Research Article

FORGETTING TO REMEMBER:
The Functional Relationship of Decay and Interference

Erik M. Altmann[1] and Wayne D. Gray[2]

[1]Michigan State University and [2]George Mason University

Abstract—*Functional decay theory proposes that decay and interference, historically viewed as competing accounts of forgetting, are instead functionally related. The theory posits that (a) when an attribute must be updated frequently in memory, its current value decays to prevent interference with later values, and (b) the decay rate adapts to the rate of memory updates. Behavioral predictions of the theory were tested in a task-switching paradigm in which memory for the current task had to be updated every few seconds, hundreds of times. Reaction times and error rates both increased gradually between updates, reflecting decay of memory for the current task. This performance decline was slower when updates were less frequent, reflecting a decrease in the decay rate following a decrease in the update rate. A candidate mechanism for controlled decay is proposed, the data are reconciled with practice effects, and implications for models of executive control are discussed.*

In silicon memory, losing information is as simple as deleting a file or overwriting a variable. In human memory, the mechanisms of forgetting are not so clear. Early in the cognitive revolution, one view was that short-term memory is vulnerable to decay (Brown, 1958; Peterson & Peterson, 1959): Information "evaporates" with time unless it is actively maintained. A competing view was that forgetting is caused by interference from distracting elements (Keppel & Underwood, 1962; McGeoch, 1932; Waugh & Norman, 1965): Retrieving one fact among many is something like searching for a friend in a crowded airport.

Interference has dominated the memory literature, but decay is not entirely forgotten. Support for decay comes from evidence of forgetting in the absence of interference (Baddeley & Scott, 1971; Muter, 1980; Reitman, 1974). Further evidence comes from an unlikely source: A data set widely cited in support of interference (Waugh & Norman, 1965) seems to show distinct decay effects as well (Hintzman, 1978; Wickelgren, 1977a). In theoretical terms, decay is arguably an adaptation to the statistical structure of the environment (Anderson & Milson, 1989). Indeed, the functional argument for some decaylike process is compelling. Consider, for example, that one would not be able to drive an automobile correctly if every change in the speed limit increased the overall level of interference in memory for speed limits. Were such interference to build up monotonically, remembering the current speed limit would quickly become impossible. Interference certainly remains a potent source of forgetting even at long delays (Keppel, Postman, & Zavortink, 1968), but unless there is some forgetting of distractors, memory would quickly fail to serve everyday needs (cf. Luria, 1968).

We propose not only that items in memory decay, but also that such decay is functional in that it mitigates interference. This functional view of decay can be understood within a signal detection framework. Figure 1 shows two items being encoded, one after the other. In the top panel, Item 1 has just been encoded. Because memory is a noisy system, the activation of Item 1 is represented by a probability density function. In the middle panel, this density function shifts gradually to the left, representing decay (loss of activation) in the interitem interval. In the bottom panel, Item 2 is encoded, representing a memory update. At this point, there is a positive d' between the two density functions because Item 1 has decayed. This d' allows the system to distinguish Item 2 (the current speed limit, say) from Item 1 (the previous speed limit). That is, a positive d' allows the system to sample the correct item. This mechanism of item discrimination is quite similar to that of strength theory (Murdock, 1974; Wickelgren & Norman, 1966), and to the mechanism of temporal distinctiveness (Baddeley & Hitch, 1993; Neath, 1993) with d' interpreted as the measure of distinctiveness. What is novel here is the specification of decay as the mechanism behind d'.

When memory must be updated frequently, functional decay theory makes two counterintuitive predictions. First, if performance depends on periodically sampling ("calling to mind") the current item, then performance should decline gradually as time passes after an update, as the current item decays and becomes harder to sample. In terms of our example, if performance at some driving-related task depends on remembering the current speed limit, then performance at that task should decline, albeit slightly, within the current speed zone. Such a decline would be adaptive—forgetting the current item would make the next item distinguishable. However, it would also be unusual, in that sampling an item (as during verbal rehearsal, say) is generally taken to increase activation and thereby improve performance (e.g., Glanzer & Cunitz, 1966; Rundus, 1971).

The second prediction is that the cognitive system adapts to changes in the update rate by varying the decay rate. That is, if the number of memory updates per unit time varies, then loss of activation per unit time should also vary. Suppose that the performance demands of a given task are met with a given d'. That is, this d' lets the system distinguish the current item accurately enough to perform the task. If the task's accuracy demands are relatively constant, then d' will be relatively constant, and the amount of decay between updates (Fig. 1) will also be relatively constant. Consequently, if the update rate (memory updates per unit time) varies, then the decay rate (activation loss per unit time) should vary as well, if the system indeed seeks a relatively constant d'. This proposal is counterintuitive in that it departs from traditional conceptions of decay as a passive, fixed-rate process (but see, e.g., Anderson, Fincham, & Douglass, 1999).

