Available online at www.sciencedirect.com

SciVerse ScienceDirect

Cognitive Systems Research 24 (2013) 96–103

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

A rational approach to memory search termination

Action editor: David Peebles

Eddy J. Davelaar [a,*], Erica C. Yu [b], J. Isaiah Harbison [b], Erika K. Hussey [b], Michael R. Dougherty [b]

[a] Department of Psychological Sciences, Birkbeck University of London, Malet Street, WC1E 7HX London, UK
[b] Department of Psychology, University of Maryland at College Park, College Park, MD 20742, USA

Available online 4 January 2013

Abstract

An important component of many, if not all, real-world retrieval tasks is the decision to terminate memory search. Despite its importance, systematic evaluations of the potential rules for terminating search are scarce. Recent work has focused on two variables: the total time spent in memory search before search is terminated and the exit latency (the time between the last retrieved item and the time of search termination). These variables have been shown to limit the number of plausible rules for terminating memory search. Here, we introduce an alternative stopping rule based on a rational moment-to-moment cost–benefit analysis and derive a closed-form expression of the exit latency function using this rational approach. We show the model's ability to capture critical latency data and make testable predictions about the influence of changing the relative costs and benefits of memory search. Results from an experiment are presented that support the model's predictions. We conclude that the decision to terminate memory search is based on moment-to-moment changes in subjective utility of retrieved memories.

Keywords: Stopping rule; Memory retrieval; Free recall

1. Introduction

One of the most influential developments in cognitive psychology and cognitive science is that of a detailed theoretical framework of memory processes. In the late 1960s, Murdock (1967) summarized a view held by many theorists in the "modal model", a model in which information (memoranda) transfers from sensory memory to short-term memory and then to long-term memory, with each subsequent system having greater memory persistence. The modal model was mainly a framework of memory encoding and the details of memory retrieval were left less-specified. Later theories explicated the retrieval processes in more detail (Anderson, 1972; Metcalfe & Murdock, 1981; Raaijmakers & Shiffrin, 1980, 1981). A common aspect of these theories is the assumption that retrieval from memory can be seen as a search process (Yntema & Trask, 1963) that takes time to complete. Importantly, in order to characterize this search process, models of memory were endowed with stopping rules that prevent the models from continuing search indefinitely. Despite the fact that theoreticians have been quick to incorporate stopping rules into models of memory, research evaluating the class of stopping rules that might characterize people's decision to terminate memory search is limited.

The evaluation of stopping rules in models of recall is of both theoretical and practical interest. From a theoretical perspective, the goal of developing a comprehensive model of memory retrieval necessitates that we specify the control systems that operate on the memory representations (Newell, 1973). Any particular memory model might yield qualitatively different predictions depending on the specification of the control structures. This is particularly true for stopping rules, since the particular stopping rule employed will affect how long the model will persist in search, which

* Corresponding author.
E-mail address: e.davelaar@bbk.ac.uk (E.J. Davelaar).

1389-0417/$ - see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.cogsys.2012.12.012

can potentially affect the output of the model (number of items retrieved) and retrieval latencies.

From a practical perspective, understanding stopping rules in the domain of memory retrieval can be informative for the development of artificial intelligence and decision support systems, as well as for cognitive models of diagnostic hypothesis generation and judgment (Thomas, Dougherty, Sprenger, & Harbison, 2008). Within these systems, different stopping rules may yield qualitatively different solutions to diagnostic problems.

In this paper, we extend the analytical work of Harbison, Dougherty, Davelaar, and Fayyad (2009) and propose a stopping rule that is motivated by a rational analysis of memory (Anderson & Milson, 1989). The predictions of the resulting rational model are tested against existing and new data.

1.1. Stopping rules

Atkinson and Shiffrin (1968, page 121) suggested a number of stopping rules, which have been implemented in models by various authors. These different stopping rules are: an internal time limit (Davelaar, 2007; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Diller, Nobel, & Shiffrin, 2001; Farrell & Lewandowsky, 2002; Metcalfe & Murdock, 1981), a strength threshold (Anderson, Bothell, Lebiere, & Matessa, 1998; Diller et al., 2001), and an event-counter that would terminate search after a prespecified number of events (Raaijmakers & Shiffrin, 1980, 1981; Shiffrin, 1970).

Given the various stopping rules employed in the literature, it is clear that little heed has been paid to how a chosen stopping rule might affect the model's retrieval dynamics. Furthermore, the empirical research on which to test candidate stopping rules has been missing. The presence of self-terminating stopping rules in models of memory is in recognition of the fact that human observers are often required to self-terminate retrieval. Yet, most empirical studies of free recall have masked the contribution of stopping rules by providing participants with a pre-set retrieval interval. The use of pre-set retrieval intervals eliminates the need for the participant to utilize a stopping rule and even if participants were to use such a rule there would be no method of measuring it.

