance is better for widely spaced presentations. Consider two groups of students preparing for a foreign language vocabulary test. What is the most efficient use of the limited time they have? The cramming students would do all of their studying on the Wednesday and Thursday before the exam on Friday. The conscientious students would study a little each week, say, the Thursday in the week preceding the exam and again on the Thursday before the Friday exam. The stylized result is that the cramming students, whose study spacing matched the one-day retention interval, would do better on the Friday exam than the conscientious ones. This would seem to vindicate all those procrastinators in college who put off studying for their exams until the last minute.

But there is a catch. If the material were tested again later, say, in a pop quiz on the following Friday, the conscientious students would outperform the crammers. That is, the forgetting rate for material learned in a massed way is faster than for material learned in a more distributed fashion. Plotting the performance of the two groups of students on the two Fridays would be expected to reveal the crossover interaction typically found in experiments that manipulate study spacing and retention lag. The results from one such experiment are graphed in Figure 6-3, illustrating this interaction at timescales of days. Participants in Keppel (1967) studied pairs of words a total of eight times. People in the massed condition studied the material eight times in 1 day, while those in the distributed condition studied the material twice on each of 4 days. Immediately after studying the material, people in the massed condition performed best, but after 8 days those exposed to distributed presentations performed best.

Spacing Effects in the Environment

What pattern in the environment would correspond to spacing effects in memory performance? Figure 6-4 shows the spacing analysis from Anderson and Schooler (1991), which was restricted to those words in *New York Times* headlines that occurred exactly twice in a 100-day window. For purposes of illustration, consider the uppermost point that corresponds to a word that, say, was mentioned on January 26 and then again on January 31. The *y*-axis plots the chances (probability) that it would be mentioned yet again on, say, February 5. The other labeled point represents words that were mentioned on, say, October 1 and not again until December 1, (p.150)

(p.151)

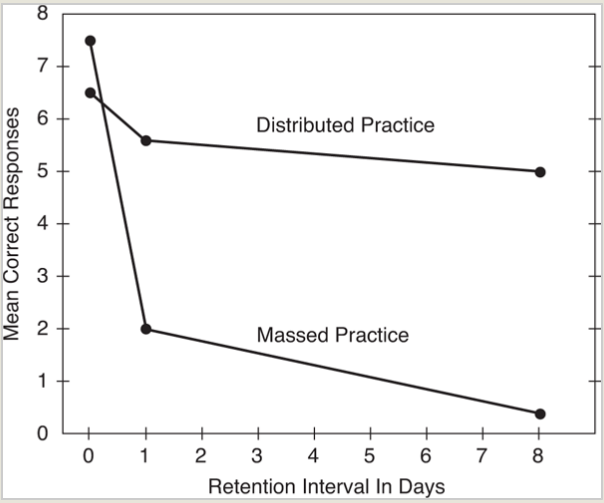

Figure 6-3: Memory performance as a function of whether learning followed a massed or distributed practice regimen (data from Keppel, 1967).

but with the interval from the last mention to February 5 now being 66 days. One way to characterize the results in Figure 6-4 is that when words are encountered in a massed way there is an immediate burst in the likelihood of encountering them again, but that this likelihood drops precipitously. In contrast, words encountered in a more distributed fashion do not show this burst, but their likelihood of being encountered in the future remains relatively constant. The difference is akin to that between the patterns with which one needs a PIN (personal identification number) for the safe in a hotel room and the PIN for one's bank account. While on vacation, one will frequently need the safe's PIN, but over an extended period one is more likely to need the PIN for the bank account. The idea is that the memory system figures the relative values of the codes over the short and long run, based on the pattern with which they are retrieved. So one can think about cramming for an exam as an attempt to signal to the memory system that the exam material will likely be highly relevant in the short term, but not so useful further in the future.

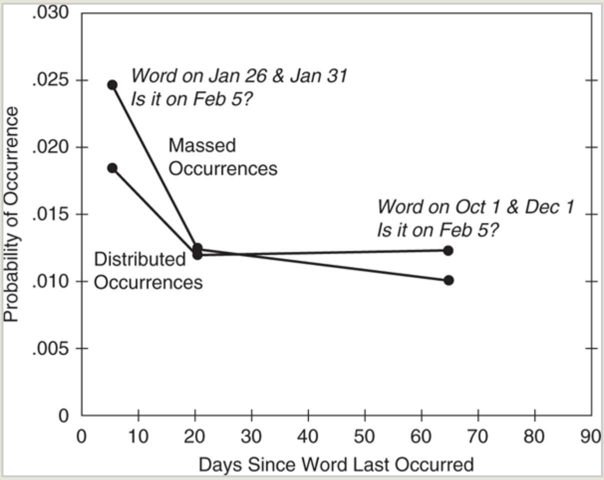

Figure 6-4: Probability of a word being used in the *New York Times* headlines as a function of number of days since it was last used, given that the word was used just twice in the previous 100 days. The steeper curve shows words whose two uses in the headlines were massed near in time to each other, and the shallower curve shows words whose occurrences were distributed farther apart (data from Anderson & Schooler, 1991).