DECAY AND INTERFERENCE IN TASK SWITCHING

We tested the predictions of functional decay theory using a task-switching paradigm in which stimuli are presented serially and inter-

---

Address correspondence to Erik M. Altmann, Psychology Department, Michigan State University, East Lansing, MI 48824; e-mail: ema@msu.edu.

Copyright © 2002 American Psychological Society **27**
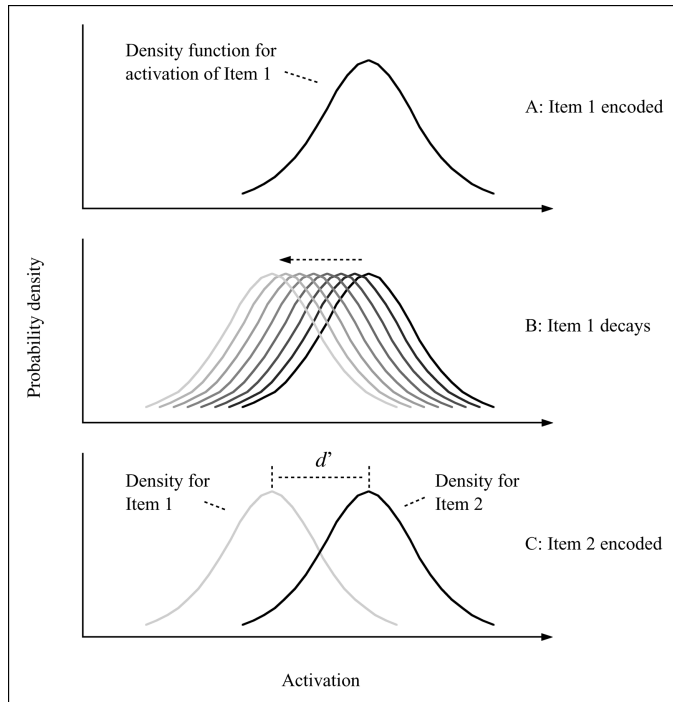
Functional Decay

Fig. 1. A signal detection formulation of functional decay theory. First, Item 1 is encoded in memory (top panel). Its activation is noisy, as shown by the bell-shaped probability density function (PDF). After being encoded, Item 1 decays (middle panel), as shown by the PDF shifting to the left. Next, Item 2 is encoded, with the same initial PDF as Item 1 (bottom panel). Because Item 1 has decayed, there is now a difference d' between the two items' activation levels. The greater this d', the less likely Item 1 is to intrude when the system attempts to retrieve Item 2.

spersed with instructional cues (or simply *instructions*) indicating what task to perform on each stimulus. To perform correctly, subjects have to remember the most recent instruction, a memory we call a *task set*. Because performance depends on the current task set, decay of that set should affect performance and hence affect behavioral measures like response time and error.

If new instructions appear in close succession (every few seconds, say), interference among task sets should build up quickly, much as proactive interference builds up quickly for items from a single category (e.g., Keppel & Underwood, 1962). After hundreds of instructions, this interference will be massive, suggesting the need for a process that mitigates interference and thereby allows the system to distinguish the task set that is currently relevant. An important attribute of our task-switching paradigm is that effects of context are not only controlled within a run, but also minimized. There are no external cues to the current task or to serial position within a run, placing the whole burden of performance on the ability to distinguish the current task set from others in memory.

We propose that the current task set decays gradually while it is current. Even in small amounts, such decay would produce important functional benefits. The argument is inductive: If a task set decays (even a little) while current, then it will be weaker than its successor, so the successor will be stronger than its predecessors and hence more likely to be sampled. Conversely, if the current task set fails to decay,

then interference will build up monotonically, and sustained performance (over hundreds of updates, say) will be impossible. Consistent with our proposal, passive decay reduces the cost of switching tasks, independently of preparatory control processes (Meiran, Chorev, & Sapir, 2000).

Functional decay contrasts starkly with other theories that might be applied to task switching. For example, viewing a trial as practice on the current task set suggests that performance should improve up to the next instruction (e.g., Newell & Rosenbloom, 1981). Alternatively, if only one measure declined from trial to trial (response time, say), this change might reflect a speed-accuracy trade-off (Wickelgren, 1977b). Decay of the current task set, in contrast, predicts that reaction time (RT) and error should increase together.

EXPERIMENT

Subjects performed several thousand trials per session, each under one of two simple instructions. Several hundred instructions were interspersed among trials, updating the current task. To test the gradual-decay prediction, we tested trends across trials between instructions. To test the variable-rate prediction, we manipulated the number of trials between instructions.

Method

Participants

Thirty-six undergraduates from George Mason University participated for credit toward a course requirement. Data from 6 were excluded because their accuracy overall was below 90%.