In order to address stopping rules in recall, one needs to allow participants to terminate their own retrieval episode. Consequently, the procedure of interest here is one in which the participants are given all the time they need for retrieval, but allowed to terminate retrieval whenever they wish (Dougherty & Harbison, 2007; Harbison et al., 2009). This paradigm yields two temporal variables anticipated by models of memory that are important for understanding search termination, but which have been obscured by the choice of experimental design in the literature. The first of these response time measures is *total time*. Total time indexes the elapsed time between the onset of a retrieval cue (i.e., the initiation of the retrieval episode) and the decision to terminate retrieval (i.e., termination of the retrieval episode). The fact that models of memory incorporate stopping rules suggests that these models yield total time predictions. Obviously, different stopping rules will yield different total time predictions, but on an intuitive level one would expect total time to be monotonically related to total number of items retrieved: Total time should increase with the number of items retrieved.

The second response time measure is what Dougherty and Harbison (2007) called the *exit latency*. Exit latencies index the amount of time between the final successful retrieval and the decision to terminate the search. In contrast to total time, there is no obvious, intuitive prediction regarding how long participants will persist in retrieval as a function of number of successful retrieval attempts. Thus, exit latencies provide a potentially diagnostic source of data for evaluating stopping rules, particularly when considered in conjunction with the total time measure (see Harbison et al., 2009).

Few published studies report data on the two temporal variables relevant for measuring termination decisions (Dougherty & Harbison, 2007; Harbison et al., 2009; Unsworth, Brewer, & Spillers, 2011). In the study by Dougherty and Harbison (2007), participants were visually presented with a cue word and 10 target words (A–X_1, A–$X_2, \ldots,$ A–X_{10}). They were told to remember the target words that were presented with each cue word. Each list of 10 target words had a unique cue word. Twelve such lists were presented in blocks of three. After each block of lists were presented, participants were given a cue word and had to report verbally as many words studied with that cue word (A–?) as they could. Responses were recorded and participants pressed the space-bar to indicate that they could not generate additional words. The total time participants spent in search was measured as the time between presentation onset of the cue for retrieval and the time of pressing the space-bar. The exit latency was measured as the time interval between the last retrieved item and the time of pressing the space-bar.

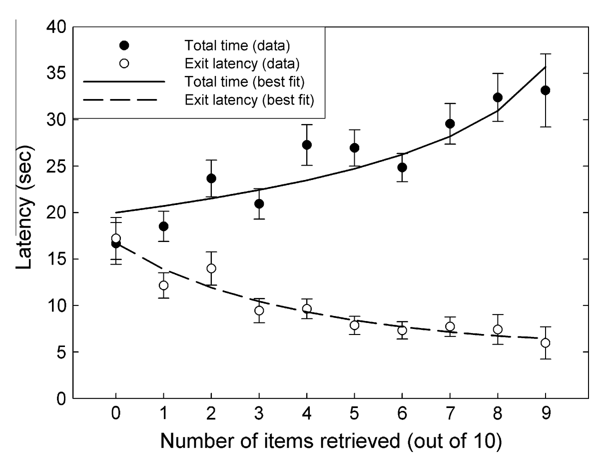

Fig. 1. Empirical data from Dougherty and Harbison (2007) with fits of the best fitting model ($R^2 = .95$). Error bars represent the standard error of the means.

Fig. 1 presents the pattern of results regarding the stopping and exit latencies as a function of the number of words retrieved in that trial. The solid lines are the best-fitting curves from the rational analysis introduced later. Fig. 1 shows that total time is an increasing function of the number of words retrieved in that trial, whereas exit latency is a negatively decelerating function of the number of words retrieved in that trial.

This pattern is consistent across experimental manipulations (Dougherty & Harbison, 2007; Harbison et al., 2009; Unsworth et al., 2011) and is a pattern that challenges the assumptions of many theoretical models of free recall (Harbison et al., 2009) as discussed next.

