

# In defence of high-speed memory scanning

Saul Sternberg

Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

(Received 15 January 2016; accepted 30 May 2016)

This paper reviews some of the evidence that bears on the existence of a mental high-speed serial exhaustive scanning process (SES) used by humans to interrogate the active memory of a set of items to determine whether it contains a test item. First proposed in the 1960s, based on patterns of reaction times (RTs), numerous later studies supported, elaborated, extended, and limited the generality of SES, while critics claimed that SES never occurred, that predictions from SES were violated, and that other mechanisms produced the RT patterns that led to the idea. I show that some of these claims result from ignoring variations in experimental procedure that produce superficially similar but quantitatively different RT patterns and that, for the original procedures, the most frequently repeated claims that predictions are violated are false. I also discuss evidence against the generality of competing theories of active-memory interrogation, especially those that depend on discrimination of directly accessible “memory-strength”. Some of this evidence has been available since the 1960s but has been ignored by some proponents of alternative theories. Other evidence presented herein is derived from results of one relevant experiment described for the first time, results of another described in more detail than heretofore, and new analyses of old data. Knowledge of brain function acquired during the past half century has increased the plausibility of SES. The conclusion: SES is alive and well, but many associated puzzles merit further investigation, suggestions for which are offered.

**Keywords:** Memory scanning; Memory search; Reaction time; Sternberg paradigm; Sternberg task; Working memory.

## Contents

1	Introduction .....	2021	2.4	Ashby task.....	2029
1.1	Attacks on high-speed memory scanning.....	2022	2.5	Novel negatives task.....	2030
1.2	Alternatives to high-speed memory scanning.....	2023	2.6	Conclusion: Procedure details matter.....	2030
1.3	Brain function and the plausibility of high-speed scanning.....	2023	3	The Lisman–Idiart–Jensen (LIJ) neurophysiological model for memory scanning and memory span.....	2031
1.4	This paper.....	2024	3.1	Nested brain oscillations.....	2031
2	Six tasks of memory interrogation .....	2024	3.2	The “skip” feature of the LIJ model.....	2032
2.1	Sv and Sf tasks.....	2025	3.3	Tests of the LIJ model.....	2033
2.2	Monsell task.....	2027	3.4	Modification of the LIJ model suggested by the rat brain.....	2034
2.3	Cued-response Monsell task.....	2029	4	Scanning rate and the memory span.....	2036
			4.1	Cavanagh’s discovery and further tests .....	2036

Correspondence should be addressed to Saul Sternberg Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104-6241, USA. E-mail: [saul@upenn.edu](mailto:saul@upenn.edu)

I thank the Experimental Psychology Society and the editor, Marc Brysbaert, for giving me the opportunity to revisit this topic; Stephen Monsell and Robert Nosofsky for their remarkably thoughtful, thorough, challenging, and helpful reviews; Stephen Checkosky and Joshua Jacobs for guidance; Derek Besner, William Hockley, Sylvan Kornblum, Ole Jensen, Uta Lass, and John Lisman for comments on the manuscript, and Paula Susan for editorial assistance. I am immensely grateful to Bassam Atallah for providing me with rat hippocampus recordings from the Atallah and Scanziani (2009) study.

4.2 Memory span versus scanning rate and the LIJ model.....2038

4.3 Two kinds of correlation between memory span and scanning rate.....2038

5 Behaviour of the RT variance .....2038

5.1 What do we expect from the serial exhaustive search model?.....2038

5.2 What is found?.....2039

6 Behaviour of the shortest reaction time.....2041

6.1 What do we expect from the serial exhaustive search model?.....2041

6.2 What is found?.....2042

6.3 The shortest RT based on estimated distributions of stage durations.....2044

6.4 Parallel comparisons and the shortest RT .....2045

7 Accessibility of the most accessible item and the mixed-recency conjecture .....2046

8 Retrieval from active versus inactive memory.....2047

8.1 Effect of  $n_{pos}$  on the activation process .....2049

8.2 Consistency with oscillator models of maintenance and search .....2049

8.3 Absence of serial position effects .....2049

8.4 Evidence against strength discrimination .....2049

9 Absence of an effect on RT of negative probe recency .....2050

9.1 The Sf task as association or category learning .....2051

10 Variation of relative response frequencies and evidence-accumulation models .....2052

10.1 The set-size one anomaly .....2056

11 Sequential comparison versus sequential activation...2056

12 Some open questions .....2060

12.1 Effect of positive set size on  $\overline{RT}$ : Comparison or activation? .....2060

12.2 Do Sv and Sf tasks elicit the same process? .....2060

12.3 Scanning rate, stimulus ensemble, and mixed sets .....2060

12.4 Is choice of process obligatory or optional? .....2061

12.5 Predicted additivity of effects of  $Pr\{pos\}$  and stimulus quality.....2061

12.6 Retrieval from active versus inactive memory.....2061

12.7 Tests of the LIJ model.....2062

13 Conclusion .....2062

References .....2063

Appendices.....2071

A Experiments: Abbreviations, sources, and sections where mentioned .....2071

B How to investigate high-speed scanning .....2072

The importance of subject motivation .....2072

Feedback and incentives.....2072

Manipulandum.....2072

Warning signal.....2072

Practice.....2072

Choice of paradigm.....2072

Design and analysis.....2072

C Analysis of the Atallah–Scanziani hippocampus recordings .....2072

D Frequency and recency effects in the Sf task and variants .....2073

Effects with brief response–stimulus intervals ....2073

Probe frequency effects in the Sf task.....2073

Sequential effects in the Sf task.....2074

E Effect of requiring recall in the Sv task .....2074

F Notes for Tables 1, 2, and 4.....2075

Table 1 .....2075

Table 2 .....2075

Table 4 .....2075

1. INTRODUCTION

A small set of digits, letters, words, or other items (the *positive set*, P-set) is presented. After a *probe delay* and a warning signal, a *probe* item appears. The subject responds (usually manually) under time pressure as to whether or not the probe is a member of the P-set (a P-probe). The *negative set*, N-set, which if possible is larger than the P-set, contains the items presented as probes (N-probes) that are not members of the P-set. The *stimulus ensemble*—the union of P-set and N-set—contains all the items in the experiment. When on each trial the members of a new P-set are presented slowly and sequentially in one location, and the probe delay is long, I shall call this procedure the *Sv task*, where “v” represents “varied”: The P-set changes from trial to trial. In the fixed-set procedure, the *Sf task*, a P-set is presented, followed by a series of perhaps 50 to 100 trials on which only probes are presented, to be judged in relation to that fixed P-set. In both tasks, we can determine the *mean reaction-time functions*,  $\overline{RT}_{pos}$  and  $\overline{RT}_{neg}$ , for correct responses on P-trials (trials with P-probes, requiring *positive responses*) and N-trials (trials with N-probes, requiring *negative responses*), respectively. These functions, which describe the increase in mean reaction time ( $\overline{RT}$ ) with the number of items  $n_{pos}$  contained in the P-set, are found to be approximately linear, consistent with the decision being based on a process of serial comparison of an internal representation of the probe to internal representations of the members of the P-set. The slopes  $\beta_{pos}$  and  $\beta_{neg}$  of these functions for positive and negative responses, from 30 to 40 ms/item for digits, indicate that the rate of such an inferred scanning process would

have to be unexpectedly high—much higher, for example, than the rate of speeded articulation (Cowan et al., 1998; Sternberg, Knoll, Monsell, & Wright, 1988) or the rate of speeded implicit speech (Landauer, 1962). Also unexpected is the finding that  $\beta_{\text{pos}}$  and  $\beta_{\text{neg}}$  are approximately equal, suggesting that instead of terminating when a match is discovered, the search is *exhaustive*, continuing through the entire set, even when the probe is a member. The interpretation proposed for these findings and others (Sternberg, 1963, 1966) was that to respond to the probe, subjects engage in a high-speed serial exhaustive search (SES) process.

Why did these findings and their interpretation, published in *Science* in 1966, arouse the remarkable level of interest that it did—the mean RTs replotted in the *Courrier Scientifique* section of *Le Monde* (Verguese, 1966), and the hundreds of related studies continuing to the current century? One can speculate about the reasons: The data are orderly and elegant, with a linear relation between straightforward measures; simple inferences from these data lead to unexpected conclusions (such as the contents of active memory not being immediately accessible); the findings are among the earliest to demonstrate that small effects on RT can be interesting, that the structure of a mental process can be revealed by details of RT data, and that memory can be investigated when it is functioning without error; the fact that the inferred process seems to be unavailable to consciousness; the implausibility of its high speed (given what was then believed about the brain); its exhaustiveness; and, perhaps most important (and the reason for its clinical applications as a measure of brain function), the revelation that the speed of a mental process might be measured separately from the complexities of perception and response.<sup>1</sup>

Scepticism about SES seemed not to be diminished by other findings that accompanied those mentioned above:

1. Qualitatively and quantitatively similar results from the Sv task (short-term memory), and the Sf task, for which sets can be recalled on

the following day (long-term memory) (Sternberg, 1963, 1966).

2. Growth of  $\overline{RT}$  with  $n_{\text{pos}}$  too great, relative to variability, to be explained by independent parallel comparisons (Sternberg, 1963, 1966).
3. Absence of an effect of the size  $n_{\text{neg}}$  of the N-set in the Sf task (Sternberg, 1963, 1966, and 1975, Figure 3).
4. Effects of  $n_{\text{pos}}$  obtained in the Sf task when  $n_{\text{pos}}$  is varied by changing only the mapping of probes onto the two responses, while altering neither the sequence of probes nor the proportion of P-trials (Sternberg, 1963, 1966, 1967c).
5. Minimal modulation of the effects of  $n_{\text{pos}}$  by relative response frequency (Sternberg, 1963; 1969b, Figure 4E).

### 1.1. Attacks on high-speed memory scanning

Since Luce (1986, p. 429) wrote “The attack on Sternberg’s interpretation of his data has been intensive and sustained”, the attack has continued:

Finding out what the measures mean is, however, a far from easy task, and in the meantime I would settle for a moratorium on interpreting their results as direct measures of mental processes such as “memory scanning”. (Baddeley, 1990, p. 279)

[Sternberg’s] claim was a startling one—that short-term memory was characterized by a limit in retrieval: Items were not immediately available, but rather were recognized by sequentially comparing a test item to all items held in short-term memory. ... Attractive though it may be, Sternberg’s model is incorrect ... [instead] retrieval from immediate memory reflects a set of parallel comparisons with an active subset of memory items. (Doshier & Sperling, 1998, pp. 239, 242)

... Sternberg ... suggested that performance was based on exhaustive serial scanning of the list items in search of the probe item. This conclusion was based chiefly on the finding that reaction times (RTs) increased linearly with list length. ... More sophisticated analyses by McElree and Doshier (1989) showed that performance is better explained by direct access than by serial scanning. (Henson, Hartley, Burgess, Hitch, & Flude, 2003, p. 1308)

STM retrieval of item information is a rapid, parallel, content-addressable process. ... Serial-scanning models fell out of favor because of empirical and modeling work showing that parallel

<sup>1</sup>This feature of the paradigm depends on the influence of  $n_{\text{pos}}$  being *selective*: It affects the search process, but not perceptual or response processes. See Sternberg (1969b) for evidence of such selectivity; see also Schweickert, Fisher, and Sung (2012) and Sternberg (1998).

processes provide a better account of the reaction time distributions in STM tasks (e.g., Hockley, 1984). (Jonides et al., 2008, p. 204)

## 1.2. Alternatives to high-speed memory scanning

As suggested above, and reviewed by McNicol and Stewart (1980, p. 256), the major alternatives to SES fall into two classes: content-addressable (direct-access) accounts, such as that of Baddeley and Ecob (1973, Model I), in which the response to a probe is determined by evaluating the strength (or level of activation) of its representation in memory relative to a criterion that varies with  $n_{\text{pos}}$ ; and parallel comparison accounts, such as that of Ratcliff (1978), in which the degrees to which the probe is similar to representations of P-set members are assessed in parallel. In his thorough analysis of the alternatives that had been proposed until then, Monsell (1978) also includes “response-association” models, such as that of Theios, Smith, Haviland, Traupmann, and Moy (1973). The alternatives also include the hybrid model of Atkinson and Juola (1974), discussed in Section 2.1. (See also Sections 5, 6, and 7.4 of Sternberg, 1975, for discussion of alternatives.)

## 1.3. Brain function and the plausibility of high-speed scanning

One contribution to scepticism about the claim of a rapid serial search process may have been ideas about the brain that were prevalent in the 1960s and 1970s. In 1973, James Anderson wrote:

It is difficult to think of cortical mechanisms capable of searches as fast as 35 milliseconds/item if we take the view that items are “taken out” of memory and “compared” with other items. This would seem to make a great many operations for a nervous system composed of what are really rather slow components. Time constants of postsynaptic events in cortex are, for the fastest excitatory postsynaptic potentials, on the order of several milliseconds, and, for inhibitory postsynaptic potentials, may be on the order of tens of milliseconds, or even more ... structures with high temporal resolution often seem to be quite

specialized for these purposes, probably because such operations are difficult to achieve with neural elements. (p. 420)

Other possible sources of implausibility include the hypothesized process being unavailable to introspection, a preference for the idea that the same mechanism underlies retrieval from active memory (AM) and long-term memory (LTM), and popularity of the idea that the brain functions as a parallel-processing machine (e.g., Anderson, Silverstein, Ritz, & Jones, 1977; Feldman & Ballard, 1982).

By the 1990s, however, considerable interest in brain oscillations in the range (20–100 Hz) had developed, and Crick and Koch (1990), Horn and Usher (1992), and Lisman and Idiart (1995) suggested that such oscillations might reflect the operation of a limited-capacity dynamic AM in which each cycle in a series of cycles corresponds to activation of the neural representation of one of the items in a memorized sequence. As discussed in Section 3, Jensen and Lisman (1998) developed this idea further and suggested that models based on it could account for “memory scanning” data. Buzsaki (2006, p. 115) wrote: “Each oscillatory cycle is a temporal processing window, signalling the beginning and termination of the encoded or transferred messages. ... In other words, the brain does not operate continuously but discontinuously, using temporal packages or quanta.”<sup>2</sup>

Also adding to the plausibility of a sequential process are recent discoveries of other neurophysiological evidence for sequential operations (e.g., Anderson, Zhang, Borst, & Walsh, *in press*; King & Dehaene, 2014; Schall, 2003; Schall, Purcell, Heitz, Logan, & Palmeri, 2011; Sigman & Dehaene, 2008).

Despite their antiquity, issues related to whether high-speed memory scanning occurs, and how to measure it if it does, discussed in the present paper, are important because of its increasing use in understanding human brain function (see, e.g., Roux & Uhlhaas, 2014) and its clinical applications. (For examples of applications in research on multiple sclerosis and Parkinson’s disease, see Archibald & Fisk, 2000; Ramsay et al., 1990;

<sup>2</sup>In an alternative approach, Amit, Sagi, and Usher (1990) developed a neural network model with three attractor subnetworks that can generate the speed, linearity, and exhaustiveness of SES.

Rao, St. Aubin-Faubert, & Leo, 1989; and Drew, Starkey, & Isler, 2009.)

#### 1.4. This paper

Section 2: There has been controversy about the data as well as their interpretation. One important reason is that the results (and the processes they reflect) vary, depending on details of the experimental procedure, details whose importance has not been sufficiently appreciated. I describe six different tasks of memory interrogation and show systematic differences in the findings produced by the most popular ones.

Section 3: I describe a neurophysiological model for memory scanning, along with tests of the model and possible modifications suggested by measurements of the rat brain.

Section 4: The finding of a linear relationship between immediate memory span and the memory scanning rate, across classes of items, has been replicated several times. The neurophysiological model described in Section 3 can explain it, and can also explain how there might be no correlation of span and rate across subjects.

Sections 5 and 6: It has been claimed that some predictions from the SES model are violated. Some of the claimed predictions—about the effects of  $n_{\text{pos}}$  on the RT variance (Section 5) and on the shortest RT (Section 6) depend on assumptions outside the model, and, furthermore, they are not violated in either the Sv or Sf tasks.

Section 7: According to one alternative to the SES model,  $n_{\text{pos}}$  should have a minimal effect on the accessibility of the most accessible item in the P-set. A rigorous test has not yet been devised.

Section 8: Results of an experiment in which retrieval from AM is compared to retrieval from inactive memory may tell us about the process of activation and bears on other interesting issues.

Section 9: Effects on RT of the recency of an N-probe is an important indicator of the use of memory strength in responding. Any such effects are absent in two experiments using the Sf task.

Section 10: In plausible interpretations of evidence-accumulation models,  $n_{\text{pos}}$  influences the rate of accumulation while the probability of a

response influences the amount of evidence required (the decision criterion). One implication is that  $n_{\text{pos}}$  and  $Pr\{pos\}$  should interact negatively. Findings from a factorial experiment violate the prediction.

Section 11: Remarkably, there is little evidence that permits us to decide that the SES model describes a process of serial comparison, rather than merely one of serial activation.

Section 12: Despite the large number of relevant studies since 1975, many interesting questions remain, some of which are mentioned, that can be addressed by new experiments.

Appendix A defines the abbreviations for 23 of the experiments mentioned in this paper, as well as providing numbers of the sections in which each experiment is mentioned.

Appendix B provides evidence about the importance of motivating subjects to perform well, by providing feedback and performance-dependent rewards, and includes other suggestions about how to investigate high-speed scanning.

Appendix C describes my analysis of the Atallah–Scanziani recordings of gamma oscillations in the rat hippocampus.

Appendix D describes procedure variants in which there are pronounced sequential and frequency effects, and the evidence for such effects in the Sf task.

Appendix E discusses the effect of requiring recall after the speeded response in the Sv task.

Appendix F describes sources of the information in Tables 1, 2, and 4.