These isomorphisms between regularities in memory and in the statistical structure of environmental events exemplify the thesis that human memory uses the recency, frequency, and spacing with which information has been needed in the past to estimate how likely that information is to be needed now. Because processing unnecessary information is cognitively costly, a memory system able to prune away little-needed information by forgetting it is better off. In what follows, we extend the analysis of the effects of forgetting on memory performance to its effects on the performance of simple inference heuristics. To this end, we draw on the research program on fast and frugal heuristics (Gigerenzer, Todd, & the ABC Research Group, 1999) and the ACT-R research program (Adaptive Control of Thought–Rational—see Anderson & Lebiere, 1998). The two programs share a strong ecological emphasis.

The research program on fast and frugal heuristics examines simple strategies that exploit informational structures in the environment, enabling the mind to make surprisingly accurate decisions without much information or computation. The ACT-R research program also strives to develop a coherent theory of cognition, specified to such a degree that phenomena from perceptual search to the learning of algebra might be modeled within the same framework. In particular, ACT-R offers a plausible model of memory that is tuned, according to the prescriptions of the rational analysis of memory, to the statistical structure of environmental events. This model of memory will be central to our implementation of the *recognition heuristic* (Goldstein & Gigerenzer, 2002) and the *fluency heuristic* (Hertwig, Herzog, Schooler, & Reimer, 2008), both of which depend on phenomenological assessments of (p.152) memory retrieval. The former operates on knowledge about whether a stimulus can be recognized, whereas the latter relies on an assessment of the fluency, or speed, with which a stimulus is processed. By housing these memory-based heuristics in a common cognitive architecture, we aim to provide models that allow us to analyze whether and how loss of information—that is, forgetting—fosters the performance of these heuristics. We begin by first describing the recognition heuristic, the fluency heuristic, and the ACT-R architecture; then we turn to the question of whether the recognition and the fluency heuristic benefit from smart forgetting.

How Recognition Enables Heuristic Inference: The Recognition Heuristic

The recognition heuristic illustrates the interplay between the structure of the environment and core capacities of the human mind (Goldstein & Gigerenzer, 2002; see chapter 5 for a detailed discussion). In short, the recognition heuristic uses the information about whether objects are recognized or not to make inferences about their values on some quantitative criterion dimension. Its policy goes like this:

> 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 of interest. (Goldstein & Gigerenzer, 2002, p. 76)

To successfully apply the recognition heuristic, the probability of recognizing objects needs to be correlated with the criterion to be inferred. This is the case, for example, 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). One reason why objects with larger criterion values are more often recognized is that they are more often mentioned in the environment (see chapter 5).

To be applied, the recognition heuristic requires that a person does not recognize too much or too little: One of the alternatives needs to be recognized, but not the other. If a person recognizes too few or too many objects, then recognition will be uninformative because it will rarely discriminate between the objects. Consider a die-hard fan of the National Basketball Association who will not be able to use the recognition heuristic to predict the outcome of any game, simply because she recognizes all of the teams. In contrast, an occasional observer of basketball games may recognize some but not all teams, and thus can more often use the recognition heuristic. The fact that the recognition heuristic feeds on partial (p.153) ignorance implies the possibility that forgetting may boost this heuristic's performance. Before we investigate this odd possibility, let us consider what a person does who recognizes all the teams. In this case, more knowledge-intensive strategies, such as the take-the-best heuristic, can be recruited (Gigerenzer et al., 1999). Take-the-best sequentially searches for cues that are correlated with the criterion in the order of their predictive accuracy and chooses between the objects on the basis of the first cue found that discriminates between them (Gigerenzer & Goldstein, 1996). But there is a potentially faster alternative to this knowledge-based strategy—namely, the fluency heuristic.

How Retrieval Fluency Enables Heuristic Inference: The Fluency Heuristic

When two objects to be decided between are both recognized, the fluency heuristic (see, e.g., Jacoby & Brooks, 1984; Toth & Daniels, 2002; Whittlesea, 1993) can be applied. It can be expressed as follows:

> Fluency heuristic: If one of two objects is more fluently processed, then infer that this object has the higher value with respect to the criterion of interest.