1.2. A rational stopping rule

Harbison et al. (2009) conducted a simulation study to compare several of the stopping rules suggested by Atkinson and Shiffrin (1968). They used the Search of Associative Memory (SAM; Raaijmakers & Shiffrin, 1981) and implemented the different stopping rules. The models were evaluated on their fit to data. Of the rules tested, only the total number of failures rule fitted the data both qualitatively and quantitatively. This is the rule that was used in the original SAM paper. In SAM, every retrieval attempt is comprised of two stages. In the first stage, an item is sampled probabilistically from episodic memory in proportion to the relative strength of the memory trace. This is akin to locating the item in the memory space. After locating the item, in the second stage, the item is recovered. The recovery probability is a monotonic function of the absolute strength of the memory trace. In SAM, retrieval failures are defined as either sampling an item that was sampled before or failing to recover a sampled item. As the retrieval phase ensues, the probability of resampling increases and thus the rate of retrieval failure increases, too. The stopping rule puts an upper limit on the number of failures such that when this limit is reached the simulation terminates the retrieval phase. The total number of failures rule is a special case of an iterative rule that is only concerned with the current sample from memory and the total accumulated number of failures. This lends itself to a rational analysis of the same rule which can make novel predictions.

We see memory retrieval as a form of information sampling for which a cost is incurred with every sampling attempt and a benefit is obtained for successful retrievals. We propose a rational stopping rule in which the decision to terminate search depends only on the information available at the last time-step. We converged on the following rule (cf. Anderson & Milson, 1989):

> *Terminate search when the additional cost of retrieving the next item starts to outweigh the relative or marginal benefit of having retrieved that item.*

We define a memory value function in which the total net value during the retrieval phase is a function of the total number of items retrieved at the elapsed retrieval time. We assume that a cost, a, is incurred with every sampling attempt, t, and a benefit, b, is obtained with every successful retrieval. We define the memory value function as:

$$V(t) = Q + bN(t) - at \quad (1)$$

where b and a are the benefit and cost parameters. $N(t)$ is the total number of items retrieved at time t. The net_value, $V(t)$, has a constant, Q, which is a (currently unconstrained) free parameter that allows the initial Q/a seconds to be without retrieval and is therefore related to the maximum first recall latency. Possible interpretations are that Q is related to factors such as motivation or time-pressure, neither of which is tested here.

This stopping rule is based on the additional cost in the next Δt seconds of retrieving the next item compared to the relative benefit of having retrieved that item. In other words:

$$\mathrm{cost}(t + \Delta t) - \mathrm{cost}(t) > b/V(t)$$

$$a * \Delta t > b/V(t) \quad (2)$$

This equation states that when the difference in cost at time t and time $t + \Delta t$ is greater than the relative benefit, the memory search will be terminated.

We implemented this rule in SAM, replacing the retrieval failures rule. Fig. 2 (top panel) shows the latency functions for the original SAM model. The retrieval failures rule captures both the increase in total time with total number of words recalled and the convex exit latency function. The bottom panel of Fig. 2 shows the latency functions of

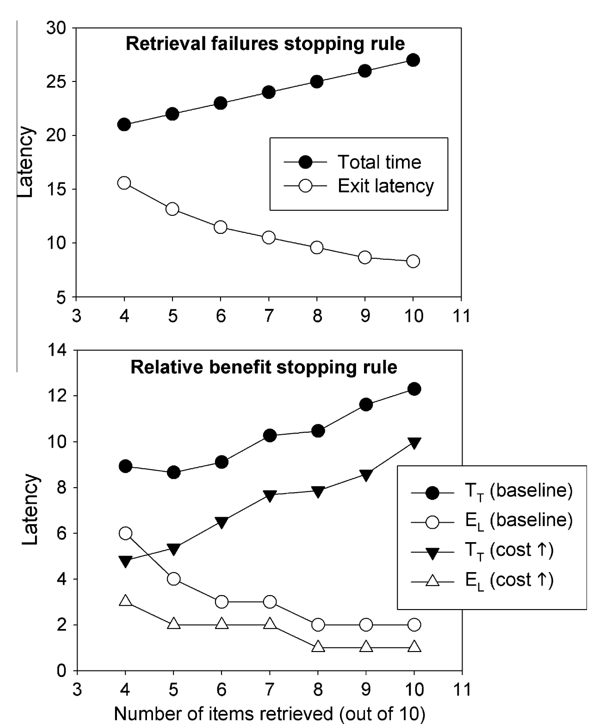

Fig. 2. Simulation results of SAM with a retrieval-failures (top) and a relative-benefit stopping rule (bottom), showing predictions for total time (T_T) and exit latency (E_L) when the cost of each memory sampling attempt increases.

SAM with relative benefit stopping rule. This model also captures the typical data patterns. In addition, when the relative cost is increased, the model predicts that both latency functions are lowered. That is, increased cost decreases the total time spent in memory search and decreases the time spent after the last item before deciding that further retrieval is futile. Importantly, these changes are independent of the total number of items retrieved. This is a nontrivial prediction. Intuitively, one would expect that increasing the cost of retrieval would lead to less time spent retrieving items to minimize the total cost. This would then lead to lower total recall. However, the rational model predicts that only the total retrieval times and the stopping decisions are affected.