Materials

Each stimulus was one of the digits 1 through 9 except 5, and an instruction was one of the strings "Even Odd" or "High Low." Stimuli and instructions were presented in the center of a computer screen. Under the even/odd instruction, subjects classified digits as even or odd, and under the high/low instruction, subjects classified digits as greater or less than 5. The same response keys ("C" and "M") were used for both tasks. Stimulus presentation and response recording were controlled by software (Schneider, 1996).

Design and procedure

Trials were self-paced, and participants were asked to work as quickly and accurately as possible. A trial began with the appearance of a stimulus and ended with a key press, at which point the next trial began immediately. Trials were grouped into *runs*, with each run preceded by a 400-ms instruction indicating the task for that run. Two runs made up a *block*, after which the participant was given speed and accuracy feedback and a self-paced rest period. The participant ended the feedback-rest period by pressing the space bar, and the next block began immediately.

To test the variable-rate prediction, we manipulated run length (average number of trials per run) within participants. The two levels, short and long, were blocked, with each level assigned to one half of an experimental session (counterbalanced across participants). Short runs varied randomly from 7 to 13 trials, constrained to sum to 20 trials per block, for a total of 256 instructions interspersed among 2,560 trials. Long runs varied randomly from 17 to 23 trials, constrained to
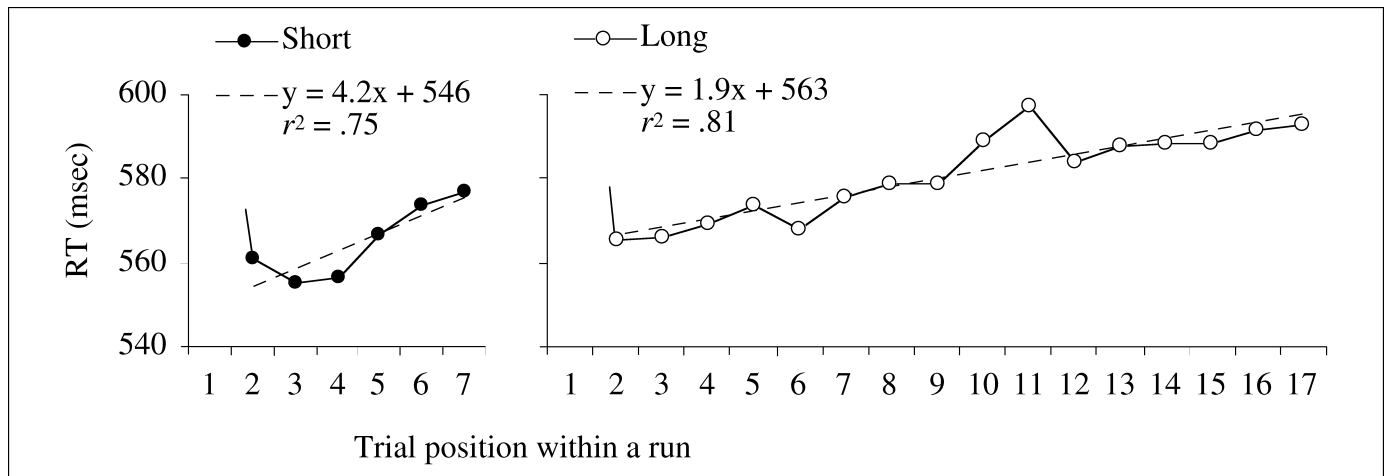
Erik M. Altmann and Wayne D. Gray

Table 1. *Response times and error rates by condition and stimulus position*

Run length	Task	Continuity	Position 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
			Response time (ms)																
Short	Even/odd	No switch	888	574	576	577	584	588	583										
		Switch	903	585	575	574	584	592	603										
	High/low	No switch	784	531	532	537	541	561	560										
		Switch	826	554	538	538	558	555	563										
	Mean		850.5	561.0	555.1	556.7	566.8	574.0	577.2										
Long	Even/odd	No switch	999	583	593	599	598	586	591	601	601	600	628	599	608	607	610	605	606
		Switch	1,031	582	572	586	583	581	588	593	598	596	606	617	612	601	590	608	614
	High/low	No switch	809	549	541	555	552	556	553	561	556	578	568	556	567	581	592	568	565
		Switch	896	549	560	536	564	549	572	562	561	583	587	565	564	566	560	585	588
	Mean		933.8	565.8	566.4	569.1	574.0	567.9	576.0	579.2	579.1	589.4	597.2	584.1	587.7	588.6	588.2	591.6	593.1
			Error (percentage)																
Short	Even/odd	No switch	3.6	3.4	5.2	4.8	4.6	5.1	4.1										
		Switch	2.6	4.7	6.0	5.4	5.9	6.4	5.3										
	High/low	No switch	2.9	3.2	3.1	3.7	4.2	3.2	2.9										
		Switch	2.9	3.3	3.9	3.8	4.1	5.1	4.6										
	Mean		2.97	3.67	4.56	4.42	4.69	4.97	4.22										
Long	Even/odd	No switch	1.9	3.6	5.2	5.7	6.4	4.8	4.0	3.6	5.2	4.8	3.8	4.8	6.2	5.2	5.2	3.8	5.2
		Switch	3.1	4.5	5.5	4.8	3.6	4.8	7.1	7.6	7.9	6.4	5.7	5.5	6.0	6.0	6.0	5.7	6.4
	High/low	No switch	1.4	2.9	4.5	2.9	4.5	3.8	4.3	5.0	7.9	5.5	6.0	3.6	2.4	3.3	6.2	4.3	4.3
		Switch	1.7	3.1	4.3	3.6	5.5	3.8	6.2	4.0	3.3	3.8	5.0	5.0	6.2	4.0	4.8	3.8	5.0
	Mean		2.02	3.51	4.88	4.23	5.00	4.29	5.42	5.06	6.07	5.12	5.12	4.70	5.18	4.64	5.54	4.40	5.24