Like the recognition heuristic, the fluency heuristic considers only a single feature of the objects: the fluency with which the objects are processed when encountered. In numerous studies, this processing fluency, mediated by prior experience with a stimulus, has been shown to function as a cue in a range of judgments. For example, more fluent processing due to previous exposure can increase the perceived fame of nonfamous names (the *false fame effect*; Jacoby, Kelley, Brown, & Jasechko, 1989) and the perceived truth of repeated assertions (the *reiteration effect*; Begg, Anas, & Farinacci, 1992; Hertwig, Gigerenzer, & Hoffrage, 1997).

In the literature, one can find many different variants of fluency, including *absolute*, *relative*, *conceptual*, and *perceptual* fluency, to name a few. Fluency has also been invoked in explaining a wide range of judgments, including evaluative and aesthetic judgments (e.g., Winkielman & Cacioppo, 2001; see Reber, Schwarz, & Winkielman, 2004, and Winkielman, Schwarz, Fazendeiro, & Reber, 2003, for reviews), and confidence and metacognitive judgments (e.g., Kelley & Lindsay, 1993; Koriat & Ma'ayan, 2005). One can also, although less frequently, come across the notion of a fluency heuristic, prominently in the work of Kelley and Jacoby (1998), Whittlesea (1993), and Whittlesea and Leboe (2003). Abstracting from the different meanings of the term fluency heuristic across articles, the gist appears to be that people attribute the fluent (p.154) processing of stimuli to having experienced the stimuli before. The ACT-R fluency heuristic, as proposed by Schooler and Hertwig (2005; see also Hertwig et al., 2008; Marewski & Schooler, 2011), aims to exploit the subjective sense of fluency in the process of making inferences about objective properties of the world.

The fluency heuristic, in contrast to the recognition heuristic, does not exploit partial ignorance but rather graded recognition. Nevertheless, it may also benefit from forgetting because fluency is more easily applicable if there are large detectable differences in fluency between objects—and forgetting could create such differences. To investigate the role of forgetting in memory-based heuristics and to model the relation between environmental exposure and the information in memory on which heuristics such as recognition and fluency feed, we implement them within the ACT-R architecture, which we now describe.

A Brief Overview of ACT-R

ACT-R is a theory of cognition constrained by having to account for a broad swath of human thought. The core of ACT-R is constituted by a declarative memory system for facts (*knowing that*) and a procedural system for rules (*knowing how*). The declarative memory system consists of records that represent information (e.g., facts about the outside world, about oneself, about possible actions). These records take on activations that determine their accessibility, that is, whether and how quickly they can be retrieved. A record's activation A_i is determined by a combination of the base-level strength of the record, B_i, and the S_{ji} units of activation it receives from each of the j elements of the current context:

$$A_i = B_i + \sum_j S_{ij}$$

A record’s base-level strength is rooted in its environmental pattern of occurrence. The activation of a record is higher the more frequently and the more recently it has been used; activation strengthens with use and decays with time. Specifically, B_i is determined by how frequently and recently the record has been encountered in the past (e.g., studied) and can be stated as follows:

$$B_i = \ln\left(\sum_{k=1}^{n} t_k{}^{d}\right),$$

where the record has been encountered n times in the past at lags of t_1, t_2,…,t_n. Finally, d is a decay parameter that captures the amount of forgetting in declarative memory and thus determines how much (p.155) information about an item’s environmental frequency is retained in memory over time, as reflected in the corresponding record’s activation. Typically, d is set to –0.5, which has been used to fit a wide range of behavioral data (Anderson & Lebiere, 1998).

The procedural system consists of if–then rules that guide the course of action an individual takes when performing a specific task. The *if* side of a production rule specifies various conditions, which can include the state of working memory, changes in perceptual information such as detecting that a new object has appeared, and many other inputs. If all the conditions of a production rule are met, then the rule fires, and the actions specified in the *then* side of the rule are carried out. These actions can include updating records, creating new records, setting goals, and initiating motor responses. This combination of components makes ACT-R a good framework within which to implement decision-making strategies, in cognitively plausible ways (Todd & Schooler, 2007).

Do the Recognition and Fluency Heuristics Benefit From Smart Forgetting?

Bettman, Johnson, and Payne (1990) explored the relative cognitive complexity and effort that various decision strategies require by representing them in production rules consisting of simple cognitive steps, such as *read, add*, and *compare*. They termed them elementary information processes. Building on this work, we show how implementing the recognition and fluency heuristics in ACT-R enables us to explore how properties of the cognitive system, such as forgetting, affect the heuristics' performance in specific environments. According to Goldstein and Gigerenzer (2002), the recognition heuristic works because there is a chain of correlations linking the criterion (e.g., the strength of an NBA basketball team), via environmental frequencies (e.g., how often the team is mentioned in the media), to recognition. ACT-R's activation tracks just such environmental regularities, so that activation differences reflect, in part, frequency differences. Thus, it would be possible in principle that inferences—such as deciding which of two players is better or which of two cities is larger—could be based directly on the activation of associated records in memory (e.g., player or city representations). However, this possibility is inconsistent with the ACT-R framework for reasons of psychological plausibility: Subsymbolic quantities, such as activation, are assumed not to be directly accessible, just as people presumably cannot make decisions by directly observing differences in their own neural firing rates. Instead, though, the system could capitalize on activation differences associated with various objects by gauging how it responds to them. The simplest measure of the system's response is (p.156) whether a record associated with a specific object can be retrieved at all, and we use this to implement the recognition heuristic in ACT-R.