To summarize, a SAM implementation in which the decision to stop memory search is based on a moment-to-moment cost/benefit analysis predicts that when the cost increases (or benefit decreases) the search will terminate sooner. We test this prediction in the following experiment.

Before addressing the experiment, there is a technicality that needs to be addressed. In the simulations, the total recall is seldom zero or one, whereas participants do have trials in which they could not retrieve anything or at most one item. In order to test the rational stopping rule against real data, we derived a closed-form expression for the exit latency.[1] This derivation is based on the value function in Eq. (1). We replaced the total number of items retrieved, $N(t)$, with:

$$N(t) = L(1 - e^{-\lambda(t-\delta)}) \tag{3}$$

with list length, L, rate of cumulative retrieval, λ, and offset for starting retrieval, δ. This equation has been shown to provide good approximations to observed data (Wixted & Rohrer, 1994) and follows from a sampling-with-replacement model (Indow & Togano, 1970). To compare, a sampling-without-replacement model would retrieve all items in a free recall task irrespective of the number of items to be retrieved. This counters actual observed data. We assume that the participant aims to obtain the maximal possible net_value. That is the participant stops search when $V(t)$ is maximal, i.e., $dV(t)/dN(t) = 0$.

$$V(t) = Q + bN(t) - at$$

$$dV(t)/dt = bL\lambda e^{-\lambda(t-\delta)} - a$$

$$t_{\text{stop}} = -\lambda^{-1}Ln(a/bL\lambda) + \delta \tag{4}$$

Substituting t_{stop} for t in (3) and solving for N_{stop} gives:

$$N_{\text{stop}} = L - a/b\lambda$$

The derivation of the exit latency function is based on the additional cost of retrieving the next item compared to the relative benefit of having retrieved that item:

$$a\Delta t = b/V(t)$$

$$at_{\text{exit}} = b/(Q + bN_{\text{stop}} - at_{\text{stop}})$$

$$t_{\text{exit}} = \frac{b}{a(Q + bN_{\text{stop}} - at_{\text{stop}})}$$

$$t_{\text{exit}} = \frac{b}{a\left(Q + bN_{\text{stop}} + a\left(\frac{1}{\lambda}Ln(1 - N_{\text{stop}}/L) - \delta\right)\right)} \tag{5}$$

Fig. 1 shows the fits of this model to the data by Dougherty and Harbison (2007). As Eq. (5) has several free parameters, it is only used here to test the prediction that changes (from a baseline condition) in retrieval costs or benefits have a noticeable impact on total time and exit latency.

2. Experiment

2.1. Methods

2.1.1. Participants

Forty-five college-aged participants were recruited from the University of Maryland subject pool and received performance-based compensation ($15 or $20) for participation in the study. Two participants were removed from analysis due to data collection errors.

2.1.2. Design and materials

The design used a delayed free recall paradigm whereby participants studied word lists, completed distractor math problems, and verbally recalled words from the most recent list using a PC-based microphone. The session was presented in two blocks. The first was a baseline block of 16 trials with the same payoff structure across participants (+100 for a correct recall, −100 for each second spent and incorrect recall). In the second block, cost and reward were varied between participants: one group was given an increase in reward (+150) for a correct recall and a simultaneous decrease (−50) for each second spent and each incorrect recall; the other group was given the inverse (+50 reward, −150 cost). The retrieval protocol followed the self-terminated search paradigm used by Dougherty and Harbison (2007): participants were instructed that they had unlimited time to recall words and could end the recall period at any time by pressing the spacebar. The experimenter monitored the participant's recall and updated the participant's score in real-time, providing feedback to the participant on screen. Thirty-two lists of monosyllabic words were randomly created for each participant. List length was varied between 5, 7, 9, and 11 words and presentation order was randomized to prevent strategy use.

2.1.3. Procedure

Participants were informed they would complete a verbal recall task. The study words were presented sequentially in the center of the computer monitor for 2 s each. Following each study list, a distractor task was presented, which consisted of two simple, timed math problems.

[1] The use of closed-form expressions facilitates identification of misfits that are due to theoretical misfits instead of sampling noise. They also speed up the fitting procedure, whereas simulation models require extremely large samples to fit the latencies at very low and very high levels of total recall.