Some of the argument depends on assuming that the same process is elicited in Sv and Sf tasks. Support for this assumption is provided in Sections 2.2, 2.4, 5.2, and 6.2. Other sections show that data from the Sf task are incompatible with alternatives to SES according to which effects of  $n_{\text{pos}}$  are due to its influence on the memory strengths of probes (Sections 6, 8, and 9) or due to its influence on the rate of evidence accumulation (Section 10).

## 2. SIX TASKS OF MEMORY INTERROGATION

When some aspect of an experimental procedure is changed, is this another way to study the same

process, or does it elicit a different process? Or, if the subject has a choice of strategy (conscious or not), do the details of procedure influence the subject's choice? And how might we know which of these is the case?

## 2.1. Sv and Sf tasks

The procedure difference between the two experiments reported in Sternberg (1966), one using varied sets (the Sv task) and the other fixed sets (Sf task), had little effect on the pattern of mean RTs:  $\overline{RT}$ s on P-trials and N-trials in both experiments increased approximately linearly, with approximately equal slopes of about 38 ms per item, and with zero intercepts of less than 400 ms. Features of the Sv task include: a small stimulus ensemble; sequential presentation of the P-set at a slow rate (about 1 s/item); a different P-set, possibly different in size, on each trial; a long delay between P-set and probe (about 2 s); a warning signal before the probe; and paid subjects, with additional monetary incentives based on speed and accuracy, with a high cost of errors.<sup>3</sup>

In the Sf task, presentation of the P-set is followed by a series of perhaps 50 to 100 trials (divided into blocks of 20 to 40) on each of which a probe is presented, calling for a positive or negative response. Because it is important to avoid a consistent mapping of stimuli to responses and the learning of associations that it produces (Kristofferson, 1972a; Schneider & Shiffrin, 1977), some or all of the P-set members in one condition become N-set members in the next, and nesting of different P-sets for the same subject is avoided. Also, to promote the independence of responses from one trial to the next, the response–stimulus (R–S) interval is long (e.g., 4 s or more), long enough to provide feedback on at least the correctness of the response, a delay, and a warning signal. Kristofferson (1972b) has shown that under such conditions the main

features of the data (linearly increasing functions with  $\beta_{\text{pos}} \approx \beta_{\text{neg}} \approx 36$  ms/item when the items are digits) remain invariant across more than 4000 trials of practice.<sup>4</sup>

Perhaps it was the similarity of the phenomena generated by these two procedures, despite their differences, that encouraged the belief that other procedural variants would also elicit the same process. On the other hand, the work of Atkinson and his collaborators (Atkinson, Herrmann, & Wescourt, 1974; Atkinson & Juola, 1974; Banks & Atkinson, 1974; see Sternberg, 1975, Section 7.4) have made it clear that alternative strategies are available—in particular, discrimination of the strength or level of activation of the probe's representation in memory (its familiarity), which can be accessed directly, rather than interrogation of AM itself. I shall call this “strength discrimination”. And Gaffan (1977) has argued (see also Browning, Baxter, & Gaffan, 2013) that familiarity discrimination and P-set interrogation depend on different brain mechanisms, based partly on the finding that the lesions associated with amnesia impair only the former.

According to the hybrid model developed by Atkinson and his collaborators, the first process after a probe is encoded is a stage whose duration is independent of  $n_{\text{pos}}$ , in which the strength of its memory representation is evaluated and a decision made based on this evaluation. Only if the strength is neither above a high criterion (causing generation of a positive response), nor below a low criterion (causing generation of a negative response)—criteria that are independent of  $n_{\text{pos}}$ —is that stage followed by serial comparisons to members of the P-set. The RT is thus a mixture of  $RT_{\text{strength}}$  and  $RT_{\text{strength}} + RT_{\text{scan}}$ , with only  $RT_{\text{scan}}$  increasing with  $n_{\text{pos}}$ .<sup>5</sup> Whereas most of the evidence for the hybrid model has involved recognition of items from long lists, two experiments (Darley & Arabie, described by Atkinson & Juola, 1974, pp. 269–276;

<sup>3</sup>Appendix B contains evidence of the extent to which the performance of many subjects depends on providing feedback and tangible incentives.

<sup>4</sup>In Appendix D, I describe variants of the Sf task in which the R–S interval is brief, which produce substantial effects of probe recency and frequency, and I consider the available evidence for such effects when the R–S interval is long.

<sup>5</sup>As formulated, if the serial comparison process occurs, it follows the decision based on memory strength of the probe. An alternative is that the two processes occur in parallel, with  $RT_{\text{strength}} < RT_{\text{scan}}$ , so that when scanning is called for,  $RT = RT_{\text{scan}}$ . One way in which these alternatives might be discriminated is mentioned in Section 12.4.

Hockley & Corballis, 1982, Exp. 1), have shown that it can also account for the behaviour of  $\overline{RT}$  with short lists.<sup>6</sup>

Differences of memory strength among probes can be regarded as a by-product of the experimental procedure: Presentation of the P-set not only creates a representation of the set, but also temporarily increases the strengths of the memory representations of its members; rehearsal of the set is believed to do the same,<sup>7</sup> and presentation of a probe temporarily increases the strength of its memory representation. These representations do not carry information about set membership per se, but under some conditions their strengths can be used to make accurate inferences about it. Those who have argued for the importance of strength discrimination have not attempted to explain the similarity of the  $\overline{RT}$  functions produced by Sv and Sf procedures, despite what would appear to be large differences in their consequences for memory strength. Monsell (1978, pp. 496–497) seems to argue that Sf elicits a different process from Sv, because results that differ from those of Sf are produced by other procedures involving a fixed set. (Examples are preventing maintenance of a short-term representation of a well-learned P-set, or providing practice with a consistent mapping, which may permit the use of S-R associations.) But another possibility is that the similarity of results from Sv and Sf tasks means that strength discrimination or performance of S-R associations are seldom used in those tasks.

Atkinson et al. (1974) provide a useful discussion of the determinants of whether scanning occurs. They point out that the use of small stimulus ensembles implies little strength difference between N-probes and P-probes (because each item is seen more than once, sometimes as a member of the P-set, sometimes of the N-set), hence increasing the chance that scanning will be required. In contrast, with large ensembles, memory-strength differences

between P-probes (just seen, or just rehearsed) and N-probes (last seen many trials earlier, and not rehearsed) may be large, permitting scanning to be avoided. They also suggest that strength discrimination is more likely to be used under conditions that emphasize speed over accuracy (as in McElree & Doshier's cued-response task, discussed below). In a similar vein, Schneider and Shiffrin (1977, pp. 31–32) have emphasized the idea that subjects are flexible, that they may have at their disposal more than one strategy, that tasks may elicit mixtures of strategies, and that previous experience and the current mixture of conditions may influence which strategy they adopt.

Nonetheless, some investigators seem to have assumed that radically different tasks all elicit the same process, and when one of these tasks gives rise to data that are inconsistent with SES, they have asserted that the latter is wrong in general. Thus, based on data from the cued-response task (see below), Doshier and Sperling (1998, p. 242) claim that “Attractive though it may be, Sternberg’s model is incorrect”, rather than “Attractive though it may be, Sternberg’s model is incorrect for the cued-response task.”

I shall call the two most influential alternative paradigms the *Monsell task* (Monsell, 1978, Exps. Monsl78.1, Monsl78.2), also exemplified by Corballis, Kirby, and Miller (1972), Johns and Mewhort (2011), and Exps. Rcliff78, McEl89.p, McEl89.2, and Nosof11; and the *Ashby task*, (Ashby, Tein, & Balakrishnan, 1993), also exemplified by Chase and Calfee (1969), Ellis and Chase (1971), Franklin and Okada (1983), Klatzky and Smith (1972), Klatzky, Juola, and Atkinson (1971), Oberauer (2001), Nickerson (1966), Schneider and Shiffrin (1977, Exp. 2, frame size 1), Smith (1967), Exps. Ashby93, Schnei77, and Frank83, and a number of fMRI studies.<sup>8</sup> As these tasks use a new P-set presented on each trial, they should be compared with the Sv task.

<sup>6</sup>In the hybrid model, errors can arise only during the strength-discrimination stage. If this assumption is relaxed, support for the model is also provided by Banks and Atkinson (1974). The conclusion from the Hockley and Corballis experiment should be regarded as tentative, because they used a brief (500 ms) R–S interval (see Appendix D).

<sup>7</sup>But in relation to serial recall, the roles of decay, and of rehearsal in counteracting its effects, are controversial (Lewandowsky & Oberauer, 2015).

<sup>8</sup>Examples are Bunge, Ochsner, Desmond, Glover, and Gabrieli (2001) and Schneider-Garces et al. (2009).

## 2.2. Monsell task<sup>9</sup>

In experiments using this task the presentation rate is about 0.5 s/item, the probe delay is 0.3–0.5 s, and the stimulus ensemble is sometimes large. Thus, in McElree and Doshier's (1989) RT experiments, Exps. McEl89.p and McEl89.2, on which their cued-response task is based, imitating a feature of Monsell's (1978) Exp. 2 (Exp. Monsl78.2), there were two stimulus ensembles, each containing 50 words, one used on odd trials, the other on even trials. (Monsell's ensembles each contained 36 words.) Assume that memory strength of a probe declines as the number of trials,  $\Delta trials$ , between its last appearance (or rehearsal) and the current trial increases. For P-probes  $\Delta trials = 0$ , but for N-probes it depends on the size of the ensemble from which the probe is drawn and whether the ensembles alternate from trial to trial. For the RT version of the Monsell task, as implemented by McElree and Doshier (1989), simulations of the trial sequence for their "pilot study" (Exp. McEl89.p) produced a range of  $\Delta trials$  for N-probes of from 2 to 184 and a mean of 22.4. (Corresponding simulations for the Sv task with an ensemble of 10 items gave a range of  $\Delta trials$  of 1 to 18 trials and a mean of 2.5, a difference of almost an order of magnitude.) The great difference in recency between probes on P-trials and N-trials in this version of the Monsell task, compared to the Sv task, is likely to increase the utility of strength discrimination.

That a different process underlies performance in Monsell and Sv tasks is suggested by the data in Table 1, which compares five experiments using Sv or Sf tasks with seven experiments using the Monsell task or a variant. Among the Sv and Sf tasks, Hockley's (1984, Exp. Hock84) subjects were slower and more error prone, perhaps because they were given neither incentives nor feedback about speed and/or they were required to switch between memory search and visual search tasks every 40 trials. Their data were pooled over

all trials in the memory search task, neglecting the possibility that the data from trials immediately after a task switch were unusual. Also, data were pooled over all 11 sessions, without regard for possible practice effects. Perhaps these are also reasons for the unusually high error rates (Table 1) and variances (Section 5.2).

The Monsell task produces flatter  $\overline{RT}$  functions and greater zero intercepts than the Sv or Sf tasks (with the exception of Exp. Hock84, which has unusual features; see above), which produce similar values. Error rates are higher in the Monsell task, despite, in some cases, larger amounts of practice. These properties are those interpreted by Atkinson and Juola (1974) as resulting from a large proportion of trials on which the mechanism is strength discrimination of the probe, rather than search of the P-set. Also, as shown in the  $t_3$  column,  $\overline{RT}$ 's tend to be greater, again despite additional practice. Monsell (1978, p. 496) and Diener (1988, p. 375) have mentioned another possibility for why a scanning process might not be used in the Monsell task—that too little time is available in that task to form a representation of the P-set that can be scanned. If such a representation is never formed for a P-set consisting of complex pictures, then this conjecture is supported by Gaffan's (1977) important experiments using the Sv task to compare P-sets of such pictures with P-sets of disyllabic words. Even with a very slow presentation rate (4.5 s/item) and a very long (6 s) probe delay, the decisions about the pictures, but not about the words, appear to have been based on strength discrimination: For example, only the serial-position functions of the former showed a facilitating effect of recency on P-trials.

Two other features of the data distinguish the Monsell task and should be regarded as diagnostic of strength discrimination: One, reported by Gaffan (1977, Exp. 2) and Monsell (1978), is the "N-probe recency effect": N-trials with probes that have been seen recently elicit longer RTs.<sup>10</sup>

<sup>9</sup>Stephen Monsell was the first investigator to use this task with an adequate number of subjects under conditions that produced an acceptably low error rate. His experiments, but not the others described below, included instructions to subjects not to rehearse the P-set.

<sup>10</sup>Using the Sv task, Gaffan (1977, Exp. 2) found a strong N-probe recency effect averaging about 150 ms with P-sets of complex pictures, but no such effect with P-sets of words. In Exp. McEl89.2, using the Monsell task with  $n_{pos} = 3, 4, 5,$  and 6 words, the effect



**Table 1.** Comparison of Sv and Sf tasks with Monsell task.

Task	Experiment	Items	MnTrls	N <sub>sub</sub>	n <sub>pos</sub>	Z-Intercept	Slope	t <sub>3</sub>	Mean Percent Error (n <sub>pos</sub> )					
									(1)	(2)	(3)	(4)	(5)	(6)
Sf	Stern66.2	d	300	6	1,2,4	369	38.3	484	0.4	1.4		1.8		
Sf	Stern67int	d	180	12	1,2,4	351	36.5	461	0.6	1.7		1.4		
Sf	Stern69b.4eq	d	230	12	1,2,4	378	35.9	486	1.3	1.0		1.3		
Sv	Stern66.1	d	96	8	1-6	397	37.9	511	0.0	0.5	2.1	2.6	1.0	1.6
Sv	Hock84	c	1760	6	3-6	512	45.2	648			2.3	2.7	4.3	7.4
M	Nosof11	c	2375	4	1-5	381	29.2	469	1.8	2.3	2.7	4.2	5.9	
M	Monsl78.1	c	834	8	1-4	430	28.7	516	1.7	3.1	3.6	6.2		
M	Monsl78.2	w	924	8	2-5	436	23.7	507		1.4	2.7	4.6	7.4	
M	Rcliff78	d	2400	2	3-5	559	20.0	619			1.7	1.8	2.2	
M	McEl89.p	w	?	10	3-5	613	21.3	677			6.0	7.0	7.5	
M	McEl89.2	w	298	18	3-6	633	18.9	690			6.5	10.3	15.5	17.5
M?	Jacbs06	c	576	18	2,4,6	571	27.3	653		2.7		5.2		10.3

Note: Tasks: Sf = Sternberg, fixed set; Sv = Sternberg, varied set; M = Monsell; M? = Monsell variant.

Items: d = digits, c = consonants, w = two-syllable words. MnTrls = Average number of test trials per subject (not available for Exp. McEl89.p). N<sub>sub</sub> = number of subjects. n<sub>pos</sub> = size of P-set. Z-icpt = zero intercept of mean  $\overline{RT}$  function. Slope = slope of mean  $\overline{RT}$  function (mean of  $\beta_{pos}$  and  $\beta_{neg}$ , ms/item). t<sub>3</sub> = Z-icpt + (3 × slope): Estimated  $\overline{RT}$  for n<sub>pos</sub> = 3, a measure of  $\overline{RT}$  that facilitates comparison of experiments (ms). See Appendix G for sources.

If this effect occurred in the Sf task, then we would expect that as n<sub>neg</sub> is reduced, with n<sub>pos</sub> fixed, the resulting increase in N-probe recency would cause  $\overline{RT}_{neg}$  to increase. But, as discussed in Section 9, this has not been found. The other feature is the existence of pronounced serial-position effects on P-trials, mainly characterized by recency effects (an RT advantage when more recently presented items are probed).

Donkin and Nosofsky (2012) provide an additional source of evidence for the difference between the processes that underlie Monsell and Sv tasks. They found that data from the Monsell task (Exp. Nosof11) but not the Sv task (Exp. Donk12)

could reject the SES model in favour of parallel self-terminating and global familiarity models.<sup>11</sup>

It is possible that the critical distinction between Sv and Monsell tasks is the probe delay and not the presentation rate. In experiments with varied probe delay, with slow presentation rate (Clifton & Birenbaum, 1970; probe delay from 0.8 s to 4.8 s, but with strategy differences across subjects) and fast presentation rate (Forrin & Cunningham, 1973; probe delay from 0.5 s to 3.5 s, but with very high error rates), the magnitude of the effect of serial position on P-trial RTs has been found to decrease as the probe delay is lengthened.<sup>12</sup> Also, whereas the presentation rate in Exp. Jacobs06 (see Table 1) was

increased slightly with n<sub>pos</sub>, and averaged about 50 ms: Equations of linear functions fitted to  $\overline{RT}$  for P-probes, distant N-probes, and recent N-probes,  $587 + 19.8 n_{pos}$ ,  $670 + 13.6 n_{pos}$ , and  $687 + 22.3 n_{pos}$ , respectively. Mean error rates for these probe types were 13%, 5%, and 19%, respectively. Results of Monsell's (1978) corresponding experiments, in which the effect also increased with n<sub>pos</sub> and averaged about 25 ms, were qualitatively similar, but slopes of the fitted linear functions were substantially greater, zero-intercepts substantially smaller, and error rates much lower.

<sup>11</sup>In considering the results of Exp. Donk12 it is worth noting that although subjects worked for ten sessions, they received feedback about the accuracy but not the speed of their responses.

<sup>12</sup>One exception is the study by Burrows and Okada (1971) who compared the Sv task and the Monsell task within subjects, with n<sub>pos</sub>-values of 1, 2, 3, and 4, and found surprisingly small and non-significant differences between the magnitudes of serial-position effects between tasks. However, their subjects, who received no feedback, were highly error-prone (5.7% in the Sv task, compared to 1.3% for the same n<sub>pos</sub> values in Exp. Stern66.1), with a slower  $\overline{RT}$  = 607 ms (compared to 492 ms in Exp. Stern66.1), and with a mean slope of 23.7 ms/digit, more characteristic of the Monsell task, so they may not have been performing optimally.

slow, as in the Sv task, the probe delay was only about 250 ms.<sup>13</sup> And the data resemble those from the Monsell task more closely than those from the Sv task.