First, our model learned about large German cities based on artificial environments that reflected how frequently the cities were mentioned in an American newspaper (see Schooler & Hertwig, 2005, for details). Second, recognition rates for the model were calibrated against the empirical recognition rates that Goldstein and Gigerenzer (2002) observed. In accordance with previous models of recognition in ACT-R (Anderson, Bothell, Lebiere, & Matessa, 1998), recognizing a city was considered to be equivalent to retrieving the record associated with it. Third, the model was tested on pairs of German cities. The model's recognition rates from the second step defined the probability that it would successfully recognize a city. 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. Such a decision rule closely matched the observed human responses. In all other cases (both cities recognized or unrecognized), the model made a guess. With this model in hand, we can ask whether forgetting can boost the accuracy of the memory-based inferences made by the recognition heuristic.

Does Forgetting Benefit the Recognition Heuristic?

To address this question, we varied the decay rate *d* and observed how the resulting changes in recognition affect inferences in the city population task. The upper bound of the decay rate, 0, means no forgetting, so that the strength of a memory record is strictly a function of its frequency. Negative values of *d* imply forgetting, and more negative values imply more rapid forgetting. Using a step size of 0.01, we tested *d* values ranging from 0 to −1, the latter being twice ACT-R's default decay rate. In Figure 6-5, the solid line shows the recognition heuristic's average level of accuracy on pairwise comparisons of all German cities it knew, including pairs in which it had to guess because both cities were recognized or unrecognized. Three aspects of this function are noteworthy. First, the recognition heuristic's performance assuming no forgetting (56% correct) is substantially worse than its performance assuming the "optimal" amount of forgetting (63.3% correct). Second, ACT-R's default decay value of –0.5 yields 61.3% correct, only slightly below the peak performance level, which is reached at a decay rate of –0.34. Third, the accuracy curve has a flat maximum, with all decay values from –0.13 to –0.56 yielding performance in excess of 60% correct. These results demonstrate that forgetting enhances the performance of the recognition heuristic, and the amount of (p.157) forgetting can vary over a substantial range without compromising the heuristic's good performance. However, as *d* approaches −1 and there is too much forgetting (resulting in a situation in which most cities are unrecognized), the performance of the recognition heuristic eventually approaches chance level.

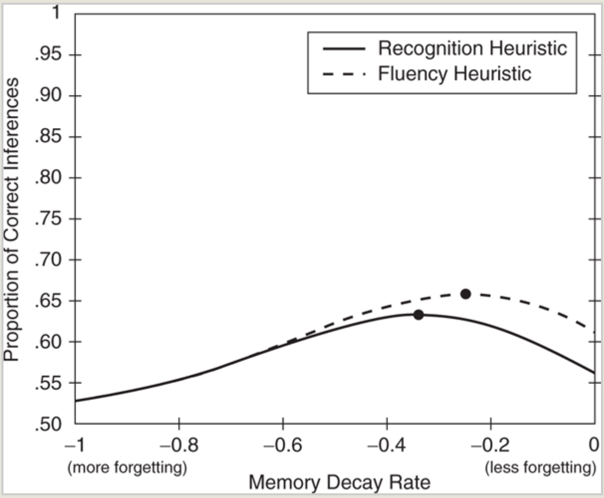

Figure 6-5: Performance of the recognition and fluency heuristics as a function of memory decay rate, *d*. Maxima are marked with dots. (Adapted from Schooler & Hertwig, 2005.)

How Does Forgetting Help the Recognition Heuristic's Performance?

Two quantities shed more light on the link between forgetting and the recognition heuristic. The first is the proportion of comparisons in which the recognition heuristic can be used as the basis for making a choice, that is, the proportion of comparisons in which only one of the cities is recognized. In Figure 6-6, the solid line shows that for the recognition heuristic this *application rate* peaks when *d* equals –0.28, an intermediate level of forgetting. The second quantity is the proportion of correct inferences made by the recognition heuristic in those choices to which it is applicable. As shown in Figure 6-7, this *recognition validity* generally increases with the amount of forgetting, peaking when *d* equals −1. The performance (Figure 6-5) and application rate (Figure 6-6) peak at nearly the same forgetting rates of −0.34 and −0.28, compared to the peak of −1 for the validity curve (Figure 6-7).