Functional Decay

Fig. 2. Response time (RT) data from Table 1 showing within-run performance decline for short (left) and long (right) runs. The dashed lines are least-squares regression lines. Data for Position 1 are excluded to remove variance due to preparatory processing (see Fig. 3).

sum to 40 trials per block, for a total of 128 instructions interspersed among 2,560 trials. The transition between conditions was marked by a message indicating that, for the duration of the session, instructions would occur half or twice as often as in the first half of the session.

Half of all blocks were switch blocks, in which the first and second instructions were different, and half were no-switch blocks, in which the first and second instructions were the same (cf. Gopher, Armony, & Greenshpan, 2000; Kramer, Hahn, & Gopher, 1999). The order of switch and no-switch blocks was randomized, subject to the constraint that successive groups of eight consecutive blocks contained an equal number of all configurations of switch and no-switch blocks. We refer to these eight-block groups as *subsessions*.

The first subsession in each condition was excluded from analysis to eliminate the period of most rapid learning. To control for effects of resting or processing the feedback at the ends of blocks, we took all measures on the second run. RT data are means of participants' medians on correct trials from blocks on which accuracy was at least 90%. Error data are means of participants' means. The data are shown in Table 1.

RT Results

RTs relevant to our two predictions are shown in Figure 2. To test for performance decline on short runs, we conducted a 6 × 2 × 2 repeated measures analysis of variance (ANOVA) on position (P2–P7), task (even/odd, high/low), and continuity (no switch, switch). P2 through P7 are the second through seventh trials of the second run of a block. The RT cost of a switch is typically borne on the first trial after the switch (De Jong, Berendsen, & Cools, 1999; Gopher et al., 2000; Rogers & Monsell, 1995), so P1 was excluded from analysis to remove switch cost as a source of variance.

The main effect of position was significant, $F(5, 145) = 5.6$, $p < .001$, $MSE = 1,779$. A contrast by the method of orthogonal polynomials showed a significant linear trend, $F(1, 145) = 21.1$, $p < .001$, which accounted for 75% of the variance attributed to position. No higher-order trends were significant. The main effect of task was also significant, $F(1, 29) = 54.9$, $p < .001$, $MSE = 4,154$. The main effect of continuity was not significant, $p > .1$, and neither were any interactions.

To test for performance decline in the long condition, we conducted a similar analysis with 16 levels of position (P2–P17) instead of 6. The main effect of position was significant, $F(15, 435) = 3.8$, $p < .001$, $MSE = 3,387$. The linear trend was significant, $F(1, 435) = 45.6$, $p < .001$, and accounted for 81% of the variance attributed to position. No higher-order trends were significant.[1] The main effect of task was significant, $F(1, 29) = 59.6$, $p < .001$, $MSE = 9,326$. The main effect of continuity was not significant, $F < 1$, and neither were any interactions.

To test the variable-rate prediction, we compared the slopes of the RT curves. The slope from P2 through P7 for the short condition (4.2 ms/trial) and the slope from P2 through P17 for the long condition (1.9 ms/trial) differed by a factor of 2, $t(28) = 2.2$, $p < .04$, corresponding neatly to the twofold difference in the update rate. These slopes were computed across the full minimum length of the runs in each condition, making them the best measure of rate of decline given our assumption that d' at run's end should be constant. However, this test did confound run length and position, so we also compared the slopes for the long and short conditions on P2 through P7. The slope for the long condition (1.7 ms/trial) again differed from the slope for the short condition (4.2 ms/trial), $t(28) = 2.0$, $p = .058$. We conclude that run length affected the rate of performance decline, with slower decline on long runs than on short runs.

For completeness, we analyzed P1 as well, with a 2 × 2 × 2 ANOVA on run length (short, long), continuity (switch, no switch), and task (even/odd, high/low). Figure 3 shows the P1 data by run length and continuity, with P2 as a baseline. RT was substantially slower for P1 than P2 regardless of continuity (a point we return to in the Discussion). P1 was 83 ms slower in the long condition than in the short condition, $F(1, 29) = 5.5$, $p < .03$, $MSE = 76,506$. The switch cost on P1 (switch – no switch) was 44 ms, $F(1, 29) = 3.4$, $p = .08$, $MSE = 34,639$. Task had an effect, $F(1, 29) = 11.7$, $p < .003$, $MSE = 81,715$, but no interactions were significant.