### 2.3. Cued-response Monsell task

As mentioned above, in McElree and Doshier's version of the Monsell task (1989, Exps. McEl89.p and McEl89.2), the mean recency difference between P-probes and N-probes was very large, increasing the utility of strength discrimination relative to the Sv task. In their variant of this task (1989, Exps. 1 and 3), subjects are forced to respond, on presentation of a tone-burst cue, at various times before processing is complete.<sup>14</sup>

### 2.4. Ashby task<sup>15</sup>

In this task, the P-set is presented simultaneously, as a row of items. This invites (but does not require) the subject to represent the set as a visual image, which I believe is unlikely when items are presented sequentially in one location.<sup>16</sup> What might we expect on those trials on which the search is of an image? If the search were serial and self-terminating, we would expect  $(\beta_{\text{neg}}/\beta_{\text{pos}}) \approx 2$ ; if the search were from left to right, we would expect monotonically increasing serial-position curves.<sup>17</sup> If the search

were a mixture of some trials on which search was exhaustive, and others on which it was self-terminating, we would expect  $1 < (\beta_{\text{neg}}/\beta_{\text{pos}}) < 2$ .

Reports of search of the display or image of a row of characters are in conflict. Nickerson (1966), Sternberg (1967a), and Hockley (1984) concluded that such search is serial and self-terminating. Atkinson, Holmgren, and Juola (1969) disputed this, as did Townsend and Roos (1973). Van Zandt and Townsend (1993), in a thorough review, argue that search in neither visual displays nor memory is exhaustive, but, referring to whether  $\beta_{\text{neg}}$  is greater than  $\beta_{\text{pos}}$ , they report that "Memory search most often yields parallel set-size functions, whereas visual search is more likely to yield significant slope differences" (p. 567). Data from Group 3 of Briggs and Johnsen (1972), who used a fixed-set procedure<sup>18</sup> to study display search with 12 subjects, support self-termination, with  $\beta_{\text{pos}} = 22.7 \pm 2.5$  ms/letter,  $\beta_{\text{neg}} = 39.9 \pm 3.1$  ms/letter, and a slope ratio of  $1.82 \pm 0.24$ .<sup>19</sup> On the other hand, the slopes of the display-search (frame size 1) data in Schneider and Shiffrin's (1977) varied set procedure, with four subjects experienced with other tasks, are small and approximately equal, with  $\beta_{\text{pos}} = 16.7$  ms/letter and  $\beta_{\text{neg}} = 15.1$  ms/letter.

Results from Sv and Sf tasks are compared to those from the Ashby task in Table 2.<sup>20</sup> They

<sup>13</sup>Also, for  $n_{\text{pos}} < 6$ , subjects did not know  $n_{\text{pos}}$  until the probe appeared.

<sup>14</sup>McElree and Doshier (1989) used their cued-response task, rather than the Meyer, Irwin, Osman, and Kounios (1988) "speed-accuracy decomposition" task, in which the process is permitted to go to completion on a random subset of trials (rather than the response being cued on all trials), which permits testing whether the cued-response procedure is eliciting the same process as in a standard RT experiment.

<sup>15</sup>The most thorough analysis of data from this task is provided by Ashby et al. (1993).

<sup>16</sup>The figure that Townsend and Ashby use to illustrate the "memory scanning task" (1983, Figure 6.1) actually shows the Ashby task. However, Ashby et al. (1993, p. 543) comment that "the use of simultaneous presentations makes the memory scanning task more similar to visual search (Atkinson et al., 1969) and in visual search it is thought that subjects search through a visual short-term store (Townsend & Roos, 1973)". Because the Ashby task does not *require* the subject to use the visual array representation, what the subject actually chooses to do may depend on previous experience and level of practice, and we may find substantial individual differences. Whether or not a subject is using a visual representation may be revealed by the RT pattern (see Table 2), but mixed strategies present problems for such inference.

<sup>17</sup>If responses are made by left and right hands, such a pattern might be complicated by the Simon effect (Kornblum, Stevens, Whipple, & Requin, 1999; Lu & Proctor, 1995), which, e.g., would shorten RTs when the target is further to the right if the positive response is made by the right hand.

<sup>18</sup>The memory sets were fixed for 48 trials, but they called this a "varied-set" procedure.

<sup>19</sup>Standard errors are based on differences across four sessions

<sup>20</sup>Ashby et al. (1993, Exp. Ashby93) displayed the memory set for  $n_{\text{pos}}$  s. Schneider and Shiffrin (1977, Exp. Schnei77) displayed it for as long as the subject wished on each trial. Franklin and Okada (1983, Exp. Frank83) displayed it for 150 ms, and used a probe delay of only 500 ms.

Table 2. *Sf and Sv tasks versus Ashby task.*

<i>Task</i>	<i>Exp</i>	$n_{pos}$	$N_{sub}$	$\beta_{neg}$ (ms/item)	$\beta_{pos}$ (ms/item)	$\beta_{neg}/\beta_{pos}$	$\beta_{neg} - \beta_{pos}$ (ms/item)
Sv	Stern66.1	1–6	8	$33.2 \pm 3.7$	$42.3 \pm 4.3$	$0.79 \pm 0.06$	$-9.1 \pm 2.4$
Sv	Stern66.1	2–6	8	$35.4 \pm 4.2$	$38.2 \pm 5.8$	$1.09 \pm 0.25$	$-2.8 \pm 4.0$
Sf	Stern66.2	1,2,4	6	$38.0 \pm 6.3$	$39.6 \pm 6.1$	$0.98 \pm 0.10$	$-1.6 \pm 3.0$
Sf	Stern67int	1,2,4	12	$37.7 \pm 2.8$	$32.9 \pm 2.5$	$1.22 \pm 0.14$	$+4.8 \pm 3.5$
Sf	Stern69b.4eq	1,2,4	12	$35.8 \pm 3.0$	$36.0 \pm 4.9$	$1.11 \pm 0.10$	$-0.2 \pm 2.9$
Ashby	Ashby93	2–5	4	$59.2 \pm 6.9$	$36.6 \pm 3.5$	$1.62 \pm 0.16$	$+22.4 \pm 5.6$
Ashby	Schnei77	1,2,4	4	$52.6 \pm 10.0$	$37.6 \pm 8.3$	$1.43 \pm 0.09$	$+15.0 \pm 3.7$
Ashby	Frank83	2–5	24	50.0	32.5	1.56	+17.5

*Note.* Stimulus ensembles consisted of the 10 digits, except for Exps. Ashby93 (11 consonants) and Schnei77 (9 digits for two subjects, 9 consonants for two subjects). See Appendix F for sources.

show that whereas  $\beta_{neg} \approx \beta_{pos}$  in the former,  $\beta_{neg} > \beta_{pos}$  in the latter, with a ratio that is greater than 1:1, but less than 2:1, consistent with a mixture of exhaustive and self-terminating search.<sup>21</sup>

However, given plausible added assumptions, such a mixture would lead us to expect the RT variance to grow more rapidly with  $n_{pos}$  for P-trials than for N-trials—the opposite of what was observed (Ashby et al., 1993, Figure 6). Like the pattern shown in Table 1 of Sternberg (1975), the unequal slopes are associated with a higher value of  $\beta_{neg}$ , suggesting a slower search rate (if the search is sequential).

The idea that AM for a spatial array and for a sequence may be represented differently has a long history (e.g., Fougny, Zughni, Godwin, & Marois, 2015; Scarborough, 1972) and is supported by brain measurements: According to Roux and Uhlhaas (2014, p. 21–22), “... theta activity occurs preferentially in tasks that involve sequential coding of multiple [working-memory] items, such as during the Sternberg paradigm ... whereas alpha oscillations tend to occur during tasks that require maintenance of simultaneously presented visual or spatial information.” It is plausible that if representations differ, then the processes used to interrogate those representations might also differ, as indicated by the data in Table 2.

## 2.5. Novel negatives task

Huesmann and Woocher (1976) and Roeber and Kaernbach (2004), as well as the Wickens, Moody, and Dow (1981) and Wickens, Moody, and Vidulich (1985) experiments with words, used a variant of the Sv task in which the stimulus ensemble, and hence the recency of N-probes, was effectively infinite: No N-probe occurred more than once in the experiment, while P-probes occurred once as set members and once as probes. Banks and Atkinson (1974) compared this condition with the standard Sv task and found differences, such as a reduction in  $\bar{\beta}$  from 45 to 27 ms/item (in the accuracy conditions), consistent with the greater use of strength discrimination, as in the hybrid model (Section 2.1). Using fixed sets, but with an R–S interval of only 500 ms (see Appendix D), Hockley and Corballis (1982, Exp. 1) made the same comparison, with similar findings: For  $n_{pos} \leq 6$ ,  $\bar{\beta}$  was reduced from 38 to 20 ms/item, again consistent with the hybrid model.

## 2.6. Conclusion: Procedure details matter

Subjects are flexible; details such as timing, emphasis on speed versus accuracy, availability of more than one internal representation of information, and previous experience may influence which

<sup>21</sup>Slope ratios in Exp. Ashby93 (1.33, 1.36, 1.88, and 1.90) differed markedly across the four subjects, with two subjects having values close to 2.0.

process or mixture of processes is selected to deal with the challenges in an experiment. Thus, while we might hope that varying the procedure that gave rise to a hypothesized process may be a new way to study that process, it may, instead, cause subjects to use a different process, or a mixture of processes such as the one embodied in the Atkinson and Juola (1974) hybrid model.

### 3. THE LISMAN–IDIART–JENSEN (LIJ) NEUROPHYSIOLOGICAL MODEL FOR MEMORY SCANNING AND MEMORY SPAN

#### 3.1. Nested brain oscillations

As mentioned in Section 1.3, new ideas about brain function have increased the plausibility of a rapid sequential process such as SES. In this section, we consider a proposed neural process that might underlie serial exhaustive scanning.

Lisman and Idiart (1995) suggested that the memory of a series of items could be maintained by a two-level hierarchy of brain oscillations, with each cycle of the faster gamma oscillation (20–100 Hz) corresponding to the (re)activation of the neurons in a cell assembly that represents one of the items, and with the sequence of faster cycles nested within each cycle of the slower theta oscillation (5–8 Hz).<sup>22</sup> What I shall call the “LIJ model” of a time-compressed dynamic memory<sup>23</sup> was further developed by Jensen and Lisman (1996a, 1996b, 1998), who suggested that such a process could maintain the memory of short lists of either known items (with representations in LTM) or novel items, and could also drive an exhaustive process of sequential comparison. (See also Jensen & Lisman, 2005, and Lisman & Jensen, 2013, and

the references therein). Because each item’s representation is activated at a different phase of the theta oscillation (“cross-frequency coupling”), such a process could represent the order of the items, as well as their identities (e.g., Siegel, Warden, & Miller, 2009). The capacity of this buffer memory—the maximum number of items that it can contain, corresponding to the memory span—is the number of gamma periods that fit, along with a fixed amount of “dead time”, within a theta period.

Two variants of the LIJ model were offered to explain performance when  $n_{\text{pos}}$  is less than the capacity: In one variant (*adapting-theta*), the theta period is shorter when fewer items are being maintained. In the other variant (*resetting-theta*), the theta phase is reset to zero once the probe has been identified and any ongoing activation sequence has been completed (which contributes a variable increment to the RT). Jensen and Tesche (2002) used magnetoencephalography to examine theta oscillation during an Sv task and found that whereas theta power increased with  $n_{\text{pos}}$ , theta frequency did not change, evidence against the *adapting-theta* variant.<sup>24</sup> Also, using intracranial recording and an Sv task with letters, Rizzuto et al. (2003) found evidence for phase resetting, especially upon probe presentation. To explain the exhaustiveness of the search, Jensen and Lisman (1998) assumed that responses could occur only at the trough of the theta cycle. (They did not discuss the possibility, considered in Section 11, that what is serial is an activation process that is followed by simultaneous comparisons, and that therefore must be exhaustive.)

In recent years, the idea that brain oscillations underlie the maintenance of information in working memory has received increasing support.<sup>25</sup> A critical feature of the LIJ model is the idea that the sequential maintenance process also

<sup>22</sup>A similar proposal was made by Horn and Usher (1992).

<sup>23</sup>“Time compressed” because the rate of item activations in the memory might be as much as 50 times greater than the rate at which the items were presented; “dynamic” because the memory is maintained by a repeating sequential activation process.

<sup>24</sup>However, evidence supporting theta adaptation in the rat is provided by Geisler et al. (2010).

<sup>25</sup>See, for example, Sederberg, Kahana, Howard, Donner, and Madsen (2003), Siegel et al. (2009), Axmacher et al. (2010), Fuentemilla, Penny, Cashdollar, Bunzeck, and Duzel (2010), Kawasaki, Kitajo, and Yamaguchi (2014), Lisman and Jensen (2013), Lisman (2010), Roux and Uhlhaas (2014), and references therein. See also Schon, Newmark, Ross, & Stern (2016), who found activity in the parahippocampal region that increased with  $n_{\text{pos}}$  during the probe delay, but who unfortunately used the novel negatives task.

plays an important role in information retrieval, and that “readout” from the dynamic, time-compressed buffer, and comparison of its contents, can occur at the same rate as the maintenance process (and does so in high-speed memory scanning).<sup>26</sup>

Several elaborations of the model would be needed for a complete and persuasive account. First, a comparison process or processes would have to be specified by which members of the P-set are compared to the probe. If the current version of the model is to explain the RT data, then, irrespective of whether such a comparison process is serial, parallel, or overlapping, any delay that it adds to the RT would have to be independent of  $n_{\text{pos}}$ . (See Section 11 for a discussion of serial comparison versus serial activation, and why it would be critical for the model whether and how the RT is influenced by the similarity of an N-probe to one or more members of the P-set.) Second, one would need evidence that supports the idea, assumed to explain the exhaustiveness of the search, that motor responses can occur only at the trough of the theta cycle.<sup>27</sup> Third, we would have to understand how the time-compressed memory could support slower sequential processes, such as ordered recall of the sequence of items, or the substantially slower scanning processes that are employed when sequence or order information must be retrieved from short lists.<sup>28,29</sup>

### 3.2. The “skip” feature of the LIJ model

In both variants of the LIJ model, the probability,  $p_a$ , that a single scan produces a response can be less

than 1.0; if there is no response, the scan is repeated. Because  $p_a$  (which is not influenced by  $n_{\text{pos}}$ ) does not change with repeated scans, the number of scans required for a response has a geometric distribution. This “skip” feature contributes to the increases, with  $n_{\text{pos}}$ , of the variance and higher moments of the RT. Without this feature, neither variant of the model can produce a third central moment that increases with  $n_{\text{pos}}$ , as observed. Jensen and Lisman (1998) fitted their models to two sets of data: Mean values over the 12 subjects of the first three central moments for N-trials in Exp. Stern67int1 (shown in Figure 1, panels H1, H2, and H3), which used the Sf task, and RT distributions pooled over P-trials and N-trials for each of the four subjects in Exp. Ashby93.<sup>30</sup> For the latter data set, the estimates of  $p_a$  for the resetting-theta variant of the model range from .20 to .69 over subjects. The value  $p_a = .20$  corresponds to a mean of five scans before the response occurs and hence a slope of the  $\overline{RT}$  function of five times the gamma period. The large individual differences in the estimates of  $p_a$ , together with the fact that it is influenced by neither  $n_{\text{pos}}$  nor the number of scans already executed, reduce the plausibility of the model for these data. For the former data set,  $\hat{p}_a = 0.78$ , which corresponds to a mean of 1.3 scans per trial, a more plausible value.

One of the surprising findings about memory scanning shown by Kristofferson (1972b) is that the rate does not increase with practice.<sup>31</sup> In Sternberg (1975) I suggested that this could be a result of the process being practised in everyday life, and so the rate is at asymptote when subjects enter the laboratory. The LIJ model provides an

<sup>26</sup>The study by Zarahn, Rakitin, Abela, Flynn, and Stern (2006), in which it is concluded that maintenance and search involve different brain regions, used the Ashby task and reported the mean slope of the  $\overline{RT}$  function to be 61 ms/letter, substantially greater than what is found for Sv or Sf tasks. It is important to ask the same question of those tasks.

<sup>27</sup>It is hard to believe that execution of an unrelated response must await the trough. If not, why must this response do so?

<sup>28</sup>Examples are recency judgements (Hacker, 1980; McElree & Doshier, 1993; Sternberg, 1969a, Exp. 8), and context recall (Sternberg, 1967b; 1969a, Exps. 6, 7).

<sup>29</sup>Other challenges to the current version of the model, in particular, the exhaustiveness of the process, include the partial selectivity of search of categorized word lists (Naus, 1974; Naus, Glucksberg, & Ornstein, 1972; see Sternberg, 1975, Section 7.2) and the fact that in a procedure that usually elicits fast exhaustive search, the process for some subjects is slower and self-terminating (see Sternberg, 1975, Section 7.3).

<sup>30</sup>As discussed in Section 2.4, because Exp. Ashby93 used simultaneous visual presentation of the P-set, and also because of properties of the mean data, it can be argued that the process underlying those data differ from that underlying the Sf or Sv tasks.

<sup>31</sup>See Sternberg (1998), Fig. 14.15, for additional evidence.

alternative: If the scanning process is driven by a gamma oscillation, which serves other functions that include maintenance of information in AM, then prior search practice need not be postulated. However, if something does have an effect on the scanning rate, then the LIJ model suggests that we may find a corresponding change in the gamma period.<sup>32</sup> Multiple sclerosis (MS) provides an example of an effect on the scanning rate: Using an Sv procedure, Archibald and Fisk (2000) found that, relative to controls, the scanning rate for their subjects with MS is 80% slower than for normal, and Rao, St. Aubin-Faubert, and Leo (1989) found it to be 47% slower.<sup>33</sup> Other studies (e.g., Cover et al., 2006; Nistico et al., 2013) have shown that MS alters oscillatory patterns, but I have found no reports of increased gamma period.