So, the decay rate of −0.34 can be thought of as the best trade-off between the effects (p.158)

(p.159)

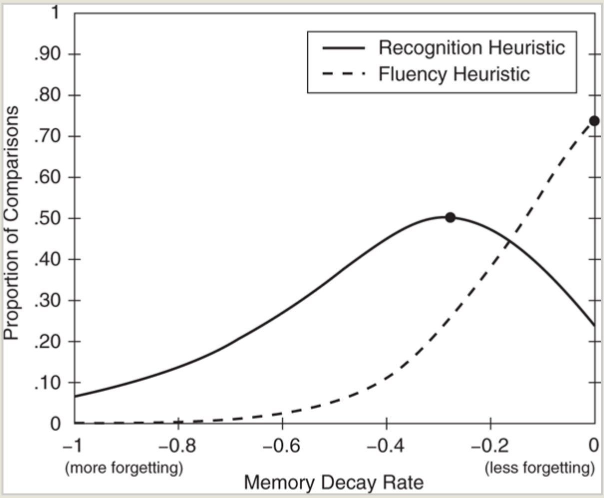

Figure 6-6: The application rate of the recognition heuristic (the proportion of all comparisons in which one city is recognized but the other is not) and of the fluency heuristic (the proportion of all comparisons in which both cities are recognized), as a function of memory decay rate, *d*. Maxima are marked with dots. (Adapted from Schooler & Hertwig, 2005.)

of forgetting on application rate and validity, with the application rate having the greater sway over performance. Thus, intermediate amounts of forgetting increase the performance of the recognition heuristic mostly by sharply increasing its applicability and, to a lesser extent, by increasing its validity.

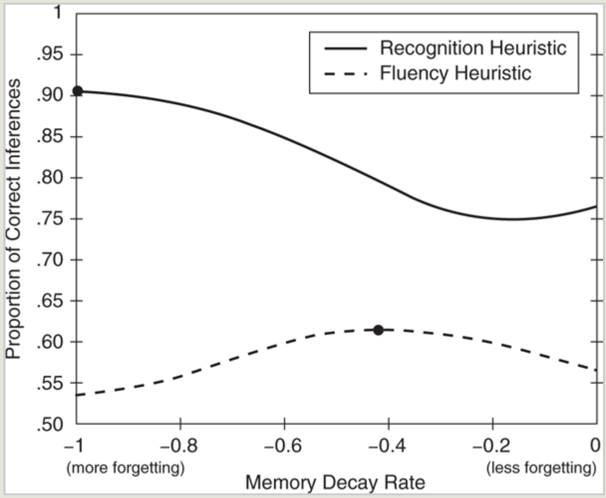

Figure 6-7: The validity of the recognition heuristic and of the fluency heuristic (the proportion of correct inferences that each heuristic makes when it can be applied) as a function of memory decay rate, *d*. Maxima are marked with dots. (Adapted from Schooler & Hertwig, 2005.)

Does Forgetting Help the Fluency Heuristic?

Loss of some information—a loss that is not random but a function of a record's environmental history—fosters the performance of the recognition heuristic. But is this benefit of forgetting limited to the recognition heuristic? To find out whether an inference strategy that makes finer distinctions than that between recognition and nonrecognition can benefit from forgetting, we now turn to 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 unrecognized (and not retrieved). But this heuristic essentially passes up information (for better or worse) whenever two objects are both recognized but the record associated with one has a higher activation than the other. The recognition heuristic ignores this difference in activation. But could this activation difference be used to decide between the two objects? Within ACT-R, recognition could also be assessed in a continuous fashion, namely, in terms of how quickly an object's record can be retrieved. Differences in retrieval time are a proxy of differences in the subsymbolic quantity of activation. The fluency heuristic exploits differences in retrieval time by inferring that if one of two objects is more swiftly retrieved, this object has the higher value with respect to the criterion.