---

1. The peak at P11 (see Fig. 2) hints that subjects expected an instruction by then, an expectation they could have held over from the short condition. However, on separating the data, we found that long-short sessions, and not short-long sessions, drove the effect.
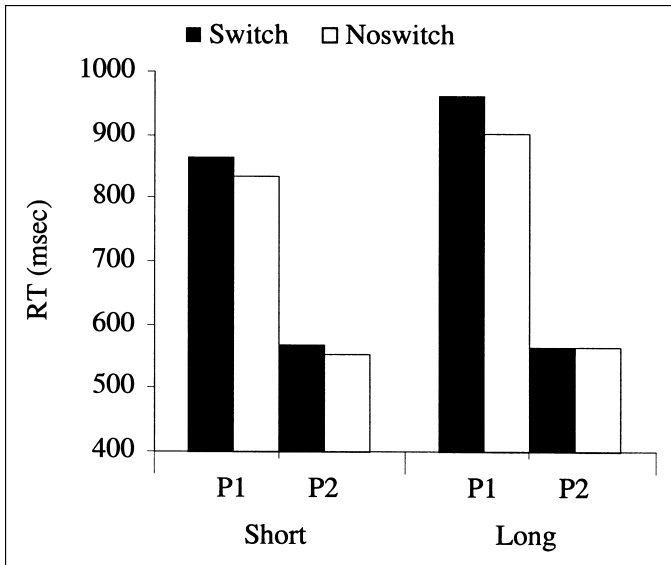
Fig. 3. Response time (RT) data from Table 1 showing the abrupt slowdown on the first trial after an instruction (P1) for short (left) and long (right) runs. Data for the second trial after the instruction (P2) are included as a baseline.

Error Results

Error rates are given in Table 1 and plotted in Figure 4. The analysis parallels that for RT, except that P1 was included, on both empirical and theoretical grounds. Empirically, the effect of task switching on error is not systematic; indeed, error sometimes declines immediately after a switch (e.g., Allport & Wylie, 1999). Theoretically, switch cost has been attributed to preparatory processes that affect P1 RT but not P1 error (De Jong et al., 1999).

In the short condition, the main effect of position was significant, $F(6, 174) = 3.9, p < .002, MSE = 14.5$. The linear trend was significant, $F(1, 174) = 14.5, p < .001$, and accounted for 54% of the variance due to position. The quadratic trend was also significant, $F(1, 174) = 8.8, p < .004$,

and accounted for 38% of the variance due to position. Task had an effect, $F(1, 29) = 10.3, p < .004, MSE = 27.3$, as did continuity, $F(1, 29) = 5.7, p < .03, MSE = 19.0$, but no interactions were significant. The continuity effect is a novel switch cost in the task-switching literature, reflected in terms of error rate but not RT, and prolonged for the duration of the entire switch run rather than localized to the vicinity of the switch trial.

In the long condition, the main effect of position was significant, $F(16, 464) = 3.6, p < .001, MSE = 27.3$. The linear contrast was significant, $F(1, 464) = 16.7, p < .001$, and accounted for 29% of the variance due to position. The quadratic trend was also significant, $F(1, 464) = 19.3, p < .001$, accounting for 33% of the variance, as was the cubic trend, $F(1, 464) = 7.0, p < .01$, which accounted for 12% of the variance. The main effect of task was significant, $F(1, 29) = 4.3, p < .05, MSE = 93.3$, and though errors trended higher for switch than for no-switch runs, the main effect of continuity was not significant, $p > 1$. The only significant interaction was among task, continuity, and position, $F(16, 464) = 2.0, p < .02, MSE = 29.8$.

To test the variable-rate prediction, we compared the slopes of the error curves. The slope from P1 through P7 for the short condition (0.23 errors/trial) and the slope from P1 through P17 for the long condition (0.10 errors/trial) again differed by a factor of 2, $t(28) = 1.8, p = .06$. These are the slopes relevant to our assumption of a constant d' at run's end. However, we also compared the slopes for the long and short conditions on P2 through P7. Unexpectedly, the slope for the long condition (0.42 errors/trial) was twice (and not half) the slope for the short condition (0.23 errors/trial), $t(28) = 2.1, p < .05$. Thus, run length again affected rate of decline independently of position, though in a more complex way for errors than for RTs. More research will be needed to determine what strategic or other factors might cause the rate of decline itself to fluctuate within a run.