### 3.3. Tests of the LIJ model

Supporting the model, Kaminski, Brzezicka, and Wrobel (2011) found that individual differences in memory span vary with the “theta–gamma ratio”—the number of gamma periods (which ranged from 21 to 39 ms) that fit into a theta period (which ranged from 118 to 222 ms). However, doubts have been expressed about this conclusion.<sup>34</sup> In an attempt to test the account that the LIJ model provides of the memory span, Vosskuhl, Huster, and Herrmann (2015) determined each subject’s theta frequency that

was most strongly coupled to oscillation in the gamma range, and attempted to drive it at a lower frequency (chosen to increase the forward digit span by one) by delivering alternating current stimulation (tACS) to the scalp at that lower frequency. They succeeded in increasing the digit span during stimulation, supporting the LIJ model.<sup>35</sup>

In a daring attempt to test the LIJ model by modulating the gamma frequency  $f_\gamma$ , and hence the scanning rate, Burle and Bonnet (2000) delivered click trains in a procedure pioneered by Treisman, Faulkner, Naish, and Brogan (1990), while subjects performed the Sf task with  $n_{\text{pos}} = 1, 2, \text{ and } 4$  digits. For an interval of about 500 ms preceding the probe, subjects heard a train of clicks of frequency  $f_c$  at or near a subharmonic (half) of the assumed gamma frequency. Other findings from this procedure indicate that when twice the click frequency  $2f_c$  is only slightly above or below  $f_\gamma$ , the latter changes in the direction of  $2f_c$ . (Thus, the point at which the effect changes sign provides an estimate  $\hat{f}_\gamma$  of the normal  $f_\gamma$ .) Analysis of the RT data showed that the scanning rate did indeed change as expected: The effects on  $\overline{RT}$  of the click train tended to increase with  $n_{\text{pos}}$ . Also,  $\bar{\beta}$  (36 ms/digit) was less than  $\hat{f}_\gamma$  (about 41 ms/digit), as expected if skipping occurred on some trials (but not if there is no skipping; see Section 3.4). If this finding were persuasive, it would be strong evidence for the seriality of the SES model, as well as support for the relation,

<sup>32</sup>An alternative is an effect on  $p_n$ , if that feature of the model is retained.

<sup>33</sup>Given the large effects that feedback and reward can have on performance in these tasks, described in Appendix B, there may be special concerns about differences in motivation between clinical populations and their controls. Also, the possibility of strategy differences being responsible for slope effects cannot be overlooked. As discussed in Sternberg (1975, Section 7.3), greater mean slopes are sometimes associated with  $\beta_{\text{neg}} \approx 2\beta_{\text{pos}}$ , suggesting a self-terminating search strategy. Both of these reports of slower scanning rates by subjects with MS claim no significant interaction of slope with response type, but neither of them provides the data separately for P-trials and N-trials.

<sup>34</sup>“... determinations of oscillation frequencies were very noise sensitive, raising doubts about the conclusion. Rigorous testing of this relationship will require resolution of the controversy about which brain regions are responsible for short-term memory maintenance and better methods for noninvasive measurement of the oscillatory frequencies at those locations.” (Lisman & Jensen, 2013, p. 1005)

<sup>35</sup>That they actually lowered the ongoing theta frequency was inferred indirectly, as they were not able to measure it during the stimulation. They assumed that their stimulation would not change the relevant gamma frequency. In an improved version of such an experiment, theta and gamma would be measured and an attempt would also be made to drive theta at a higher frequency, as well, chosen to *decrease* the span.

embodied in the LIJ model, between the gamma oscillation and the scanning process.<sup>36</sup>

The most promising test thus far was recently carried out by Bahramisharif, Jensen, Jacobs, and Lisman (2016), who analysed intracranial recordings while subjects performed an Sv task with  $n_{\text{pos}} = 3$  letters. During the 2 s probe delay, while the letter sequence was being maintained in memory, they recorded activity at brain sites “tuned” to one of the letters (Jacobs & Kahana, 2009), for sequences in which that letter occupied the first, second, and third serial position. They found that the theta phase at which gamma power was maximum at those sites depended systematically on the serial position of that letter in the three-letter sequence, with the mean phases associated with the first and third positions differing by  $180^\circ$ . This remarkable discovery supports the idea that maintenance of a P-set involves sequential activation of its members.

For other tests that bear on the LIJ model, see Sections 11 and 12.

### 3.4. Modification of the LIJ model suggested by the rat brain

In the LIJ model, the gamma period has zero variance. However, Atallah and Scanziani (2009) have shown that in the hippocampus of the rat, the periods of gamma oscillations of the local field potential have a very large range: Their paper provides examples of periods from 11 to 35 ms in an awake, moving rat, and from 12 to 44 ms and 11 to 47 ms in two different anaesthetized rats. I am not suggesting that details of such data agree with details of gamma oscillations in awake humans performing a particular task. But because these are among the best data we have just now that inform us about gamma-period varia-

bility, they should be considered. In addition, intracranial recordings from humans also indicate considerable variability in the gamma period.<sup>37</sup> Atallah and Scanziani (2009) find that the duration of the period that follows a peak is positively correlated with the height of the peak, and conclude that a higher peak reflects more neural activity, which generates more inhibition, resulting in a longer delay before the next peak. This *peak-delay property* helps to explain the effect of item complexity on the scanning rate, as discussed in Section 4.

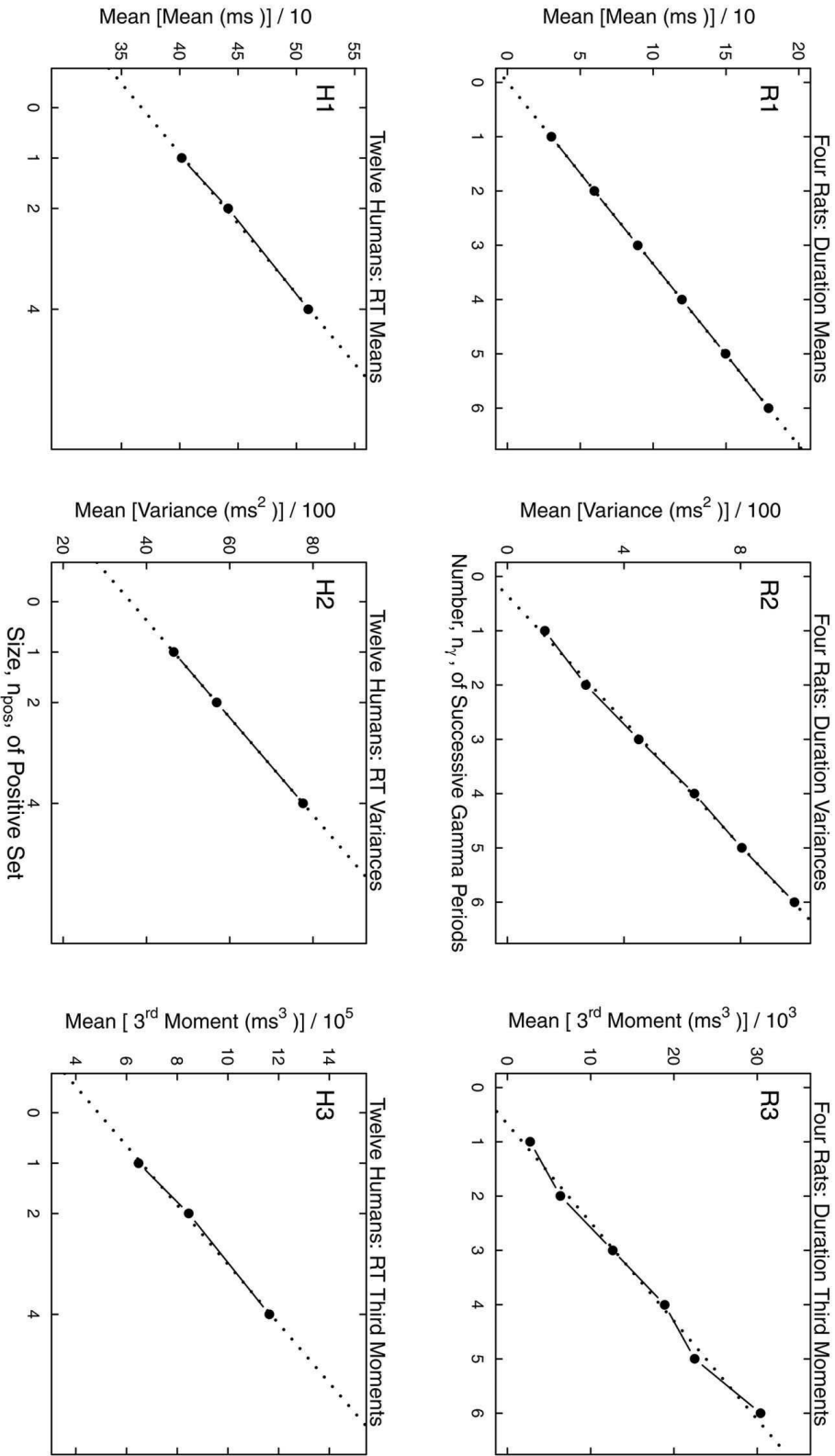
To determine whether gamma-period variability could contribute appropriately to the observed changes in the RT distribution with  $n_{\text{pos}}$ , thus obviating the need for the “skip” feature, I analysed the data collected by Atallah and Scanziani from four anaesthetized rats.<sup>38</sup>

If the process in memory scanning that depends on  $n_{\text{pos}}$  is driven by a sequence of  $n_{\text{pos}}$  successive gamma periods, then the RT should include the duration of such a sequence. The distributions of the cumulated durations of a set of successive gamma periods depends on any covariances among them as well as on variability of a single period. One way to avoid the complexity this implies is to consider the distributions of the cumulated durations of different numbers of successive periods. Figure 1 shows the relation between the number of successive gamma periods,  $n_\gamma$ , and means over the four rats of each of the first three moments of their cumulated durations. For any  $n_\gamma$ , the estimated duration of  $n_\gamma$  successive periods depends on precisely locating two peaks. Because it is plausible that error in their locations is independent of  $n_\gamma$  and that the true cumulated duration and any measurement error are approximately uncorrelated, any measurement error should only add constants to the estimated moments, independent of  $n_\gamma$ .

<sup>36</sup>However, the effects were small; the analysis was complicated by requiring correction for a global effect of  $f_c$  on RT; the foreperiod as well as the number of clicks the subject heard before seeing the probe were confounded with  $f_c$ ; it appears to have been assumed (or found) that there were no individual differences in  $f_c$ ; and, as in the Vosskuhl et al. (2015) study, the evidence that the oscillation frequency was actually influenced by  $f_c$  was indirect. In an improved version of such an experiment, tACS might be used (Helfrich et al., 2014), as Vosskuhl et al. did, instead of clicks; if possible,  $f_c$  values would be determined in relation to measurements of individual gamma frequencies; and the effect of the manipulation on  $f_c$  would be measured.

<sup>37</sup>Personal communication, Joshua Jacobs, December 2014.

<sup>38</sup>See Appendix C for an outline of the analysis, and some of the findings for individual rats.



**Figure 1.** Panels R1, R2, and R3: Three moments of the distributions of summed durations of  $n_y$  consecutive gamma periods in rat hippocampus, as a function of  $n_y$ . Panels H1, H2, and H3: Three moments of the distributions of RTs on  $N$ -trials in human data from Exp. Stern67int1, as a function of  $n_{pos}$ . (These are the data from the S task used by Jensen & Lisman, 1998, for fitting the LIJ model.) Note that not all experiments using the S task produce functions as linear as these. For the rat data, equations of the fitted lines in the three panels are R1:  $0 + 30n_y$  (ms); R2:  $-58 + 172n_y$  (ms<sup>2</sup>); R3:  $-2,983 + 5,202n_y$  (ms<sup>3</sup>). For the human data, they are H1:  $367 + 36n_{pos}$  (ms); H2:  $3,618 + 1,035n_{pos}$  (ms<sup>2</sup>); H3:  $488,501 + 170,073n_{pos}$  (ms<sup>3</sup>).



The mean values of the first three moments, shown in the top row of Figure 1, increase remarkably linearly for  $1 \leq n_\gamma \leq 6$ , consistent with the contributing gamma periods being identically distributed and stochastically independent. Also, the direct estimates of the covariance of successive periods shown for individual rats in Table C1 in Appendix C are negligible. The plots of the first three moments of human RT data in the second row of Figure 1 also increase remarkably linearly, but with  $n_{\text{pos}}$ ; they are there to suggest that a modification of the LIJ model in which there is no “skipping”, but where the gamma period is variable, might be equally effective in accounting for human RT data.<sup>39</sup>

## 4. SCANNING RATE AND THE MEMORY SPAN

### 4.1. Cavanagh’s discovery and further tests

Scanning rates for items drawn from different ensembles differ systematically. The remarkable linear relationship that Cavanagh (1972) discovered between these variations in scanning rate and variations in the span of immediate memory for items from these ensembles has aroused considerable interest and has invited several replications (Brown & Kirsner, 1980; Lass, Lüer, Becker, Fang, & Chen, 2004; Puckett & Kausler, 1984) or similar attempts (Puffe, 1990).<sup>40</sup> The most extensive such replication is that of Lass et al. (2004), in which performances with each of seven stimulus ensembles in each of two different languages (German and Wu Chinese) were studied in different groups. As shown in Table 16.1 of Lass et al. (2004), for each of these 14

data sets, measurements were made of a large number of subjects: from 48 to 144 for span, and from 48 to 96 for scanning rate.<sup>41</sup>

Cavanagh considered the relationship between the time per item (slope) in memory scanning and the reciprocal of span (which can be considered space per item if the span reflects a fixed “space” or “capacity”). However, span is measured in different ways in different experiments—for example, the maximum length of a list recalled with 100% accuracy, or with 50% accuracy, or the list-length to which a staircase procedure converges. It follows that the origin of the scale of memory span is arbitrary, which means that taking its reciprocal may distort its relationship with other measures. An alternative is to consider the relationship between memory span itself and the scanning rate in items per second (the reciprocal of the slope of the  $\overline{RT}$  function, a scale whose origin is not arbitrary), as Lass et al. (2004) have done. Their data are shown in Figure 2, along with Cavanagh’s, plotted in the same way. Parameters of the fitted linear functions are shown in Table 3, along with corresponding parameters of the data assembled by Cavanagh.<sup>42</sup>

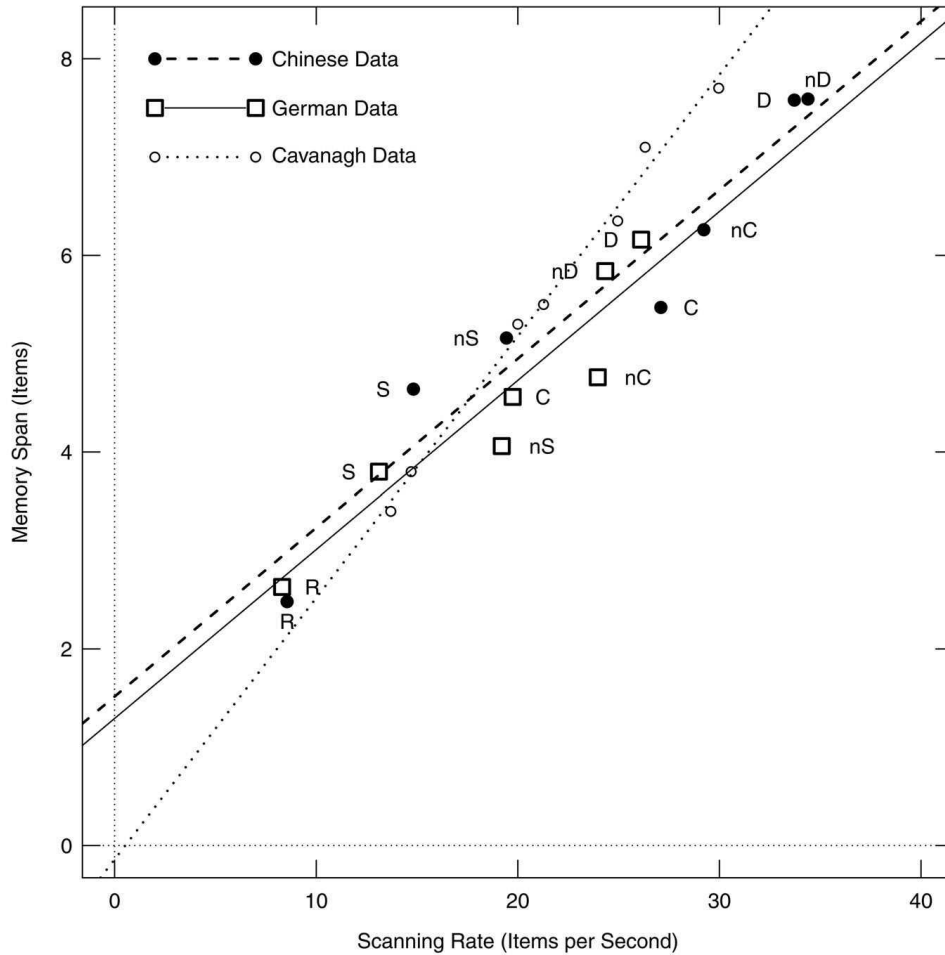
As can be seen in Figure 2, for the six ensembles whose members are nameable, memory spans and scanning rates are both higher for Chinese than for German subjects. Nonetheless, the fitted linear functions have identical slopes and almost equal zero-intercepts: The same law appears to apply to both language groups. The difference between the zero-intercepts of the Lass et al. and Cavanagh functions may result from differences in the definitions of the memory span. The absence of feedback and/or rewards for speed in the studies used by Cavanagh might contribute to their lower scanning rates for the ensembles associated with faster rates,

<sup>39</sup>Although both distributions are positively skewed for all values shown of  $n_\gamma$  and  $n_{\text{pos}}$ , and the Pearson moment coefficient of skewness decreases as both  $n_\gamma$  and  $n_{\text{pos}}$  increase, the rate of decrease is slower for the RTs.