The predictive accuracy 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 (dashed) line in Figure 6-5. At the same time, the fluency heuristic provides an overall additional gain in performance above the recognition heuristic. Figure 6-6 (dashed line) shows that the applicability of the fluency heuristic does not benefit from forgetting but rather decreases as forgetting increases. Part of the explanation for how the fluency heuristic does benefit from forgetting is illustrated in Figure 6-8, which shows the exponential function that relates a record's activation to its retrieval time. To appreciate the explanation, let us first point out that neither our ACT-R model of the fluency heuristic nor actual people can reliably discriminate between any minute difference in two retrieval times. In fact, the difference in retrieval times needs to be at least 100 ms for people to be able to reliably discriminate between them (Hertwig et al., 2008). The beneficial impact of forgetting on (p.160) the fluency heuristic is related to this *just noticeable difference* (JND). Specifically, forgetting lowers the range of activations to levels that correspond to retrieval times that can be more easily discriminated. For illustration, consider retrieval times of 200 and 300 ms, which correspond to activations of 1.99 and 1.59, respectively. For these relatively low activations, only a small difference of 0.4 units of activation suffices to yield the 100 ms JND in retrieval time. In contrast, the same 100 ms difference in retrieval time between 50 and 150 ms corresponds to a difference of 1.1 units of activation. Thus, by shifting the activation range downward, forgetting helps the system settle on activation levels corresponding to retrieval times that can be more easily discriminated. In other words, a given difference in activation at a lower range results in a larger, more easily detected difference in retrieval time than the same difference at a higher range. In the case of the fluency heuristic, memory decay prevents the activation of (retrievable) records from becoming saturated.

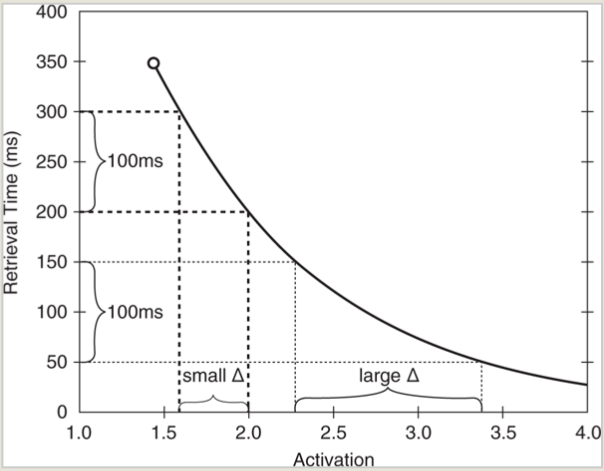

Figure 6-8: The relationship between a memory record's activation and its retrieval time. (Adapted from Schooler & Hertwig, 2005.)

Both the recognition and the fluency heuristic can be understood as means to indirectly tap the environmental frequency information locked in the activations of records 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 exploring the sensitivity of the recognition and fluency heuristics to changes in the rate (p.161) of memory decay within ACT-R, we demonstrated that forgetting actually serves to improve the performance of these heuristics by strengthening the middle links of the chain of correlations on which they rely.

Do People Use the Fluency Heuristic?

Up to this point, our analysis of the fluency heuristic has been mostly theoretical in nature. Is there empirical evidence that the retrieval fluency is a valid indicator of environmental quantities, and that fluency guides people's inferences about those quantities? To find out, we performed ecological and empirical analyses of fluency. We first analyzed the validity of the fluency heuristic in five real-world environments by measuring actual retrieval fluency (as recognition speeds) and using a quantitative criterion (Hertwig et al., 2008, Study 1): (a) the 118 U.S. cities with more than 100,000 inhabitants in 2002; (b) the 100 German companies with the highest revenue in 2003; (c) the 106 most successful music artists in the United States, in terms of the cumulative U.S. sales of recordings from 1958 to 2003; (d) the 50 richest athletes in the world in 2004; and (e) the 100 wealthiest people in the world in 2004. The validity of retrieval fluency in each environment was defined as the mean proportion of pairs where the object with the smaller mean retrieval time scored higher on the respective criterion (averaged across 40 participants, excluding pairs where the difference in mean retrieval times was below the JND of 100 ms). In all five environments, fluency validity exceeded chance level (.50), ranging from .66 in the cities environment to .58 in the companies and music artists environments. In addition, fluency validity was related to the size of the differences in mean retrieval time. Figure 6-9 shows that there is a clear tendency, manifest across all five environments, for larger differences to be associated with higher fluency validity. This tendency can also be explained within the ACT-R framework: Objects with larger criterion values tend to occur more frequently in the environment, and thus their memory records tend to have higher activations and be more quickly retrieved. Consequently, large differences in retrieval times are likely to correspond to pairs of objects in which one object has a large criterion value and the other has a small value. For such pairs, fluency can be expected to be quite valid. In an extensive ecological analysis of fluency, we replicated and extended these results across more than 20 diverse domains (Herzog & Hertwig, in press).