Problems contained three digits and two operands (e.g., $3 * 2 + 1 = ?$) and always resulted in a single-digit answer (digits 0–9). A question mark prompted the participant to enter an answer. Components of the math problem were presented sequentially for 1 s each. After two math problems, participants were prompted to begin verbally recalling words from the most recent study list and press the spacebar when they were finished retrieving. After the spacebar press, participants were prompted to press the spacebar again to begin the next study list when they were ready.

2.1.4. Coding

Using PennTotalRecall audio-analysis software, verbal retrieval data were retrospectively analyzed with millisecond accuracy. Two coders independently coded: (1) all words that were produced by each participant on each trial, (2) the time stamps of the verbal onset of all generated words, and (3) the time stamps of retrieval termination (i.e., times associated with spacebar presses). From these data, number of items retrieved, number of intrusions including repetitions and intra- and extra-list false alarms, inter-retrieval times, and exit latencies (i.e., time between end of final word retrieved and retrieval termination) were calculated. Each subject's trials were averaged before summarizing across subjects.

2.2. Results

A 2×2 mixed design included an initial baseline control environment (+100 correct recall, −100 s spent or incorrect recall) and a second payoff environment varied between subjects (favorable: +150, −50; unfavorable: +50, −150). Due to steep learning curves in each new environment (see Fig. 3B), only the last 8 of the 16 trials in each block were included in the following repeated measures ANOVA analyses.

The net points (rewards for correct recalls less the penalties for incorrect recalls and time spent) were updated in real-time for participants to use as feedback to monitor their own retrieval performance. As predicted, net points earned in each block (see Fig. 3A) increased over time [$F(1,41) = 6.77$, $p < .013$, $\eta_p^2 = .14$] and the participants for whom the rewards increased and costs decreased earned more points overall [$F(1,41) = 15.23$, $p < .001$, $\eta_p^2 = .27$]; while net points in the baseline block were equivalent across conditions (favorable: $M = -23.21$, $\mathrm{SE} = 41.04$; unfavorable: $M = -35.80$, $\mathrm{SE} = 40.10$), performance splits drastically in the second block (favorable: $M = 281.85$, $\mathrm{SE} = 54.97$; unfavorable: -161.08, $\mathrm{SE} = 53.71$; condition × time: 38.80, $p < .001$, $\eta_p^2 = .49$), showing that the manipulation worked.

Total number recalled, including intrusions and repetitions, did not vary due to time, payoff environment, or an interaction of the two [conditions: $F(1,41) = 1.61$, ns, $\eta_p^2 = .04$; time: $F(1,41) = 3.36$, ns, $\eta_p^2 = .08$; condition × time: $F(1,41) = 3.84$, ns, $\eta_p^2 = .09$]. Overall, the rate of intrusions was low (0.3 intrusions per list).

The retrieval rate (see Fig. 3B) showed an asymmetry with the rate remaining at the baseline level (1.18 words/s), but increasing in the unfavorable condition [condition: $F(1,41) = 5.40$, $p < .05$, $\eta_p^2 = .12$; block: $F(1,41) = 47.97$, $p < .001$, $\eta_p^2 = .54$; condition × block: $F(1,41) = 27.35$, $p < .001$, $\eta_p^2 = .40$].

Temporal measures were sensitive to learning during the experiment: total time and exit latency both decreased significantly for all participants [total time: $F(1,41) = 22.19$, $p < .001$, $\eta_p^2 = .35$; exit latency: $F(1,41) = 12.95$, $p < .001$, $\eta_p^2 = .24$]. This performance change came primarily from the participants for whom the rewards decreased and the costs increased: the interaction between time and payoff structure was significant for both measures [total time: $F(1,41) = 29.01$, $p < .001$, $\eta_p^2 = .41$; exit latency: $F(1,41) = 9.98$, $p < .003$, $\eta_p^2 = .20$], but the main effects of condition were not significant [total time: $F(1,41) = 1.14$, ns, $\eta_p^2 = .03$; exit latency: $F(1,41) = 2.54$, ns, $\eta_p^2 = .06$].

Fig. 4 shows the data on the retrieval latencies broken down by block and condition. Only those levels of number of items retrieved for which there were sufficient datapoints were included for the model fit. The solid lines are the best