<sup>40</sup>In an effort to collect memory-span and memory-scanning data in one procedure, Puffe used a large range of P-set sizes, fitted bilinear  $\overline{RT}$  functions to the data as Burrows and Okada (1975) had done, and assumed that the breakpoint of the bilinear function is a measure of the memory span. Unfortunately, no measure of precision of the estimated spans was provided.

<sup>41</sup>Another reason to trust these results (in addition to the large sample sizes) is that all conditions included feedback on speed and accuracy as well as performance-dependent payoffs, whose importance is discussed in Appendix B.

<sup>42</sup>These data have been treated as if only the span measure is subject to error.



**Figure 2.** Immediate memory span versus estimated memory scanning rate for Chinese-speaking and German-speaking subjects working with seven stimulus ensembles: R = Random shapes; S = Geometrical shapes; nS = Names of geometrical shapes; C = Colours; nC = Names of colours; D = Digits; nD = Names of digits. Based on Table 16.1 of Lass et al. (2004). Also shown are values for the data assembled by Cavanagh (1972), from diverse studies, 32 providing measures of scanning rate and 13 providing measures of memory span, plotted in the same way. In order of increasing scanning rate, the ensembles in Cavanagh’s data are nonsense syllables, random forms, words, geometric shapes, letters, colours, and digits. Scanning rates are reciprocals of the means of reported RT function slopes for each stimulus ensemble.

and the resulting greater slope. It is not obvious what distinguishes ensembles whose items occupy more or less of the capacity of AM: Is it familiarity, or,

**Table 3.** Parameters of fitted linear functions: memory span (items) versus scanning rate (items/s).

Data	Slope		Zero intercept	
	Mean	Interval	Mean	Interval
Lass et al.: German	0.17	(0.09, 0.31)	1.3	(-1.3, 2.2)
Lass et al.: Chinese	0.17	(0.09, 0.30)	1.5	(0.4, 2.7)
Cavanagh	0.26	(0.21, 0.32)	-0.3	(-1.3, 0.8)

Note. Intervals are 95% bootstrap confidence intervals.

as suggested by Cavanagh (1972), is it the size of the internal representation—the number of features required for its specification? If scanning time per item and space per item are both proportional to size, as he suggested, and memory capacity is measured in features, as he suggested, this would explain the relationship he discovered. I shall call this aspect of the items in an ensemble their “complexity”.

Together with the invariance of linearity and  $\beta$  across 4000 trials of practice in the Sf task (Kristofferson, 1972b), the relation of  $\beta$  and memory span supports the idea that even though

many laboratory procedures do not elicit SES (Section 2), it represents a fundamental operation of human information processing.

#### 4.2. Memory span versus scanning rate and the LIJ model

Recall the peak-delay property described by Atallah and Scanziani (2009): The gamma period that follows a higher peak in the local field potential is prolonged. Suppose (a) this property also applies to humans, together with (b) a more intensive process generates more neural activity and hence a higher peak, and (c) activation (or comparison) of more complex items requires a more intensive process. The gamma period will then increase with item complexity. If we also assume that complexity does not influence the theta period, then, as pointed out by Jensen and Lisman (1996c), the LIJ model can explain the linear relation between memory span and scanning rate across levels of item complexity. However, for this explanation to work, we must also assume that the skip probability,  $p_a$ , is not only invariant across  $n_{\text{pos}}$  values, as mentioned in Section 3.2, but is also invariant across levels of item complexity.

#### 4.3. Two kinds of correlation between memory span and scanning rate

Stimulated by Cavanagh's discovery of the correlation *across stimulus ensembles*, several investigators have searched without success for correlations *across subjects* of measured scanning rates and memory spans. These include Brown and Kirsner (1980), Puckett and Kausler (1984), Cowan et al. (1998),<sup>43</sup> and, most recently, Lass et al. (2004), in an experiment with 48 subjects. Hulme, Newton, Cowan, Stuart, and Brown (1999) found a correlation across subjects between the memory-scanning rate for one-syllable words and memory

span, but their memory-scanning data were highly atypical, with a slope of about 67 ms/word for one-syllable words, and a zero intercept of about 760 ms.<sup>44</sup>

Whether to expect such a correlation across subjects depends on the mechanism that connects scanning rate and memory span, and how different features of that mechanism vary from one subject to the next. For example, consider the LIJ model, and suppose that, as mentioned above, the gamma period, but not the theta period, increases with item complexity, thus explaining the linear relationship between memory span and scanning rate, *across stimulus ensembles*. However, if the primary variation *across subjects* was in the theta period and not the gamma period, we would not expect subjects with larger memory spans to have faster scanning rates. Or, if the two periods varied proportionally across subjects, then the scanning rate would vary across subjects, but not the memory span.

### 5. BEHAVIOUR OF THE RT VARIANCE

#### 5.1. What do we expect from the serial exhaustive search model?

Suppose, as indicated in Figure 4 of Sternberg (1975), that the serial-comparison process is one of four stages in item recognition, and that the RT is the sum of their durations. How do we expect the RT variance,  $\text{var}(RT)$ , to be influenced by  $n_{\text{pos}}$  and by response type (positive or negative)?

Unlike predictions based on the SES model about the behaviour of  $\overline{RT}$ , two additional requirements must be satisfied to be certain about what has been claimed to be its predictions for  $\text{var}(RT)$ . First, we must elaborate the model by adding the assumption that durations of the individual comparisons within the serial-comparison process are mutually uncorrelated as well as being uncorrelated with the durations

<sup>43</sup>They believe that the correlation they found was between a slower memory-search process and memory span.

<sup>44</sup>A different connection between memory span and scanning rate was suggested by Cowan et al. (1998), Cowan et al. (2003), and Hulme et al. (1999), who have proposed that the time intervals between items during immediate memory recall reflect a rapid sequential activation process akin to high-speed memory scanning.

of the other stages.<sup>45</sup> Second, even if the SES model is elaborated in this way, we have to estimate variances in such a way as not to introduce spurious correlations. (For example, because practice might influence the durations of more than one stage, the data that provide variance estimates should not be pooled over levels of practice.)

If we measured the RT variability separately for each P-set and probe combination, and then combined those measures, the above requirements would be sufficient for the predictions to be valid, but I am unaware of any analyses that do this. The result is that there are additional issues related to possible differences among P-sets of the same size, and to differences among probes. Even in analyses of data from the Sf task, I know of none in which variances (or means) are estimated separately for different probes. If probe encoding for recognition shares properties with probe encoding for naming, then priming effects, which may depend on  $n_{\text{pos}}$  (Kirsner, 1972), may mean that the magnitude of the contribution to  $\text{var}(RT)$  from probe encoding may not be independent of  $n_{\text{pos}}$ . Furthermore, if different P-sets of the same size produce different RTs, in addition to any effects of probe differences, then in the Sv task as normally used, contributions to  $\text{var}(RT)$  from P-set differences cannot be separated from effects of  $n_{\text{pos}}$  itself. Because variances in the Sv task but not the Sf task must include this source of variation, we may find greater variances in the former.

Only if the elaborations of the SES model are valid, and any effects of probe differences and P-set differences depend minimally on  $n_{\text{pos}}$ , and data are appropriately analysed, can we predict that the increase of  $\text{var}(RT)$  with  $n_{\text{pos}}$  should be linear, as claimed by Doshier and McElree (1992, p. 401), and that the rates of increase should be the same

for P-probes and N-probes, as claimed by McElree and Doshier (1989, p. 347).<sup>46</sup>

If the variance is estimated from data pooled over probes, as is typical, and probes differ in encoding duration, then as  $n_{\text{pos}}$  grows, and especially if the number of different N-probes decreases as  $n_{\text{pos}}$  increases,  $\text{var}(RT)$  would be expected to increase faster with  $n_{\text{pos}}$  for P-probes than for N-probes; simulations suggest that the growth would be nonlinear.

## 5.2. What is found?

Even when individual RTs from my Sf experiments of the 1960s have been retained, the identity of the probe associated with each RT has not been preserved, so that a suitable within-probe estimate of  $\text{var}(RT)$  is not possible.<sup>47</sup> Also, the design of the Sv task, with a different P-set on every trial, precludes removing any RT differences among P-sets of the same size. Thus, the available RT variances include these components in addition to the variance of the scanning process itself.

Table 4 provides mean estimates of  $\text{var}(RT)$  on N-trials and P-trials as functions of  $n_{\text{pos}}$  in five experiments that used either Sf or Sv procedures, together with slopes of fitted linear functions and the percentage of variance explained by these linear functions. Where available, standard errors of the slopes are also provided. Also given are the total number of trials contributing to each tabulated value.<sup>48</sup> For Exps. Stern66.2 and Stern67int,  $n_{\text{pos}}$  was changed by altering the mapping of probes onto responses, while keeping the structure of the sequence of probes the same. Variance values in the rows labelled “all” for these experiments are weighted means of  $\text{var}(RT)$  for P-trials and N-trials, weighted by their relative frequencies of 4/15 and 11/15, so that the distribution of probes that contributes to these variances

<sup>45</sup>It is not implausible that such independence might be violated. For example, if the quality of encoding varies from trial to trial, and influences the comparison time, then this could create a negative covariance between the durations of the encoding and comparison stages, as well as a positive covariance among comparison durations.

<sup>46</sup>Schneider and Shiffrin (1977, p. 30) mention the need to elaborate the model.

<sup>47</sup>However, analyses of components of variance for Exps. Stern75a and Stern75b, in which the stimulus ensemble consisted of (highly familiar) digits, showed significant contributions of probe differences on both P-trials and N-trials.

<sup>48</sup>Error rates are sufficiently low so that the number of trials is a good approximation of the number of trials on which responses were correct.

**Table 4.** *RT variances ( $ms^2$ ) in five experiments using Sf and Sv tasks.*

Task	Experiment	Resp	Trials	$n_{pos}$						Slope ( $\gamma$ )	Pct Lin	
				1	2	3	4	5	6			
Sf	Stern66.2 6	neg	528	3088	5088		7400				$1397 \pm 260$	97.8
		pos	192	2694	5110		6643				$1238 \pm 536$	90.2
		all	720	2983	5094		7198				$1355 \pm 255$	96.4
Sf	Stern67int 12	neg	950	3578	5478		7315				$1199 \pm 419$	96.1
		pos	346	5062	6005		8374				$1115 \pm 473$	99.7
		all	1296	3974	5619		7597				$1177 \pm 383$	98.1
Sf	Stern69b.4eq 12	neg	600	5354	6621		8547				1050	99.5
		pos	600	4831	5479		8337				1206	97.4
Sv	Stern66.1 8	neg	96	5420	8756	9445	6731	34,740	14,857		$3498 \pm 1837$	35.5
		pos	96	4819	6840	18,308	8307	26,539	36,959		$5994 \pm 2589$	76.8
Sv	Hock84 6	neg	1320			37,109	51,548	69,914	89,033	17,414		99.6
		pos	1320			30,781	45,527	54,552	87,168	18,819		99.1

*Note.* Exp = experiment;  $N_{sub}$  = number of subjects; Resp = response; neg = negative; pos = positive; all = weighted means of values for the two responses; Trials = approximate number of trials contributing to each tabulated value; Slope = slope of a linear function fitted by least squares, with  $\pm SE$  where available. Pcnt Lin = percent of total variance of the means explained by the fitted function. See Appendix F for sources.

is independent of  $n_{pos}$ .<sup>49</sup> The large standard errors of the estimated slopes are reflections of the large sampling error of the variance and its sensitivity to outliers. Except for the experiment with the smallest sample sizes, the growth of  $\text{var}(RT)$  is well approximated by a linear function, especially when one considers the large sampling variance of  $\text{var}(RT)$ . Excluding Exp. Stern66.1, the mean ratio  $\hat{\gamma}_{neg}/\hat{\gamma}_{pos}$  is  $1.05 \pm 0.07$  indicating that the slopes are close to equal. The data suggest that even for the smaller  $n_{pos}$  values, the Sv task produces higher variances, possibly indicating the effect of P-set described above. The extremely high variances in Exp. Hock84 may be a consequence of the special features of that experiment and its analysis, described in Section 2.2.

From Table 4 we can conclude that for the Sf task,  $\text{var}(RT)$  grows approximately linearly with

$n_{pos}$  and with approximately equal slopes for positive and negative responses. However, for the Sv task, the data (except for the smallest  $n_{pos}$  values) are too variable to draw any conclusion with confidence. As shown in Sections 2.2, 2.4, 5.2, and 6.2, there is evidence for the Sf and Sv tasks eliciting the same process. Why, then, should there be differences in  $\text{var}(RT)$  between them? One possibility is that in the Sv task, interference from the P-set presented on one or more previous trials occasionally lengthens RTs.<sup>50</sup>

To support their claim that the  $\text{var}(RT)$  grows more rapidly for positive than for negative responses, Townsend and Ashby (1983, p. 124), Doshier and McElree (1992, p. 401), and McElree and Doshier (1989, p. 347) all cite Schneider and Shiffrin (1977). The most relevant data are from Schneider and Shiffrin's

<sup>49</sup>This depends on assuming that the contributions of probe differences to RT in the Sf task are unaffected by  $n_{pos}$  and are the same for a probe, whether it is a P-probe or an N-probe.

<sup>50</sup>Sample sizes are especially small in Exp. Stern66.1: only 16 trials per subject per value of  $n_{pos}$ . This experiment was continued for two additional exploratory sessions in which between-subject differences in presentation rate and probe delay were introduced—differences that appeared to have no systematic effect on  $\overline{RT}$ . It is helpful to consider the mean variances over the three sessions. For these combined data, linear functions fit better (Pcnt Lin = 99.2 for N-trials; 88.8 for P-trials), and the slopes are closer to being equal ( $4,763 \pm 1,778$  for N-trials;  $3,764 \pm 1,350$  for P-trials).

Experiment 2, the “varied set” condition, with “frame-size” (i.e., display size) 1, an Ashby task.<sup>51</sup>

## 6. BEHAVIOUR OF THE SHORTEST REACTION TIME

### 6.1. What do we expect from the serial exhaustive search model?

As  $n_{\text{pos}}$  increases, what do we expect of the shortest RT,  $\min(RT)$ , in a sample? Suppose that five conditions are satisfied:

1. The duration of a single comparison has a positive minimum.
2. The base-time distribution is not influenced by  $n_{\text{pos}}$ .
3. The durations of all the comparisons and the base time are stochastically independent.
4. The data have been analysed so as not to disturb this independence.
5. The data on all included trials have resulted from successful implementation of SES.

Then the  $\min(RT)$  to execute the process described by the SES model should increase with  $n_{\text{pos}}$ , and should do so on both P-trials and N-trials.<sup>52</sup> However, when any of these conditions is violated, it becomes less clear what to expect. Here are four examples of possible violations:

*Fast guesses.* If the errors on some trials result from fast guesses (Ollman, 1966; Yellott, 1971) rather than “stimulus-controlled responses”, then the expected number of correct responses that result from fast guesses is the same as the number of errors that do. If the error rate increases with  $n_{\text{pos}}$ , then this may mean that the proportion of correct responses that are fast guesses increases with  $n_{\text{pos}}$ . This effect could cause  $\min(RT)$  to actually become

shorter, as  $n_{\text{pos}}$  increases, hiding any increase in  $\min(RT)$  associated with “stimulus-controlled responses”. (This is one reason why the estimate of a low quantile, such as the 10% quantile, may be preferred to the estimated minimum as a characterization of the low tail of the RT distribution; another reason is that unlike a low quantile, the sample minimum is biased high, by an amount that depends on sample size.) The effect of a few fast guesses on the minimum is likely to be substantially greater than their effect on the mean or median.