DISCUSSION

Performance decline across trials between instructions was robust. It appeared in RTs and error rates, ruling out a speed-accuracy trade-off (Wickelgren, 1977b), and was not an artifact of position, task, or continuity. The effect of update rate was also apparent, with slower decline on long runs than short. These findings are strong initial support for functional decay theory.
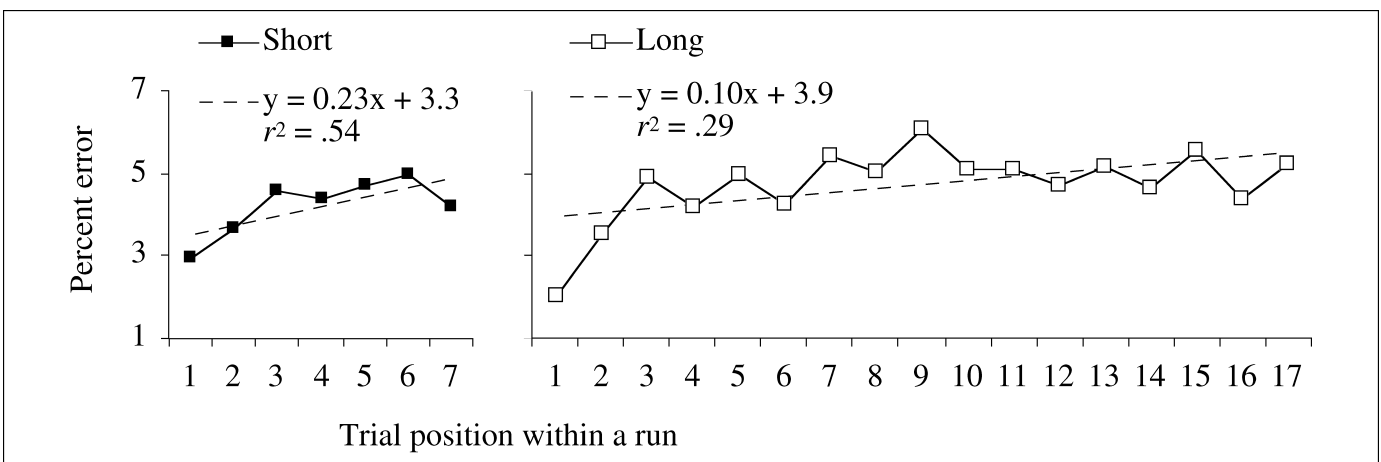

Fig. 4. Error data from Table 1 showing within-run performance decline for short (left) and long (right) runs. The dashed lines are least-squares regression lines.
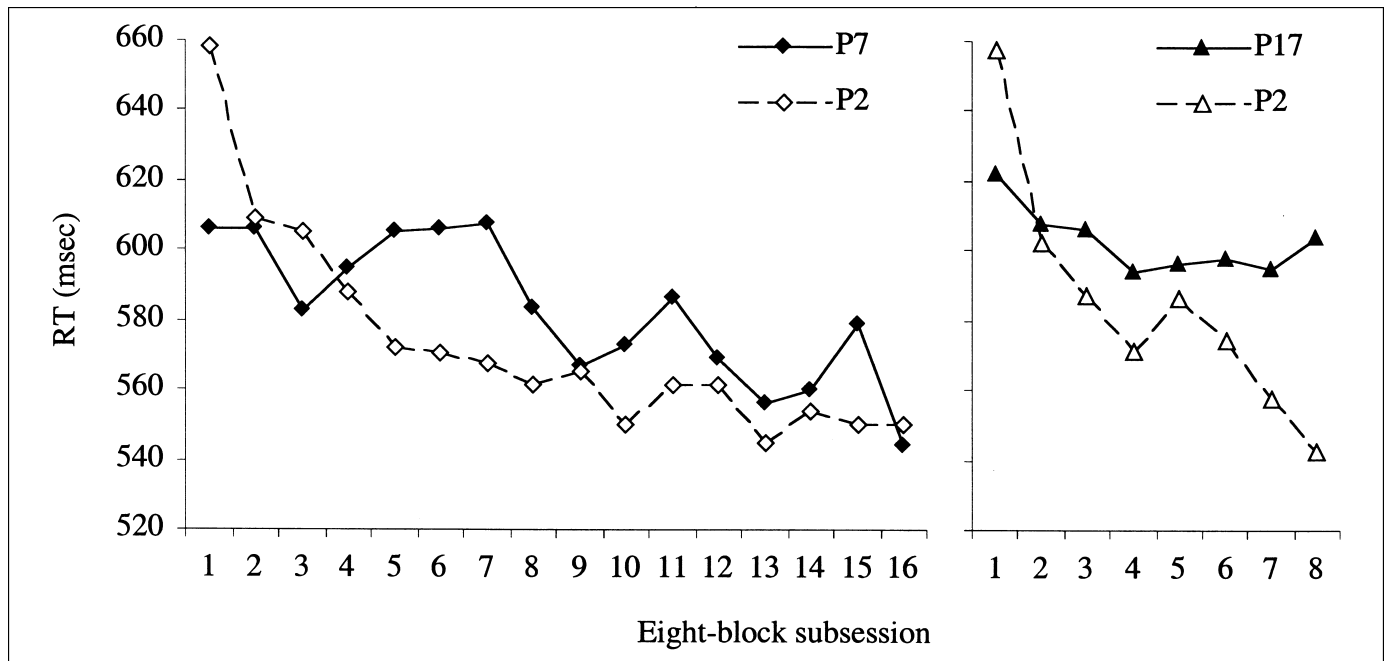
Fig. 5. Response times (RT) as a function of subsession and position (P) for short (left) and long (right) runs. The curves demonstrate both within-run performance decline and practice effects across subsessions.