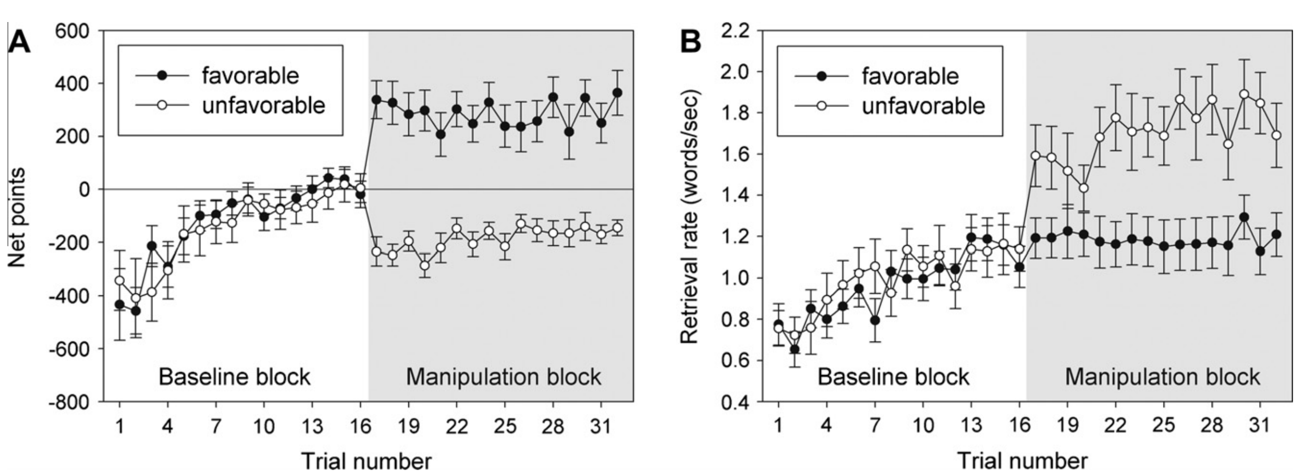

Fig. 3. Net points (A) and retrieval rate (B) over individual trials in each of the two blocks broken down by the manipulation (favorable versus unfavorable) in the second block. Error bars represent standard error of the mean.

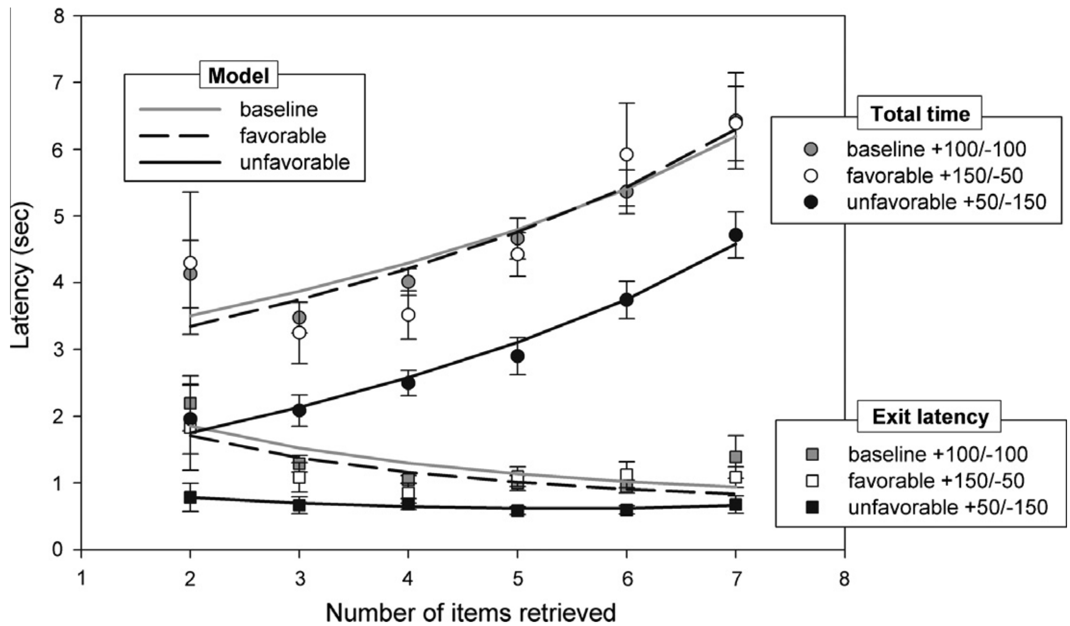

Fig. 4. Total time (circles) and exit latency (squares) functions for the baseline (both groups combined) and the second block (favorable and unfavorable condition). Only the last 8 trials of each block were used. Lines represent the best-fitting model ($R^2 = .97$). Error bars are standard errors of the mean.

fits with Eq. (5). The prediction was that increase in cost or decrease in benefit would lower the latencies. Compared to the baseline condition, making the test harder by increasing the cost and decreasing the benefit did indeed lower all retrieval latencies. Nevertheless, the opposite manipulation, decreasing the cost while simultaneously increasing the benefit, did not change the latencies compared to baseline. We address this asymmetry in the general discussion. The experiment confirmed the prediction by the rational model and showed that increasing the cost does not influence the total recall, as one would intuitively expect. Instead it only affected the rate of retrieval and the decision to terminate search.