*Differences in sample size.* For any plausible RT distribution, the expected value of the shortest RT is inversely related to the sample size. In most experiments, the numbers of trials are approximately equal for different values of  $n_{\text{pos}}$ . Therefore, as  $n_{\text{pos}}$  grows, the number of trials for each item in the set decreases, including the most accessible item. Given item differences, and depending on details of distributions, this could cause the shortest RT to increase as  $n_{\text{pos}}$  grows.<sup>53</sup>

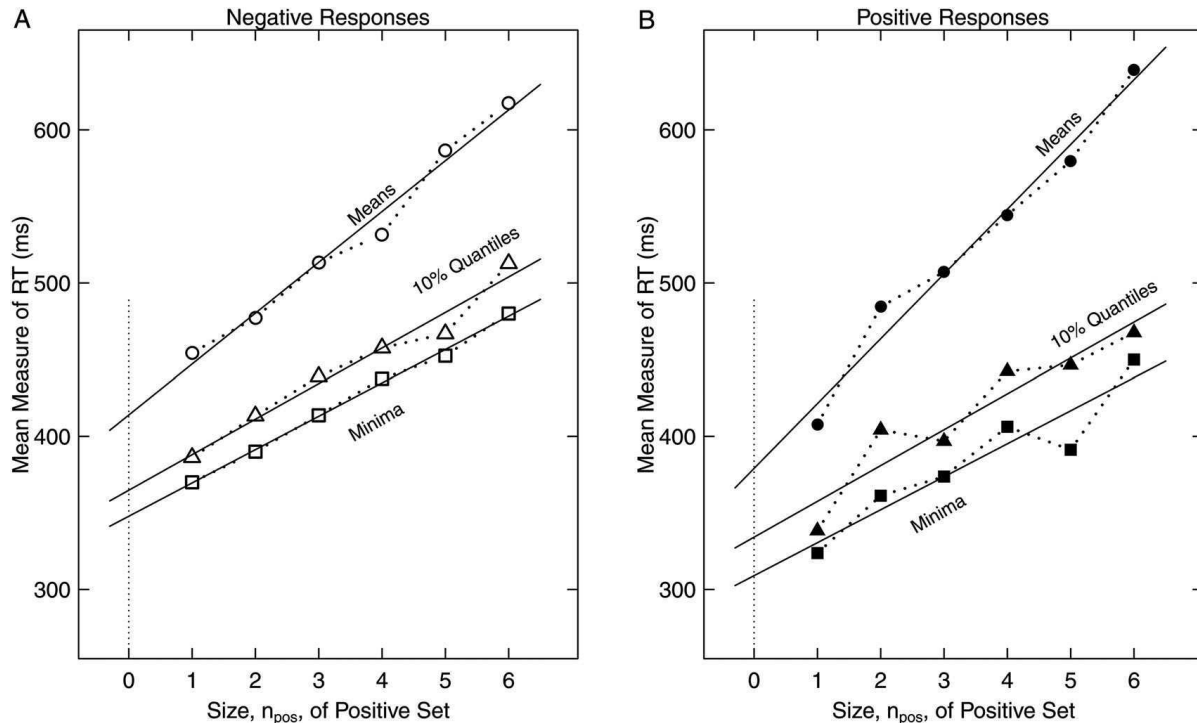
*Operation of the Atkinson–Juola hybrid model (Section 2.1).* The strength criteria in the Atkinson–Juola hybrid model do not vary with  $n_{\text{pos}}$ . Hence, even if strength discrimination is used on only a few trials, the expectation of  $\min(RT)$  will be short and independent of  $n_{\text{pos}}$ . It is worth noting that as  $n_{\text{pos}}$  increases in the Sv task, the recency of the most recent N-probe decreases, while that of the most recent P-probe remains invariant. Thus, to the extent to which the hybrid model operates and some responses are based on strength discrimination,  $\min(RT_{\text{neg}})$  should increase more slowly with  $n_{\text{pos}}$  than  $\min(RT_{\text{pos}})$  does.

*Encoding differences.* We know that  $\text{var}(RT)$  increases precipitously with  $n_{\text{pos}}$  (Table 4;

<sup>51</sup>Dosher and McElree (1992, p. 401) also cite Schneider and Shiffrin (1977) to support their assertion that “Predictions of linear increases in variability ... also fail”. However, the latter authors say (p. 30) that “On the whole, despite a certain amount of noise in the data, the variances are approximately a linear function of the load”.

<sup>52</sup>If the five conditions are satisfied, and the positive minimum mentioned in Condition 1 is large, then Dosher and McElree (1992, p. 401) are correct that the SES model predicts “that the minimum RT should depend fairly strongly on list length”.

<sup>53</sup>As examples of the effect of sample size and distributional shape and spread on the bias, for Gaussian, uniform, and exponential distributions, respectively, as the sample size grows from 10 to 100, the sample minimum shrinks by about 96%, 28%, and 9% of the standard deviation.



**Figure 3.** Mean values over six subjects of means, 10% quantiles, and minima as functions of  $n_{pos}$  for N-trials (panel A) and P-trials (panel B) in Exp. Stern66.1, which used the Sv task. Equations of the lines, fitted by least squares, are, for N-trials and P-trials, respectively,  $348 + 21.8n_{pos}$  and  $309 + 21.5n_{pos}$  (minima);  $365 + 23.2n_{pos}$  and  $334 + 23.4n_{pos}$  (10% quantiles); and  $414 + 33.2n_{pos}$  and  $379 + 42.3n_{pos}$  (means).

Sternberg, 1975, p. 15 fn). Suppose that this is partly due to an increase in the base-time variance, possibly because the time to encode the probe varies across members of the stimulus ensemble. Then, as  $n_{pos}$  increases, the chance that a rapidly encoded item is included increases for P-trials, and, as a result,  $\min(RT)$  could decrease for P-trials. (If  $n_{neg}$  decreases as  $n_{pos}$  increases, we should see the reverse effect for N-trials.)

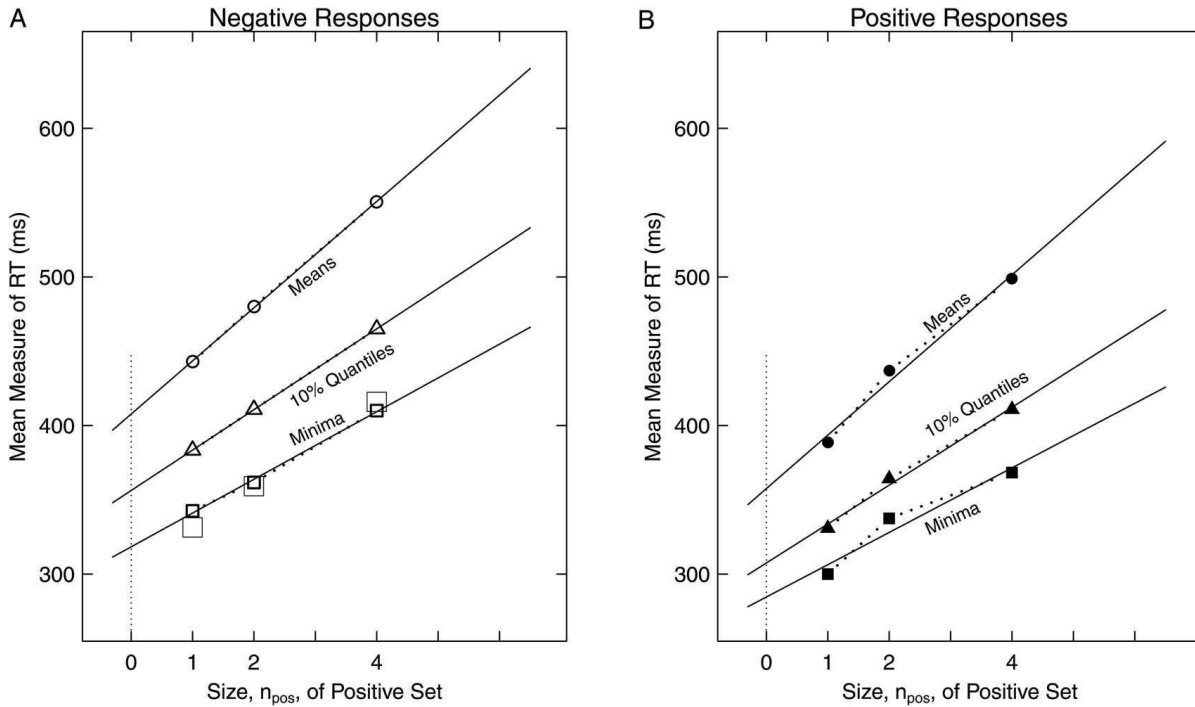
## 6.2. What is found?

Using the Sf procedure, with nested P-sets, Lively and Sanford (1972) and Lively (1972) reported that  $\min(RT)$  in their samples increased systematically with  $n_{pos}$ : For same-category

probes,  $\min(RT)$  in the former study increased at a rate that was 57% of the rate of increase of the mean; in the latter study, combining speed and accuracy conditions,  $\min(RT)$  increased at 56% of the rate of increase of the median.<sup>54</sup> Similar findings for Sv and Sf tasks (Exps. Stern66.1 and Stern69b.4eq) are presented in Figures 3 and 4. To facilitate comparison, the scales in these figures are the same. Because the sample minimum is sensitive to outliers, and is also a biased estimator of the population minimum to an extent that depends on sample size, 10% quantiles (which are median-unbiased) are also shown.<sup>55</sup> Means are also shown, because the relation between minimum and mean, and its dependence on  $n_{pos}$ , are of interest. The ratios of

<sup>54</sup>These findings were used in Section 6.1 of Sternberg (1975) as evidence against the search being self-terminating. In both studies, the rates of increase of  $\min(RT)$  with  $n_{pos}$  on P-trials and same-category N-trials were both significantly greater than zero, but the difference between these rates was not significant. Also in that section, the approximately linear and equal rates of increase of RT variance were described, with the rate of increase used as additional evidence against the search being self-terminating.

<sup>55</sup>Quantiles were estimated using the Hyndman and Fan (1996) Type 8 median-unbiased estimator.



**Figure 4.** Mean values over 12 subjects of means, 10% quantiles, and minima as functions of  $n_{pos}$  for N-trials (panel A) and P-trials (panel B) in Exp. Stern69b.4eq, which used the Sf task. Equations of the lines, fitted by least squares, are, for N-trials and P-trials, respectively,  $318 + 22.7n_{pos}$  and  $285 + 21.7n_{pos}$  (minima);  $356 + 27.2n_{pos}$  and  $308 + 26.2n_{pos}$  (10% quantiles); and  $408 + 35.8n_{pos}$  and  $358 + 36.0n_{pos}$  (means). In Panel A, the large, light squares, described by  $303 + 28.2n_{pos}$ , are means of minima based on the distribution analysis of data from Exp. Stern67int1, discussed in Section 6.3. Based on between-subject differences in Exp. Stern69b.4eq, the standard error of the mean slope of the minimum is 5.8 ms.

the rates of increase of minimum to mean are similar to those mentioned above, and to each other:  $61 \pm 7\%$  for the Sv task, and  $62 \pm 9\%$  for the Sf task.<sup>56</sup> The substantial increase of the minima with  $n_{pos}$  in Figures 3 and 4 argues that strength discrimination, as in the hybrid model, does not play an important role. Also, except for the means in Figure 3, the differences between slopes for P-trials and N-trials in all six comparisons of Sv and Sf results do not exceed 1 ms/item, supporting the idea that the same process (SES) underlies both.

However, for one of the six subjects in the Sv task of Exp. Hock84, the rate of increase with  $n_{pos}$  of the minimum RT on N-trials was smaller, possibly substantially smaller, while the rate of increase of the mean RT was greater.<sup>57</sup> Based, apparently, on only these data, it has been concluded that, as  $n_{pos}$  increases, “Distributions of RTs ... show only tiny shifts of the minimum ...” (Doshier & McElree, 1992, p. 401), “the fastest times are the same” (Doshier & Sperling, 1998, p. 240), and “... for the memory search task, the leading edge stays relatively constant ...” (Hockley, 1984, p. 604).

<sup>56</sup>For each subject, these ratios were calculated separately for P-trials and N-trials, and then averaged. The values reported are the means and standard errors of these averages over subjects.

<sup>57</sup>The histograms of RTs on N-trials for this one subject have been published at least four times: by Hockley (1984, Figure 5), Doshier and McElree (1992, Figure 3b), Doshier and Sperling (1998, Figure 25d), and Hockley (2008, Figure 2). Estimates of the minima of these data have, themselves, never been reported, nor have any estimates of the precision of such estimates, nor have this subject’s error rates, nor the histograms or minima for the other five subjects. Because the bin size in these histograms is 50 ms, and the bin that represents the shortest RTs in the sample (300–350 ms) is occupied for all  $n_{pos}$  values, all that we can say with certainty about the change of the sample minimum as  $n_{pos}$  increases from 3 to 6 (and the mean  $RT$  over subjects increases by about 147 ms), is that it lies between a decrease of 50 ms and an increase of 50 ms.



What of the P-trials and the other five subjects? In the absence of the RT distributions, I used the means over the six subjects of the estimated parameters of the ex-Gaussian distributions for P-trials and N-trials that Hockley fitted to his Sv-task data (1984, Figure 4), and I simulated the distributions using a sample size of 218 (the sample size per subject per  $n_{\text{pos}}$  value per response type in the experiment). From 1000 replications, I determined the mean values of the minima and mean for each response type and each  $n_{\text{pos}}$  value, and fitted the relations between these values and  $n_{\text{pos}}$  with linear functions. The slopes (mean rates of increase with  $n_{\text{pos}}$ ) were similar for P-trials and N-trials; their means were 16.7 ms/item for minima, 35% of the 47.7 ms/item for means. Thus, the increase with  $n_{\text{pos}}$  of  $\overline{\min(RT)}$  relative to  $\overline{RT}$ , based on mean estimated ex-Gaussian parameters for the six subjects, is less than what was found in other studies, but was substantial, nonetheless.

I applied the same method to the data for  $n_{\text{pos}} = 2, 4, \text{ and } 6$  from the repeated negatives condition in Exp. 1 of Hockley and Corballis (1982), which used a variant of the Sf task with an ensemble of two-syllable nouns.<sup>58</sup> Starting with mean ex-Gaussian parameters (measured from their Figure 2), and using a sample size of 192 and, again, 1000 replications, I found that the mean rate of increase of  $\overline{\min(RT)}$  with  $n_{\text{pos}}$  was 21.6 ms/item, 54% of the rate of increase of the mean (40.2 ms/item), similar to the percentages in other experiments mentioned in Section 6.2.<sup>59</sup>

### 6.3. The shortest RT based on estimated distributions of stage durations

Let us extend the SES model by assuming that the durations of each of the  $n_{\text{pos}}$  successive comparisons

are identically distributed and stochastically independent, as well as being stochastically independent of the base time (the summed duration of other stages, unaffected by  $n_{\text{pos}}$ ). If we then let  $\kappa_r(\text{base})$  and  $\kappa_r(\text{comp})$  be the  $r$ th cumulants of the base time and comparison time distributions, respectively, and let  $\kappa_r(RT)$  be the  $r$ th cumulant of the RT, we will have:

$$\kappa_r(RT) = \kappa_r(\text{base}) + n_{\text{pos}} \times \kappa_r(\text{comp}). \quad (1)$$

Thus, because cumulants are additive under these conditions,  $\kappa_r(RT)$  will be a linear function of  $n_{\text{pos}}$ , whose zero intercept is the  $r$ th cumulant of the base time, and whose slope is the  $r$ th cumulant of the comparison time. [Just as the slope of the  $\overline{RT}$  function provides an estimate of the mean comparison time, the slopes of the  $\kappa_r(RT)$  functions provide estimates of the  $r$ th cumulants of the comparison time, and similarly for the zero-intercepts and the base time.] When I examined the mean over subjects of estimates of the first four RT cumulants for each  $n_{\text{pos}}$  value on N-trials in Exp. Stern67int1 (which used the Sf task), I found the first three to be beautifully linear.<sup>60</sup> [In these data,  $\kappa_4(RT)$  increased with  $n_{\text{pos}}$ , but not linearly.] Nonetheless, I fitted linear functions to all of the first four cumulants of the RT and thus obtained estimates of the corresponding cumulants of the base-time and comparison-time distributions (Sternberg, 1964), estimates that agreed, roughly, with those from two other experiments.<sup>61</sup> Given the first four cumulants, one may be able to find a corresponding distribution in the Pearson family of distributions, a family that includes many common distributions, including beta, exponential, gamma, Gaussian, and others. The probability

<sup>58</sup>As the R-S interval was only 0.5 s, a variant discussed in Appendix D, conclusions for the Sf task must be tentative.

<sup>59</sup>Of course, the validity of these simulated minima depends on the goodness of fit of ex-Gaussian distributions to the data, about which questions can be raised: Whereas only 4% of the memory scanning data sets in Hockley (1984) deviated significantly ( $p < .05$ ) from the ex-Gaussian distribution, 36% of the data sets from Exp. 1 of Hockley and Corballis (1982) did. And it is not clear what the relation is of the distribution based on averaged parameters to the individual distributions that gave rise to the parameters that were averaged. To support claims about the "leading edge", it would be far better to determine the minima, or low quantiles, directly.

<sup>60</sup>The first three cumulants are the same as the first three central moments, which, for this experiment, are shown in Panels H1, H2, and H3 of Figure 1. Because this method should ideally be applied to stable data from individual subjects, rather than means, the procedure described is a compromise. Nonetheless, it is worth determining where it leads.

<sup>61</sup>Also, the sets of estimated cumulants for each of the six distributions from the three experiments satisfied inequalities that are required if the associated distributions are non-degenerate, unimodal, and contain only non-negative values (Johnson & Rogers, 1951; Mallows, 1956).

density functions of the inferred distributions that I found are, for the base time,

$$f(t) = \begin{cases} c(t - 299)^{.67}(51 + t)^{-11.23}, & (t \geq 299 \text{ ms}) \\ 0, & (t < 299 \text{ ms}), \end{cases} \quad (2)$$

and for the comparison time,

$$g(t) = \begin{cases} k(t - 25.5)^{-.91}(821.0 - t)^{6.03}, & (25.5 \leq t \leq 821.0 \text{ ms}) \\ 0, & (t < 25.5; t > 821.0 \text{ ms}), \end{cases} \quad (3)$$

where  $c$  and  $k$  are normalizing constants. These functions show that there can be no base times less than 299 ms, and no comparison times less than 25.5 ms. Both density functions are positively skewed and rise steeply from zero: The mode of  $f(t)$  is at  $t = 321$  ms (22 ms above its minimum); its three quartiles are 326 ms, 350 ms, and 390 ms. Because  $g(t)$  is monotonically decreasing, its mode is at its first non-zero value ( $t = 25.5$  ms). Its three quartiles (25%, 50%, 75%) are 25.6 ms, 26.9 ms, and 38.5 ms, respectively, its median only 1.4 ms above its minimum.

Thus, given the Pearson family, if we specify the mean, spread, skewness, and kurtosis of these data sets (i.e., specify the first four cumulants), we arrive, remarkably, at estimates of the population minima of  $f(t)$  and  $g(t)$ , and, with stochastic independence of the durations of successive stages, and the additivity of their minima that follows from it, we have estimates of the growth with  $n_{\text{pos}}$  of the theoretically shortest RT. To test this analysis, one could use the inferred distributions to simulate the data, and ask whether aspects of the observed distributions, such as the minima, could be recovered. Unfortunately, the observed minima of the RTs in Exp. Stern67int1 are not readily available. I therefore compared the simulation to the  $RT_{\text{neg}}$

data from Exp. Stern69b.4eq, another Sf experiment.