Historically, a central question about forgetting has been whether interference can explain effects otherwise attributed to decay (e.g., Keppel & Underwood, 1962). Could interference explain within-run performance decline? In our paradigm, one source of proactive interference (PI) is the collection of old task sets in memory. Indeed, PI of this kind is what implies the need for functional decay. However, such PI cannot directly cause within-run performance decline. The number of old task sets in memory changes only between runs, not within a run, so the effects of PI cannot change within a run. Note that we define interference firmly in terms of number of distractors, not in terms of time in any sense. To index interference by time would simply reintroduce the notion of decay and render interference meaningless as a competing hypothesis.

Another possible source of interference is retroactive interference (RI) from trials within the current run. The number of such trials, unlike the number of old task sets, does correlate with performance decline. However, RI cannot be the sole cause of decline because the rate of decline, measured at a given position, would be a function of that position alone. We found, instead, that the rate of decline was affected by run length as well as position: The slopes of the error and RT curves, on identical positions, depended on whether runs were long or short.

Another question is whether variable decay is realistic theoretically, distinct as it is from traditional conceptions of passive fixed-rate decay. In response, we suggest that decay—defined as loss of activation—is a function of sampling as well as time. For example, in the function activation $= \ln(n\sqrt{T})$, adapted from Anderson and Lebiere (1998), activation depends on the total number of samples (n) as well as time since encoding (T). If the sampling rate fluctuates over time, activation will rise and fall, making the sampling rate (or rehearsal rate, in familiar terms) a mechanism by which the cognitive system might vary the decay rate. According to this analysis, passive fixed-rate decay is a special case: If sampling stops altogether, activation

will decay as a power function of time. Formal models based on the assumption that the system manipulates activation via the sampling rate provide close quantitative fits to within-run performance decline and to components of task-switching cost.[2]

Empirically, within-run performance decline may seem unusual, but there is converging evidence in the task-switching literature. For example, in the widely cited work of Rogers and Monsell (1995), Experiment 6 contained runs of four trials, and, indeed, RT showed a monotonic slowing trend starting with P2.[3] Similarly, in switching between languages, "RTs unexpectedly *increased* (slightly but consistently) with increasing number of 'same language' responses" (Meuter & Allport, 1999, p. 32; original emphasis). Visual inspection of data from Kramer et al. (1999, Figs. 2–4) suggests a slight slowing from "1-after" to "3-after" trials (P2 to P4, in our terms). Finally, where speedups have been reported (Allport & Wylie, 2000, Experiment 2; Meiran et al., 2000, Experiment 1), they were measured between and not within runs (a distinction we revisit shortly).

Within-run performance decline is a challenge to models of task switching, suggesting that a systems view of executive control should replace the traditional focus on the temporal locus of the switch. Task-

---

2. Activation $= \ln(n/\sqrt{T})$ predicts that sampling the current task set (once per trial, say) causes activation to increase, rather than decrease, because sampling increments n. However, the equation also predicts that this increase can be counteracted by an abrupt decrease in the sampling rate. Thus, decay within a run will follow massed sampling of the current task set at the start of the run (Altmann & Gray, 1999). This massed sampling explains the response slowdown on P1 (Fig. 3).

3. Pair-wise comparisons on trials were not significant, but were not in our data, either. Within-run performance decline is gradual enough to require testing the trend across the run.

switching models should explain how performance on the actual task unfolds over time. We also suggest that switch cost per se is less interesting an effect of executive control than the much larger slowing on P1 that occurs regardless of continuity (Fig. 3). This "restart cost" has been found by other researchers (Allport & Wylie, 2000; Gopher et al., 2000; Kramer et al., 1999), but to date there is no compelling functional account that integrates it with other task-switching effects. It seems likely that a more complete model of executive control will link both switch cost and restart cost to downstream effects like within-run performance decline.