3. General discussion

The purpose of this paper was to extend our earlier work on stopping rules by proposing a stopping mechanism that was motivated by a rational analysis of decisions made on a moment-to-moment basis. We demonstrated the comparability of predictions made by a SAM implementation of the rule. We also derived a closed-form expression of the exit latency function that fits the data presented by Dougherty and Harbison (2007) and makes testable predictions about the influence of monetary payoff structure on retrieval latencies and the decision to stop memory search.

The prediction was that making it harder to gain points would lower the retrieval latencies due to higher probability of stopping, whereas the reverse would be the case when it was easier to gain points. Interestingly, only the former prediction was borne out by the data and model fits. The results might be seen as an instance of loss aversion. In particular, during the baseline trials participants are learning a strategy by which they can minimize the amount of losses. This strategy is maintained in the second block if it does not lead to further losses, even if the payoff scheme is more favorable. Because the rational model is symmetric in the influence of costs and benefits, it fails to capture the asymmetry seen in the data. However, a heavy-side function over the difference in expected utility (EU) could rectify this discrepancy (i.e., adjust parameters when $EU_{new} < EU_{old}$), albeit in an ad hoc manner.

Anderson and colleagues provided a rational analysis of the free recall task (Anderson & Milson, 1989; Anderson & Schooler, 1991), in which each item has a need probability associated with it. Only those items are retrieved whose need probability is larger than a certain criterion, which increases with the time spent inspecting an item before accepting or rejecting it. Anderson and Milson (1989) were able to capture a number of basic memory phenomena using their adaptive perspective. However, their analysis only provided the time of the last retrieved item and not of the exact time of terminating memory search. A possibility would be to use the criterion to estimate the termination time, but this would require knowing the functional form of how the criterion changes during item inspection. Nevertheless, the success of Anderson's rational analysis and our current results warrants investigating how these can be combined and would allow analyzing the consequences of different retrieval processes on stopping rules. This also applies to research that is inspired by the animal foraging literature, such as problem solving (Payne & Duggan, 2011) and information foraging (Pirolli & Card, 1999). We leave such an endeavor for the future.

Our analysis suggests that stopping rules should play a more central role in the development and testing of models of memory. The choice of stopping rule has a major impact on the overall model behavior. Obviously, one of the ultimate goals of memory theory is to characterize memory retrieval in general, both in and out of the lab. By focusing more on how people terminate memory search, we can bring our models more in line with the type of retrieval tasks that characterize retrieval tasks outside of the free-recall paradigm.

Future work could address a number of remaining issues regarding the model. First, in our experiment we treated a single intrusion as equally costly as one second in the retrieval phase. It may require more or less than one second to utter an intrusion depending on word characteristics such as word frequency and word length. Although intrusions tend to be rare in free recall, assessing the cost of an intrusion would be one area of further inquiry. Second, and related to the first, Laming (2009) suggested that free recall terminates when the same word gets retrieved. Specifically, Laming analyzed a particular instantiation of a Markovian memory model in which retrieved items are placed back and on the head of the memory record. The retrieval process is biased towards retrieving the most recent item, irrespective of whether it was already retrieved. The retrieval bias together with the replacement at the head of the memory record makes the Markov model not retrieve all items from the list and settle on retrieving a single item repetitively. Based on this analysis, Laming suggests that the "terminal state – an inability to retrieve anything from memory except the same one word – tells the participant that he or she will not be able to recall any further words" (Laming, 2009, p. 167). However, an implementation of this rule did not produce the exit latency curve seen in the data. Although repetitions are treated in SAM as retrieval failures, not all retrieval failures are repetitions. Apart from the sporadic intrusions, there are also recovery failures. These are retrieval attempts in which a new item was indeed correctly sampled, but the strength of the memory trace is insufficient to produce an output. These weak traces contribute to retrieval failures and could lead to terminating the memory search before a single item is repeated. A formal model comparison could elucidate the status of a repetition as either an initiator of search terminations or a covariate. Third, in our subjective value function, the value of Q is arbitrary. In our current conception, we think that Q might be related to such factors as motivation to retrieve or time pressure. Finally, in the area of task interleaving, Payne, Duggan and Neth (2007) addressed stopping behavior on easy and difficult tasks. A formal model was developed based on foraging theory, which fitted the data remarkably well. Our rational analysis could be compared to stopping rules based on foraging theory (cf. Wilke, Hutchinson, Todd & Czienskowski, 2009).