The number of N-trials for each value of  $n_{\text{pos}}$  per subject was about 50 in Exp. Stern69b.4eq, and we expect the minimum RT in 50 trials to be greater than the theoretical minimum. For comparison to the obtained minima, then, random sampling from the estimated distributions was needed.

For each of 50 simulated trials, I added one sample from the base-time distribution to the sum of one, two, or four samples from the comparison-time distribution, thus simulating the corresponding convolutions, found the minimum of each sum among the 50 simulated trials, and repeated this for 10,000 simulated subjects. Means of these minima are shown in Figure 4A. Their values are 331.4, 359.2, and 415.9 ms (described by  $303 + 28.2n_{\text{pos}}$  ms), averaging about 10 ms above values of the theoretical minima of 324.5, 350.0, and 401.0 ms (described by  $299 + 25.5n_{\text{pos}}$  ms).<sup>62</sup> As shown in Figure 4, then, sampling from the inferred distributions for N-trials in Exp. Stern67int1, the means of the RT minima are found to increase with values close to the means over 12 subjects of the directly observed RT minima of samples of about 50 RTs from N-trials in a different experiment (Exp. Stern69b.4eq) that also used the Sf task. This agreement, together with the plausibility of the inferred distributions, supports the extended SES model for the Sf task.<sup>63</sup>

#### 6.4. Parallel comparisons and the shortest RT

It would be interesting to know whether any plausible model of parallel comparisons would be consistent with a linear growth of the shortest RT with  $n_{\text{pos}}$ , at a rate that is 50% or 60% of a linearly increasing mean RT.

<sup>62</sup>This simulation shows that the change in the RT distribution as  $n_{\text{pos}}$  increases from 1 to 4 is associated with an increase in the bias of the sample minimum, but that this increase is only 2.7 ms/item.

<sup>63</sup>Promising as these findings are, the experiments that provided the data used to estimate the distributions were not designed for this purpose. More suitable experiments would provide more practice to achieve better stability, and an analysis that permitted removal of any effects of nuisance factors such as probe differences.

## 7. ACCESSIBILITY OF THE MOST ACCESSIBLE ITEM AND THE MIXED-REGENCY CONJECTURE

One reason why any effect of serial position on P-trials is of special interest is that, without elaboration, a model in which scanning is exhaustive cannot explain it. Whereas the effects of serial position on RTs for positive responses are relatively small in Sv and Sf tasks,<sup>64</sup> they are large in the Monsell task. Indeed, what is striking about the RT data in this task is that, given the serial position, the effect on P-trial RTs of  $n_{\text{pos}}$  itself can be minimal: Almost all that matters is the lag between the probed item and the probe, and the effect of  $n_{\text{pos}}$  on RT averaged over serial position is almost entirely a consequence of larger sets being associated with longer average lags. This is especially true of the last serial position. As reported by McElree and Doshier (1989, Tables 1 and 5), and Monsell (1978, Figures 3b, 6c, and 7b),  $\overline{RT}$ s for the last serial position in the Monsell task are shorter than those for any other position, and, as shown in Table 5, are almost independent of  $n_{\text{pos}}$ : The mean and median of its effects are 3.0 and 1.6 ms per item, respectively. Thus, in the Monsell task the effect of  $n_{\text{pos}}$  on the accessibility of the most accessible item is minimal.<sup>65</sup>

Thus, even if a strength criterion in this task is influenced by  $n_{\text{pos}}$ , the separations between the strength of the most accessible item and the criteria for all the  $n_{\text{pos}}$  values must be sufficiently great so that the effects on  $\overline{RT}$  of the differences in separation are small. This could happen if the function relating RT to the separation decelerates.

Why is the effect of serial position reduced when the probe delay is lengthened? To explain this, at least some of those who believe that the same

mechanism is employed in Sv and Monsell tasks have proposed (without evidence) what I shall call the “mixed-recency conjecture”. McElree and Doshier (1989) suggest that whereas with a fast presentation rate and a short probe delay, “subjective and objective recency are strongly coupled” (p. 352), “longer retention intervals (>1 s), . . . allow . . . partial rehearsal to alter subjective recency” (p. 349). Baddeley and Ecob (1973, p. 230) say that “Although serial position effects are frequently absent in the tasks used by Sternberg, this is probably because there is usually a delay between presentation and test during which the subject rehearses the set. Such rehearsal might be expected to obscure any recency effects . . .” And in their discussion of Monsell’s (1978) data, Nosofsky, Little, Donkin, and Fific (2011, p. 11) say: “If rehearsal takes place, then the psychological recency of the individual memory-set items is unknown because it will vary depending on each subjects rehearsal strategy.”

These authors seem to believe that with different probe delays, the same process is invoked, but that whereas the recency effect—which is diagnostic of this process—is present regardless of probe delay, it is hidden in the Sv task, because rehearsal renders the effective recency unknown. The idea is that the recency associated with a particular probe on a particular trial depends on where the subject happens to be in covert cyclic rehearsal when the probe appears on that trial: If the probe happens to be the last item to have been rehearsed at that moment, it is, effectively, the most recent item. The absence of an effect of (objective) serial position then results from the effective recency for a probe in any serial position being a *mixture* of the possible lags 1, 2, . . . ,  $n_{\text{pos}}$ : As in the Monsell task, the accessibility of the most accessible item

<sup>64</sup>See, e.g., data in Figure 5B (discussed in Section 8), from the AM condition.

<sup>65</sup>Monsell (1978) found this for stimulus ensembles containing consonants (Exp. 1) and two-syllable words (Exp. 2), using instructions that asked subjects not to rehearse. McElree and Doshier (1989) found it for two-syllable words, in both their pilot experiment and their Exp. 2, with no instructions about rehearsal. For reasons that are unclear, this was not found in the Monsell task of Exp. Nosof11, with a stimulus ensemble containing consonants, also discussed by Donkin and Nosofsky (2012, Figure 2): An effect of  $n_{\text{pos}}$  on the RT for the most recently presented probe (lag 1) is shown by all four subjects in that experiment, with a mean of about 15 ms/item, and all four subjects show large effects of recency. However, there is a clear difference between these data and their data using the Sv task with digits (Exp. Donk12, 2012, Figure 5), in which two of their three subjects show a substantial effect of  $n_{\text{pos}}$  on the  $\overline{RT}$  for lag 1, and no subject shows much effect of recency.

**Table 5.** *Effect of  $n_{pos}$  in Monsell task on  $\overline{RT}$  (ms) for most recent item.*

$n_{pos}$	1	2	3	4	5	6	Increase/Item
Experiment							
McEl89.p			609	604	599		-0.5 ms
McEl89.2			592	596	599	596	1.5 ms
Monsl78.1, Immediate	369	378	388	396			9.1 ms
Monsl78.2, Experimental		449	456	460	458		3.2 ms
Monsl78.2, Control		431	417	426	433		1.6 ms

Note. Values for Monsell (1978) were read from graphs.

is almost independent of  $n_{pos}$ —we just do not know on any particular trial which of the items this is.

An alternative explanation for the effect of lengthening the probe delay, suggested by Monsell (1978, p. 496), and mentioned in Section 2.2, is that “The opportunity to rehearse might give the subject time to establish the list in short-term memory in a format suitable for search.”<sup>66</sup> Another possibility depends on the idea that the decline of an item’s activation decelerates, as suggested by the curvilinearity of the serial-position effect in the Monsell task (McElree & Doshier, 1989, Figures 4, 9; Monsell, 1978, Figure 3). If so, shortening the probe delay makes for a greater difference between the item’s activation level and that of less recent (negative) items, increasing the speed and accuracy of strength discrimination, which invites subjects to choose that strategy rather than scanning.

How might the mixed-recency conjecture be tested? It is tempting to say that it is falsified by the observation (Figures 3 and 4) that in both Sf and Sv tasks, an estimate of the minimum,  $\min(RT)$ , increases with  $n_{pos}$  at a rate (22 ms/item) far greater than the rate for the most accessible item in the Monsell task shown in Table 5. However, a rigorous test would require taking into account the decreasing proportion of observations in the mixture distribution on which the estimate is based as  $n_{pos}$  increases, along with the presence of an

increasing number of long RTs. Unfortunately, a satisfactory method for doing so has not yet been developed.

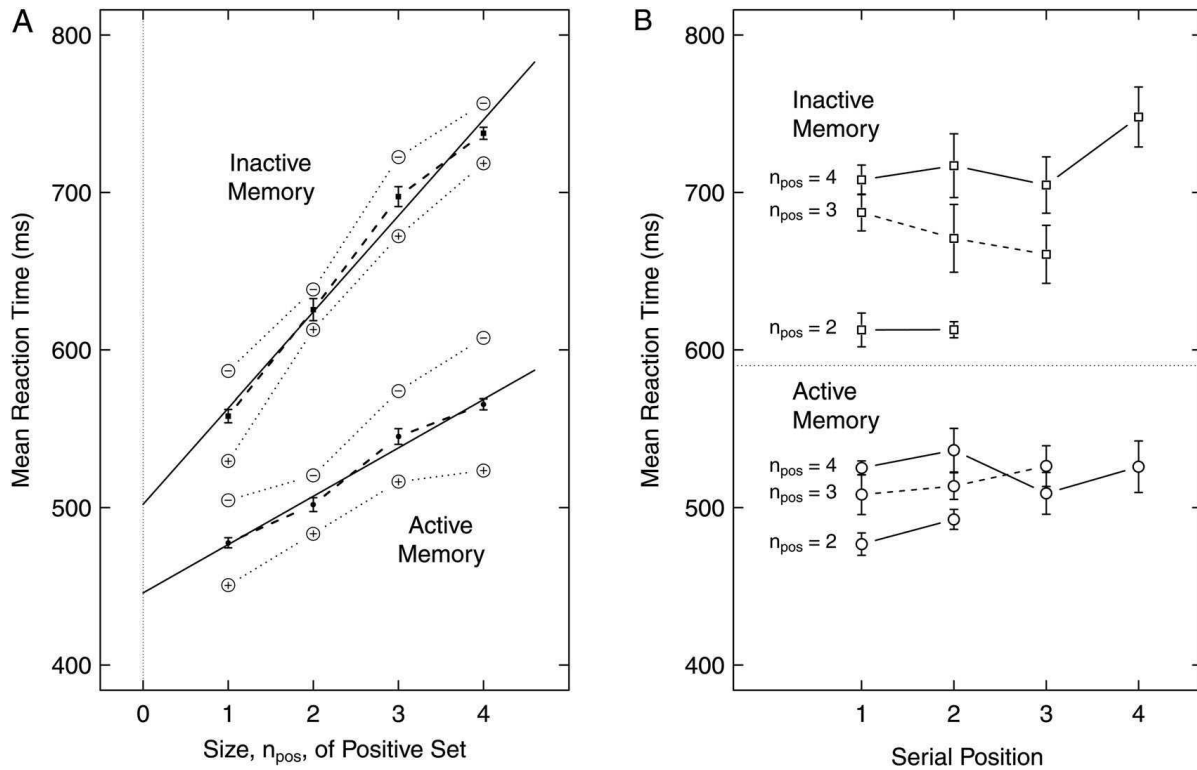
## 8. RETRIEVAL FROM ACTIVE VERSUS INACTIVE MEMORY

In contrast to the Sv procedure, the P-set in the Sf procedure is stored in LTM: Positive sets used on one day can be recalled on the next. The similarity of the findings from the two procedures suggests that in both procedures it is the same memory—the AM—that is searched: In the Sf procedure the information in LTM has been activated. This leads to the question, how is retrieval different if the information searched is not held in AM? An initial small experiment that asks this question, based on the idea that the AM is of limited capacity, was reported in Sternberg, 1969a (Exp. Stern69a).<sup>67</sup> Some of the results from an improved version of this experiment, Exp. Knoll69, are shown in Figure 5.

The data are from 12 subjects who each ran in four test sessions after two practice sessions. Each session, on a different day, contained about 200 trials and lasted for about an hour. Although the four disjoint P-sets of digits, of size 1, 2, 3, and 4, differed across subjects, for each subject they remained the same for all six sessions. (Subjects could recall them at the beginning of the second

<sup>66</sup>See also Diener (1988, p. 375) who suggests that “In the absence of the delay, the memory set may not be stored in a form that is amenable to the search that results in the typical set-size effect.”

<sup>67</sup>Roeber and Kaernbach (2004) attempted to replicate and extend Exp. Stern69a, but their replication failed; this is not surprising, as they used the novel negatives task.



**Figure 5.** Mean data from the 12 subjects in Exp. Knoll69. Panel A: Mean RT as a function of  $n_{pos}$  for active-memory and inactive-memory conditions. Means of RTs for P-trials and N-trials are shown by filled circles with  $\pm SE$  indicated, based on between-subject differences (after removing subject differences in mean and slope); equations for the unbroken lines fitted to them by least squares are  $448 + 30.7n_{pos}$  (active memory), and  $502 + 61.0n_{pos}$  (inactive memory). Mean RTs are shown separately for P-trials and N-trials by open circles containing plus and minus signs, respectively. Panel B: Mean RT on P-trials as a function of serial position for  $n_{pos} = 2, 3,$  and  $4$  in conditions of active and inactive memory, with  $\pm SE$  indicated, based on subject differences after adjusting for means over serial positions

session.) At the start of each trial in blocks of 18, in both AM and LTM conditions, the subject recited the relevant P-set, in a fixed order. In the LTM condition, the subject then saw a list of seven different consonants, different from trial to trial, presented at a rate of 3 letters/s. The probe consisted randomly of either a digit, or a signal to recall the letters in order, each on about half of the trials. If the former, the subject had to make a speeded positive or negative response, depending on its membership in the P-set. The AM condition was the same, but without presentation of a list of letters; in response to the recall

signal, the subject had to recite the first seven letters of the alphabet.

As in the earlier experiment (Exp. Stern69a), the  $\overline{RT}$  functions are both approximately linear, with the slope in the LTM condition twice that in the AM condition, and the slopes for positive and negative responses approximately equal in both conditions. On the P-trials, in neither condition is there a systematic effect of serial position: The weighted mean serial-position slopes, weighted by 1, 2, and 3 for  $n_{pos} = 2, 3,$  and  $4$ , respectively, is  $1.5 \pm 1.4$  ms/item for the AM condition, and  $0.3 \pm 2.8$  ms/item for the LTM condition.<sup>68</sup>

<sup>68</sup>Mean rates of speeded-response errors on P-trials and N-trials, respectively, were 2.4% and 1.6% (AM) and 3.0% and 4.1% (LTM). The recall score was the number of correct letters in correct positions. Mean errors per string were 1.8 (recall control) and 1.7, 1.9, 2.0, and 1.9 for  $n_{pos} = 1, 2, 3,$  and  $4$ , respectively. Lines fitted by least squares to data from P-trials and N-trials, respectively, are  $431 + 25.2n_{pos}$  and  $461 + 36.2n_{pos}$  (AM) and  $477 + 62.6n_{pos}$  and  $528 + 59.4n_{pos}$  (LTM). Corresponding lines fitted to data excluding  $n_{pos} = 4$  are  $418 + 32.9n_{pos}$  and  $464 + 34.6n_{pos}$  (AM) and  $462 + 71.4n_{pos}$  and  $513 + 67.9n_{pos}$  (LTM).

These findings are important in four ways:

### 8.1 Effect of $n_{pos}$ on the activation process

In four studies investigators obtained  $\overline{RT}$  functions for judging P-set membership under two conditions: In one, they regarded the P-set as being held in “primary” memory; in the other, which produced longer RTs, the P-set was regarded as being held in “secondary” memory. They applied the subtraction method to their data, interpreting the difference between  $\overline{RT}$  functions as the time to change the status of the list from secondary to primary memory. Because these differences were relatively constant across  $n_{pos}$  values, they argued that the time required for this change in status did not increase with  $n_{pos}$ . This would seem to contradict the findings described above, in which the difference is a function that increases monotonically with  $n_{pos}$ . However, in all four of these studies, it can be argued that the decisions in the “primary memory” conditions depended on a process other than that used in Sv and Sf tasks.<sup>69</sup>

### 8.2 Consistency with oscillator models of maintenance and search

In both the previous experiment (Exp. Stern69a) and this one, the slope of the  $\overline{RT}$  function in the LTM condition was approximately twice that in the AM condition: Combining the data from the 16 subjects in the two experiments, the mean slope ratio<sup>70</sup> is  $1.99 \pm 0.15$ . In the 1960s, the

conjecture to explain the doubling of slope was that in the LTM condition, the search was preceded by an activation process that was sequential. But it was not obvious why these two processes should occur at the same rate, which would be required for the slope to double. However, if both processes are controlled by the same gamma oscillation rate, as suggested by Vergauwe and Cowan (2014, 2015), then this would explain the doubling.<sup>71</sup>

### 8.3 Absence of serial position effects

In previous experiments where flat serial position functions were obtained, because cyclic rehearsal during the retention interval might be interrupted at a random position by the probe, the nominal and effective serial positions might differ, and there might be no effect of the nominal position, as discussed in Section 7. Here, because the sets were well learned, and recited in order at the start of each trial, and because, in the LTM condition, the subject had to retain a new list of letters before discovering that a digit had to be classified, the nominal and actual serial positions are more likely to correspond.