Thus, using fluency could lead to valid decisions—but to what extent do people's inferences actually agree with its use in the fluency heuristic? Across three of the five environments listed (p.162) above, cities, companies, and music artists, we asked participants to infer which of two objects scored higher on a quantitative dimension (Hertwig et al., 2008, Study 3). In addition, participants' retrieval times for objects in these environments were measured. Then, for each participant, the percentage of inferences that were in line with the fluency heuristic (among all pairs in which both objects were recognized) was determined. The mean accordance with the fluency heuristic was .74, .63, and .68 in the cities, companies, and music artists environments, respectively. The extent to which people's inferences conformed to the fluency heuristic was a function of differences in recognition speeds, as shown in Figure 6-10, even rising to around .8 accordance when these differences exceeded 700 ms in the cities and music artists environments. This appears to be ecologically rational use of the fluency heuristic, insofar as retrieval fluency is more likely to yield accurate inferences with larger differences in retrieval times (Figure 6-9).

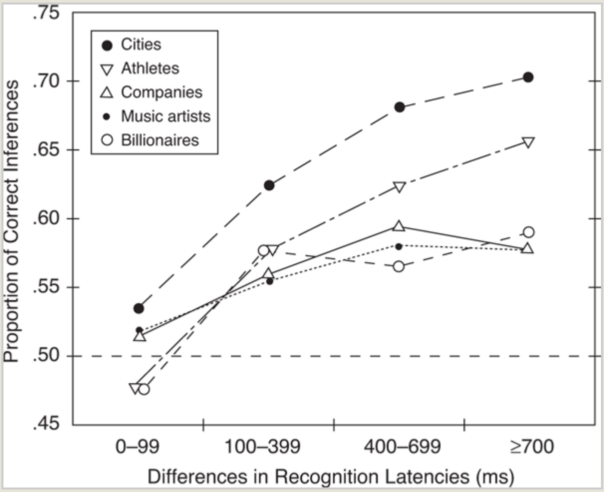

Figure 6-9: The validity of the fluency heuristic (the proportion of correct inferences that the rule makes when it can be applied) as a function of increasing differences in recognition latencies. (Adapted from Hertwig et al., 2008.)

To summarize, retrieval fluency can be a valid predictor of objective properties of the world, and to different degrees in different environments. Moreover, we found that a large proportion of people's inferences conformed to the decisions made by the fluency heuristic using this predictor. In a related analysis, Marewski and Schooler (2011) showed that the use of the fluency heuristic appears (p.163) particularly pronounced when people recognize both objects but cannot retrieve any additional cue knowledge about them.

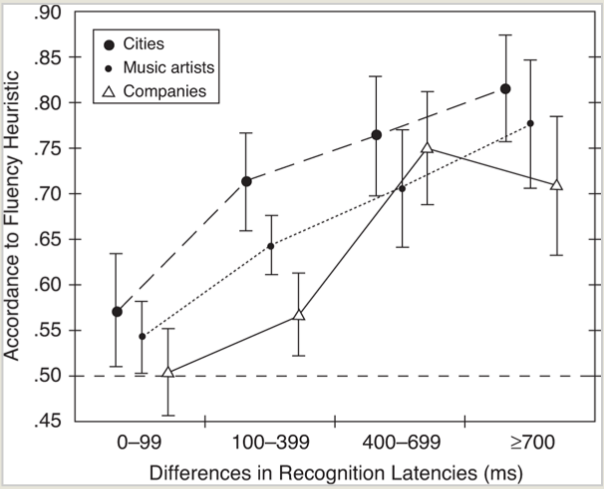

Figure 6-10: Proportion of decisions made in accordance with the fluency heuristic as a function of increasing differences in recognition latencies (bars show 95% confidence intervals of proportions aggregated across subjects). (Adapted from Hertwig et al., 2008.)

The Importance of Forgetting

Some theorists have argued that forgetting is indispensable to the proper working of memory. Building on the notion of beneficial forgetting, we demonstrated that ecologically smart loss of information—loss that is not random but reflects the environmental history of the memory record—may not only foster memory retrieval processes but may also boost the performance of inferential heuristics that exploit mnemonic information such as recognition and retrieval fluency. If human recognition memory were so lossless and exquisitely sensitive to novelty that it treated as unrecognized only those objects and events that one has truly never seen before (and not also those that were experienced long ago and since forgotten), then extensive experience could eventually render the recognition heuristic inapplicable (see Todd & Kirby, 2001). By implementing inferential heuristics within an existing cognitive architecture, we were able to analyze in detail how parameters of memory such as information decay affect inferential accuracy.

(p.164) This analysis also revealed two distinct reasons for why forgetting and heuristics can work in tandem. In the case of the recognition heuristic, intermediate amounts of forgetting maintain the systematic partial ignorance on which the heuristic relies, increasing the probability that it correctly picks the higher criterion object. In the case of the fluency heuristic, intermediate amounts of forgetting boost the heuristic’s performance by maintaining activation levels corresponding to retrieval latencies that can be more easily discriminated. In what follows, we discuss how the fluency heuristic relates to the availability heuristic and whether it is worthwhile to maintain the distinction between the fluency and recognition heuristics, and we conclude by examining whether forgetting plausibly could have evolved to serve heuristic inference.