Investigating stopping rules has important implications for understanding tasks other than free recall. For example, within the medical decision making literature, it is clear that physicians entertain costs when determining when to terminate their retrieval of diagnostic hypotheses from memory (Weber, Böckenholt, Hilton, & Wallace, 1993). More recently, Dougherty and Hunter (2003a, 2003b) showed that the perceived probability of any particular event (a hypothesis) is partially dependent on the number of alternatives retrieved from memory, which was affected by time pressure. This suggests that the decision to terminate memory search will affect his or her perceived probability of a particular hypothesis. Within the frequency judgment literature, Brown and colleagues (Brown, 1995, 1997; Brown & Sinclair, 1999; Conrad, Brown, & Cashman, 1998) have shown that participants' responses to survey questions often are a monotonically increasing function of total time spent searching memory. Thus, the magnitude of participants' frequency judgments on behavioral survey questionnaires should be affected by when they terminate search of long-term memory. Although the above tasks are all quite distinct, they serve to underscore the ubiquity of stopping rules in real-world retrieval tasks. Therefore, understanding how people terminate memory search, and the psychological and ecological variables that affect search termination, is paramount to the development of comprehensive models of memory retrieval and to understanding the dynamics of memory retrieval outside the lab.

In summary, in this paper we obtained further evidence for the view that participants are making adaptive choices to search termination that can be described using a rational analysis of the cost and benefits of memory retrieval.

Acknowledgments

This material is based on work supported by the National Science Foundation under Grants SES-0134678 and BCS-1030831. The authors thank Dave Huber and Rick Thomas for their helpful comments.

References

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. (1972). FRAN: A simulation model of free recall. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation. Advances in research and theory* (Vol. 5, pp. 315–378). New York: Academic Press.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*, 703–719.

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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation. Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1539–1553.

Brown, N. R. (1997). Context memory and the selection of frequency estimation strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 898–914.

Brown, N. R., & Sinclair, R. C. (1999). Estimating number of lifetime sexual partners: Men and women do it differently. *Journal of Sex Research, 36*, 292–297.

Conrad, F., Brown, N. R., & Cashman, E. (1998). Strategies for estimating behavioral frequency in survey interviews. *Memory, 6*, 339–366.

Davelaar, E. J. (2007). Sequential retrieval and inhibition of parallel (re)activated representations: A neurocomputational comparison of competitive queuing and resampling models. *Adaptive Behavior, 15*, 51–71.

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112*, 3–42.

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC-REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 414–435.

Dougherty, M. R., & Harbison, J. I. (2007). Motivated to retrieve: How often are you willing to go back to the well when the well is dry? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 1108–1117.

Dougherty, M. R. P., & Hunter, J. E. (2003a). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica, 113*, 263–282.

Dougherty, M. R. P., & Hunter, J. E. (2003b). Probability judgment and subadditivity: The role of working memory capacity and constraining retrieval. *Memory & Cognition, 31*, 968–982.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*, 59–79.

Harbison, J. I., Dougherty, M. R., Davelaar, E. J., & Fayyad, B. (2009). On the lawfulness of the decision to terminate memory search. *Cognition, 111*, 146–421.

Indow, T., & Togano, K. (1970). On retrieving sequence from long-term memory. *Psychological Review, 77*, 317–331.

Laming, D. (2009). Failure to recall. *Psychological Review, 116*, 157–186.

Metcalfe, J., & Murdock, B. B. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior, 20*, 161–189.

Murdock, B. B. (1967). Recent developments in short-term memory. *British Journal of Psychology, 58*, 421–433.

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*. New York: Academic Press.

Payne, S. J., & Duggan, G. B. (2011). Giving up problem solving. *Memory & Cognition, 39*, 902–913.

Payne, S. J., Duggan, G. B., & Neth, H. (2007). Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology: General, 136*, 370–388.

Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review, 106*, 643–675.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. Bower (Ed.). *The psychology of learning and motivation* (Vol. 14). New York: Academic Press.

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

Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press.

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review, 115*, 155–185.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2011). Factors that influence search termination decision in free recall: An examination of response type and confidence. *Acta Psychologica, 138*, 19–29.

Weber, E. U., Böckenholt, U., Hilton, D. J., & Wallace, B. (1993). Determinants of diagnostic hypothesis generation: Effects of information, base rates, and experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1151–1164.

Wilke, A., Hutchinson, J. M. C., Todd, P. M., & Czienskowski, U. (2009). Fishing for the right words: Decision rules for human foraging behavior in internal search tasks. *Cognitive Science, 33*, 497–529.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review, 1*, 89–106.

Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior, 2*, 65–74.