The data presented here have implications beyond functional decay theory, in that they contribute a boundary condition on practice effects. In organisms from rats to humans, and in human task domains ranging from accounting (Ericsson & Lehmann, 1996) to novel writing (Ohlsson, 1992), practice makes perfect—yet in our paradigm, practice on the current task set seems to hurt. Figure 5 indicates how this conflict is reconciled. The ordinate shows RT (as in Fig. 2), but the abscissa shows eight-block subsessions (see Design and Procedure) instead of trial position. Performance decline is evident in the vertical distance between P2 (the lower, dashed curves) and P7 and P17 (the upper, solid curves) after the first few subsessions.[4] Practice effects are evident at a coarser temporal grain, in the negative slope of RT across subsessions (for all curves). Thus, performance decline and practice effects appear in the same data but at different levels—the former across a few seconds of performance, the latter over 10s of minutes or longer.

In conclusion, we propose that decay and interference are functionally related: If a target decays, it will interfere less with future targets. Consistent with this proposal, our data show a performance decline between task-set updates that we attribute to decay of the current task set. This decline, which is more gradual with less frequent task-set updates, is easily reconciled with the universal law of practice: At a fine temporal grain, practice effects give way to low-level mechanisms for managing interference. This performance decline is also an important empirical constraint on models of executive control—one that seems critical to address as researchers develop such models into more complete and functional systems (Newell, 1990).

REFERENCES

Allport, A., & Wylie, G. (1999). Task-switching: Positive and negative priming of task-set. In G.H. Humphreys, J. Duncan, & A. Treisman (Eds.), *Attention, space, and action: Studies in cognitive neuroscience* (pp. 273–296). New York: Oxford University Press.

Allport, A., & Wylie, G. (2000). Task-switching, stimulus-response binding, and negative priming. In S. Monsell & J.S. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 35–70). Cambridge, MA: MIT Press.

Altmann, E.M., & Gray, W.D. (1999). Serial attention as strategic memory. In *Proceedings of the twenty first annual meeting of the Cognitive Science Society* (pp. 25–30). Hillsdale, NJ: Erlbaum.

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

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

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

Baddeley, A.D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory & Cognition*, *21*, 146–155.

Baddeley, A.D., & Scott, D. (1971). Short term forgetting in the absence of proactive interference. *Quarterly Journal of Experimental Psychology*, *23*, 275–283.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*, 12–21.

De Jong, R., Berendsen, E., & Cools, R. (1999). Goal neglect and inhibitory limitations: Dissociable causes of interference effects in conflict situations. *Acta Psychologica*, *101*, 379–394.

Ericsson, K.A., & Lehmann, A.C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*, 273–305.

Glanzer, M., & Cunitz, A.R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*, 351–360.

Gopher, D., Armony, L., & Greenshpan, Y. (2000). Switching tasks and attention policies. *Journal of Experimental Psychology: General*, *129*, 308–339.

Hintzman, D.L. (1978). *The psychology of learning and memory*. San Francisco: W.H. Freeman.

Keppel, G., Postman, L., & Zavortink, B. (1968). Studies of learning to learn: VIII. The influence of massive amounts of training upon the learning and retention of paired-associate lists. *Journal of Verbal Learning and Verbal Behavior*, *7*, 790–796.

Keppel, G., & Underwood, B.J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, *1*, 153–161.

Kramer, A.F., Hahn, S., & Gopher, D. (1999). Task coordination and aging: Explorations of executive control processes in the task switching paradigm. *Acta Psychologica*, *101*, 339–378.

Luria, A.R. (1968). *The mind of a mnemonist*. New York: Basic Books.

McGeoch, J.A. (1932). Forgetting and the law of disuse. *Psychological Review*, *39*, 352–370.

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology*, *41*, 211–253.

Meuter, R.F.I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*, 25–40.

Murdock, B.B. (1974). *Human memory: Theory and data*. New York: Wiley.

Muter, P. (1980). Very rapid forgetting. *Memory & Cognition*, *8*, 174–179.

Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory & Cognition*, *21*, 689–698.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Ohlsson, S. (1992). The learning curve for writing books: Evidence from Professor Asimov. *Psychological Science*, *3*, 380–382.

Peterson, L.R., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193–198.

Reitman, J.S. (1974). Without surreptitious rehearsal, information in short-term memory decays. *Journal of Verbal Learning and Verbal Behavior*, *13*, 365–377.

Rogers, R.D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207–231.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*, 63–77.

Schneider, W. (1996). MEL Professional (Version 2.1) [Computer software]. Pittsburgh, PA: Psychology Software Tools.

Waugh, N.C., & Norman, D.A. (1965). Primary memory. *Psychological Review*, *72*, 89–104.

Wickelgren, W.A. (1977a). *Learning and memory*. Englewood Cliffs, NJ: Prentice Hall.

Wickelgren, W.A. (1977b). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*, 67–85.

Wickelgren, W.A., & Norman, D.A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, *3*, 316–347.

---

4. P2 is slower than P7 for the first few blocks, but then becomes faster and stays that way. We interpret this crossover to reflect a rapid build-up of task-set interference that triggers adaptations that result in functional decay. How this triggering occurs—how the system might balance the benefits of preventing interference against costs like performance decline—is an important question for future research.