The Fluency and Availability Heuristics: Old Wine in a New Bottle?

The fluency heuristic feeds on environmental frequencies of occurrences that are related to criterion variables such as population size. It thus can be seen as another ecologically rational cognitive strategy belonging to the *adaptive toolbox* of fast and frugal heuristics (Gigerenzer et al., 1999). The fluency heuristic also shares an important property with one of the three major heuristics investigated in the heuristics-and-biases research program, namely, availability (Kahneman, Slovic, & Tversky, 1982): Both the availability heuristic and the fluency heuristic capitalize on a subjective sense of memory fluency. Tversky and Kahneman (1973) suggested that people using the availability heuristic assess the probability and the frequency of events on the basis of the ease or the frequency with which relevant instances of those events can be retrieved from memory. Thus, they proposed two notions of availability

(Tversky & Kahneman, 1973, pp. 208, 210), one that depends on the actual frequencies of instances retrieved and one that depends on the ease with which the operation of retrieval can be performed (for more on the distinction between these two notions of availability, see Hertwig, Pachur, & Kurzenhäuser, 2005, and Sedlmeier, Hertwig, & Gigerenzer, 1998).

If one understands availability to mean ease of retrieval, then the question arises of how ease should be measured. Sedlmeier et al. (1998), for example, proposed measuring ease in terms of speed of retrieval of an instance (e.g., words with a letter "r" in the third position). Interpreted in this way, availability becomes nearly interchangeable with fluency as we use it, although the fluency heuristic retrieves the event itself (e.g., the name of a disease), whereas the availability heuristic retrieves instances from the class of events (e.g., people who died of a heart attack vs. people who died of lung cancer to estimate which of the two diseases has a (p.165) higher mortality rate). We have no objection to the idea that the fluency heuristic falls under the broad rubric of availability. In fact, we believe that our 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., Fiedler, 1983; Gigerenzer & Goldstein, 1996; Lopes & Oden, 1991). In the end, how one labels the heuristic that we have called fluency is immaterial because, as Hintzman (1990) observed, "the explanatory burden is carried by the nature of the proposed mechanisms and their interactions, not by what they are called" (p. 121).

What Came First: The Forgetting or the Heuristics?

One interpretation of the beneficial effect of forgetting as identified here is that the memory system loses information at the rate that it does in order to boost the performance of the recognition and fluency heuristics and perhaps other heuristics. One could even hypothesize that a beneficial amount of forgetting has evolved in the cognitive architecture in the service of memory-based inference heuristics. Though such a causal link may be possible in theory, we doubt that evolving inferential heuristics gave rise to a degree of forgetting that optimized their performance, because memory has evolved in the service of multiple goals. It is therefore problematic to argue that specific properties of human memory—for instance, forgetting and limited short-term memory capacity—have optimally evolved in the service of a single function. Although such arguments are appealing—for an example, see Kareev's (2000) conjecture that limits on working memory capacity have evolved "so as to protect organisms from missing strong correlations and to help them handle the daunting tasks of induction" (p. 401)—they often lack a rationale for assuming that the function in question has priority over others. We find it more plausible that the recognition heuristic, the fluency heuristic, and perhaps other heuristics have arisen over phylogenetic or ontogenetic time to exploit the existing forgetting dynamics of memory. If this were true, a different set of properties of memory (e.g., different forgetting functions) could have given rise to a different suite of heuristics.

Conclusion

Analyses of cognitive limits, a well-studied topic in psychology, are usually underpinned by the assumption that these limits, such as forgetting, pose a serious liability. In contrast, we demonstrated (p.166) that forgetting might facilitate human inference by strengthening the chain of correlations that link the decision criteria, environmental frequencies, memory record, activations, and the speed and accuracy of fundamental memory retrieval processes with the decision that is ultimately made. The recognition and fluency heuristics, we argued, use the characteristics of basic retrieval processes as a means to indirectly tap the environmental frequency information locked in memory activations. In light of the growing collection of beneficial effects ascribed to cognitive limits (see Hertwig & Todd, 2003), we believe it timely to reconsider their often exclusively negative status and to investigate which limits may have evolved to foster which cognitive processes and which processes may have evolved to exploit specific limits—as we propose in the case of heuristic inference and forgetting.

Notes:

(1.) The actual predictions of the rational analyses are in terms of odds, where odds equal $p/(1-p)$. However, when probability p is very small, odds and p are quite similar. For example, a p of 0.05 corresponds to odds of 0.0526. As most people are more comfortable thinking in terms of probabilities, we use them here.