### 8.4 Evidence against strength discrimination

Based on results from experiments with the Monsell task, some investigators (e.g., Doshier & McElree, 1992; Doshier & Sperling, 1998; McElree & Doshier, 1989; Monsell, 1978) have argued or implied that not only in their experiments,

<sup>69</sup>In the Wickens et al. (1981) and Wickens et al. (1985) experiments using words, the novel-negatives task was used (N-probes were never repeated), so that subjects could base their decisions on memory strength, as in the Monsell task. Also, sets were presented simultaneously, as in the Ashby task. In the Wickens et al. (1985) experiment using consonants, sets were presented simultaneously, and subjects were slow, with mean zero-intercepts and slopes of 625 ms and 52 ms/consonant, respectively. In Conway and Engle (1994), mean slopes based on the two smallest sets ( $n_{pos} = 2,4$ ) were 194 ms/item for words, and 96 ms/item for consonants, far outside the range of slopes for high-speed scanning. Also, lists were presented simultaneously during the learning phase, as in the Ashby task. In Zysset and Pollmann's (1999) similar study, using consonants, slopes (only P-trials reported) for the two smallest sets ( $n_{pos} = 4,6$ ) in their primary memory condition were 57.2 and 57.3 ms/item. In all four studies, analyses started with median RTs; as slopes based on medians tend to be smaller than slopes based on means, all these slopes must be considered above the range for high-speed scanning.

<sup>70</sup>In Exp. Stern69a, with four subjects, linear functions fitted to the mean data are  $336 + 57n_{pos}$  (AM) and  $467 + 105n_{pos}$  (LTM); the mean ratio of slopes in the LTM and AM conditions is  $1.98 \pm 0.22$ . In the improved experiment, with 12 subjects, mean slopes in AM and LTM conditions are  $32.5 \pm 2.0$  and  $62.6 \pm 4.5$  ms/item, respectively, with the mean ratio  $2.00 \pm 0.19$ . Standard errors are based on between-subject differences.

<sup>71</sup>These authors have recently argued for a common rate for several sequential processes in memory and have related this idea to the LJJ model discussed in Section 3.

but also in experiments using the Sv task, the effect on  $\overline{RT}$  of increasing  $n_{\text{pos}}$  results from an increase in the average lag of probes on P-trials, with a resulting decrease in memory strength, along with a reduction in the strength differences between members of P-sets and N-sets. Consider this argument applied to performance in the AM and LTM conditions of the Sf task. An account in terms of memory strength might also include the idea that the RTs in the LTM condition are longer because reciting the P-set was separated from the probe by presentation of the letter list, which would increase the lag, hence decrease the memory-strength difference between P-probes and N-probes.

Some evidence against the importance of lag is of course provided by the flatness of the serial-position curves. The greater steepness of the  $\overline{RT}$  function in the LTM condition provides stronger evidence. Let  $t_{\text{lag}}$  be the separation in time units between the presentation, rehearsal, or pronunciation of an item and the probe of that item. Consider the average  $t_{\text{lag}}$  of items in P-sets of different size in the two conditions. If the reciting rate were 2 digits/s, then, in both conditions, when the probe is presented, the average  $t_{\text{lag}}$  difference of items when  $n_{\text{pos}} = 1$  and 3 would be 0.5 s. Assuming 2 s in the AM condition between the end of reciting and the probe, and 5 s in the LTM condition, the  $t_{\text{lag}}$  of digits in these  $n_{\text{pos}}$  values in the AM condition would be 2.0 and 2.5 s, and in the LTM condition would be 5.0 and 5.5 s.

Data from the Monsell task indicate that as  $t_{\text{lag}}$  increases, recognition RT increases, but at a declining rate (McElree & Doshier, 1989; Monsell, 1978). Thus, effects of  $n_{\text{pos}}$  on RT should be *smaller* in the LTM condition. It follows that whereas an account in terms of memory strength leads us to expect the RT to be longer in the

LTM condition, as observed, it also leads us to expect that the effect of  $n_{\text{pos}}$  on the RT for “yes” responses in that condition (the slope of the  $\overline{RT}$  function) should be smaller than in the AM condition. Yet the slope is about twice as great.<sup>72</sup>

## 9. ABSENCE OF AN EFFECT ON RT OF NEGATIVE PROBE RECENCY

One feature of data from the Monsell task, carefully explored by Monsell (1978) because it enables discrimination among alternative theories, is the N-probe recency effect: Presumably, presentation of an item increases its strength in memory, which then gradually declines. To the extent that memory strength is used to decide about membership in the P-set, N-probes that were displayed more recently should therefore be harder to reject. As described in Section 2.2, this effect has been reported by Monsell (1978) and McElree and Doshier (1989), using the Monsell task, and by Gaffan (1977) with pictures, but not with words, using the Sv task.

In two experiments using the Sf task,  $n_{\text{neg}}$  was varied with  $n_{\text{pos}} = 1$  (Exp. Stern75b;  $n_{\text{neg}} = 1, 2, 4, \text{ and } 8$ ;  $N_{\text{sub}} = 8$ ) and  $n_{\text{pos}} = 2$  (Exp. Stern75a;  $n_{\text{neg}} = 2, 4, \text{ and } 6$ ;  $N_{\text{sub}} = 6$ ). In both experiments, positive and negative responses were required with equal frequency, and the  $n_{\text{neg}}$  N-probes were presented with approximately equal frequency.<sup>73</sup> The data shown in Table 6 omit the conditions in which  $n_{\text{pos}} = n_{\text{neg}}$ ; under these conditions, even if the same strategy were used as in the other conditions, there is no reason why the subject might not determine, even from trial to trial, which is the effective P-set.<sup>74</sup> For a probe,  $\Delta\text{trials}$  is the inverse of its recency: the number of trials since it

<sup>72</sup>In the Monsell task, a brief filled delay was found to reduce  $\bar{\beta}$  from 42.2 to 30.5 ms/item (Exp. Monsl78.1). According to Monsell (1978, p. 481), for direct-access memory-strength models, “if decay [of strength] decelerates ... then the effective rate of decay, and hence the slope, will be smaller after a delay.” Could such a dramatic difference—from reducing the slope in his experiment to doubling it in this one—be produced by plausible adjustments of strength criteria in such models? It seems unlikely.

<sup>73</sup>In each of the four parts of Exp. Stern75b, 100 test trials were preceded by 40 practice trials. The mean error rate was 2.4%. In each of the three parts of Exp. Stern75a, 120 test trials were preceded by 48 practice trials. The mean error rate was 2.9%. In both experiments, different P-sets were used in different parts.

<sup>74</sup>Evidence favouring this possibility is provided by  $\overline{RT}_{\text{neg}} - \overline{RT}_{\text{pos}}$ , usually between 20 and 40 ms when P-trials and N-trials are equiprobable, which was 2, 32, 28, and 36 ms for  $n_{\text{neg}} = 1, 2, 4, \text{ and } 8$ , respectively, in Exp. Stern75b, and 12, 36, and 32 ms for  $n_{\text{neg}} = 2, 4, \text{ and } 6$ , respectively, in Exp. Stern75a.

**Table 6.** *Effects of probe recency on mean RT.*

Probe Type	Experiment	Stern75b ( $n_{pos} = 1$ )			Stern75a ( $n_{pos} = 2$ )		
		$n_{neg}$	2	4	8	4	6
Negative	Mean $\Delta trials$		3.9	7.9	16.3	8.3	11.7
	Pr{ $\Delta trials = 1$ }		0.23	0.05	0.00	0.07	0.05
	Mean RT (ms)		449.7	434.5	444.8	494.7	491.8
Positive	Mean $\Delta trials$		2.0	2.0	2.0	3.9	4.0
	Pr{ $\Delta trials = 1$ }		0.43	0.43	0.40	0.33	0.33
	Mean RT (ms)		417.6	406.8	408.9	459.0	460.2

was last presented. As can be seen, the mean  $\Delta trials$  for N-probes increased markedly with an increase in  $n_{neg}$ , especially in Exp. Stern75b, where the increase was by a factor of 4. And the proportion of trials for which  $\Delta trials$  was 1 (an immediate repetition) declined from about a quarter of the trials to none, in that experiment.

These experiments show that in the Sf task, changes in probe recency per se have at most negligible effects on RT. In Exp. Stern75b, the mean increase in  $\overline{RT}_{neg}$  per mean unit of  $\Delta trials$  was  $-0.2 \pm 1.0$  ms, and per N-set item it was  $-0.3 \pm 2.0$  ms. In Exp. Stern75a, these values were  $-0.9 \pm 2.3$  ms and  $-1.5 \pm 3.8$  ms, respectively.<sup>75</sup> This conclusion is supported by Hawkins and Hosking (1969) and Biederman and Stacy (1974), who found effects of relative frequency among P-probes that were negligible among N-probes. On the other hand, if we wish to attribute the difference between experiments in RTs for P-probes to their difference in recency, the data require a vastly greater effect of recency: The difference in the mean  $\Delta trials$  for P-probes from 2 to 4 trials across experiments indicates an increase in  $\overline{RT}_{pos}$  per mean trial of  $25.1 \pm 11.7$  ms. In Appendix D, further evidence about recency and frequency effects in the Sf task is described, along with a description of variants of the Sf task in which such effects are prominent.

### 9.1. The Sf task as association or category learning

Should the Sf task be regarded as learning a set of associations between items and responses (as proposed by Theios et al., 1973) or as practice in categorizing items as “positive” or “negative” (e.g., Nosofsky & Alfonso-Reese, 1999; Nosofsky & Palmeri, 1997)? The asymmetry between effects of  $n_{pos}$  and  $n_{neg}$  suggests not. As  $n_{neg}$  is increased, with positive and negative trials equally frequent, the subject has fewer opportunities to learn and practise assigning a category to each item in the N-set, or to strengthen its association with a response. Yet increasing  $n_{neg}$  does not cause  $\overline{RT}_{neg}$  to increase, as shown above. Also, with ensemble size and  $Pr\{pos\}$  fixed, an increase in  $n_{pos}$  is associated with a decrease in  $n_{neg}$ , and thus, for each N-set (P-set) member, more (fewer) opportunities to practise responding: Despite this,  $\beta_{pos}$  is not systematically greater than  $\beta_{neg}$ . (See, e.g., Figure 4.) Instead, this asymmetry of the effects of  $n_{pos}$  and  $n_{neg}$  makes it plausible that negative responses are made “by default”, when the probe fails to match any member of the P-set, a conclusion also supported by the finding of effects of relative frequency differences among P-probes but not among N-probes mentioned above (see Appendix D).

<sup>75</sup>The absence of an effect of  $n_{neg}$  was described in Sternberg (1963) and mentioned in Sternberg (1966); the data are shown in Figure 3 in Sternberg (1975). This finding appears not to have been considered by advocates of strength-based theories of performance in this task (e.g., Monsell, 1978 and McElree & Doshier, 1989, who believe that recency across trials is one determinant of RT, and Hockley & Murdock, 1987). Nor have these findings been considered by those who argue that repetition priming plays a role in the fixed-set procedure (Jou, 2014; Stadler & Logan, 1989).



## 10. VARIATION OF RELATIVE RESPONSE FREQUENCIES AND EVIDENCE-ACCUMULATION MODELS

Can data from Sv or Sf tasks be explained by evidence-accumulation models such as the diffusion model (Ratcliff, 1978), or the linear ballistic accumulator (LBA) model (Brown & Heathcote, 2008)? Van Vugt, Beulen, and Taatgen (2016) pointed out that the bulk of the neurophysiological studies in monkeys and humans that favour such models involve perceptual decision making (Heekeren, Marrett, & Ungerleider, 2008), and, using intracranial recordings in humans, they were unable to find corresponding neural support for such models applied to the Sv task. Nonetheless, it seems worth asking whether behavioural data from the Sf or Sv task might be consistent with such models. In the LBA model, the major parameters are the amount of evidence  $e_j$  required for a decision, and the rate,  $r_k$  at which it accumulates. In this connection it is interesting to consider how the effects of relative response frequency ( $Pr\{pos\}:Pr\{neg\}$ ) and size of the P-set ( $n_{pos}$ ) combine.

In the two cases in which the diffusion model has been fitted to data from Sf or Sv tasks, an increase in  $n_{pos}$  was interpreted mainly as a reduction in the “relatedness” or “resonance” of matches, and hence a reduction in  $r_k$  (“drift rate”), at least for positive responses.<sup>76</sup> In contrast, for both LBA and diffusion models, the response bias induced by differential response proportions has been represented by the amount of evidence

required: For a less frequent response, the evidence threshold, or the starting point for evidence accumulation, or both (hence,  $e_j$ ), are suitably adjusted. (Donkin, Brown, & Heathcote, 2011; Forstmann, Brown, Dutilh, Neumann, & Wagenmakers, 2010; Leite & Ratcliff, 2011; Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012; Ratcliff & McKoon, 2008, Exp. 3<sup>77</sup>). In none of these applications was the effect of differential response proportions attributed to a post-decision process (“translation & response organization”), as in the model described in Sternberg (1969b, Figure 6), based on the additivity of this effect with  $n_{pos}$  and with response type (Sternberg, 1969b, Figures 4D and 4E).

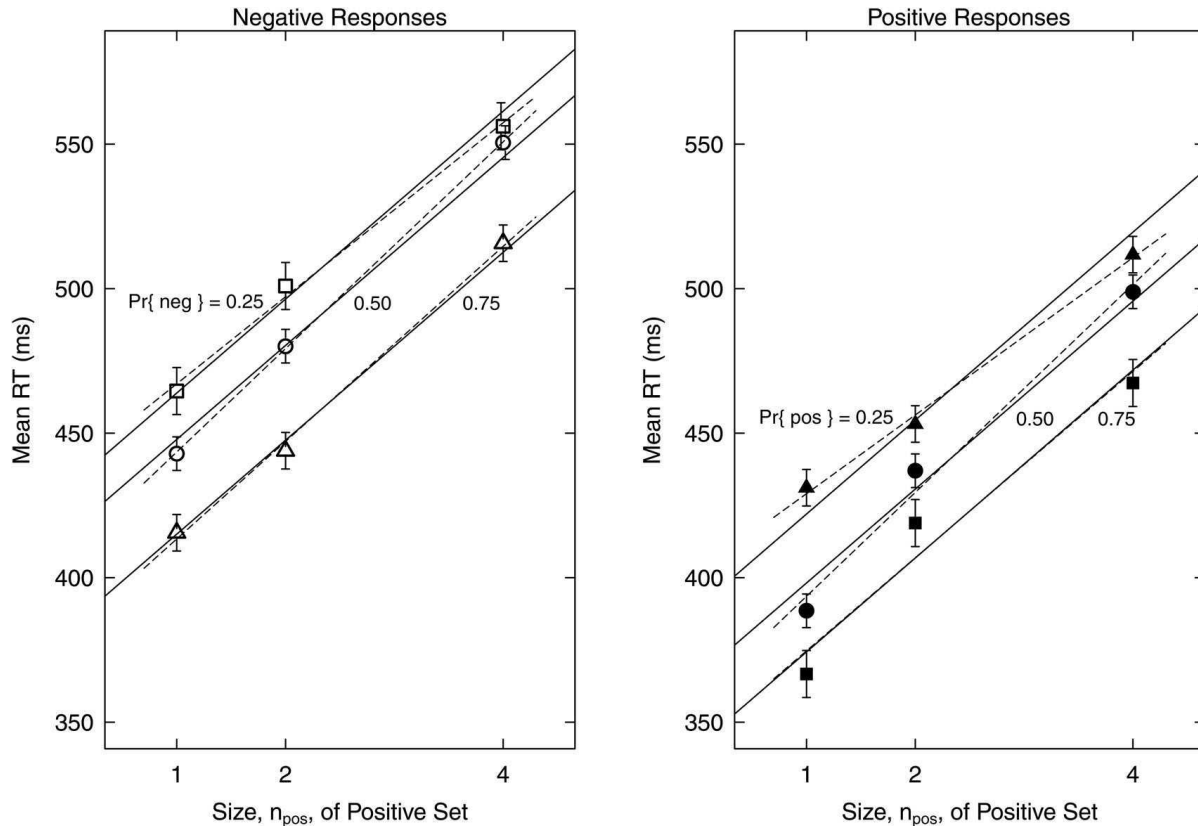
Let us idealize these findings, and assume that  $n_{pos}$  influences only the rate of evidence accumulation  $r_k$ , and that  $Pr\{response\}$  influences only the amount of evidence required for a response,  $e_j$ . We are then led to two predictions for an experiment in which  $Pr\{pos\}$  and  $n_{pos}$  are varied factorially. Specifically, as  $Pr\{pos\}$  increases, so that less evidence is required on P-trials and more on N-trials,

*Prediction 1:*  $\overline{RT}_{pos}$  will shrink and  $\overline{RT}_{neg}$  will grow; hence their difference,  $\overline{RT}_{neg} - \overline{RT}_{pos}$ , will grow.

*Prediction 2:* The effect of slowing the rate of evidence accumulation (by increasing  $n_{pos}$ ) on  $\overline{RT}_{pos}$  will shrink, while its effect on  $\overline{RT}_{neg}$  will grow. That is,  $\beta_{pos}$  will shrink, while  $\beta_{neg}$  will grow, and hence their difference,  $\beta_{neg} - \beta_{pos}$  will grow.

<sup>76</sup>In a review of the diffusion model, Ratcliff and McKoon (2008, p. 876) say “For recognition memory, for example, drift rate would represent the quality of the match between a test word and memory.” In his analysis of Exp. Hock84 (an Sv task with letters), Ratcliff (1988) found that the primary effect of the increase in  $n_{pos}$  from 3 to 6 was a systematic reduction in  $r_k$  for matches (from .405 to .294), and a smaller reduction in  $r_k$  for non-matches. Also, the separation between the starting point and the “yes” boundary (hence,  $e_j$ ) increased slightly but systematically. In their “VM blocked” condition in their Sf task with words, Strayer and Kramer (1994, Exp. 2) found that the primary effect of the increase in  $n_{pos}$  from 2 to 6 was a reduction in  $r_k$  for matches from .362 to .267, and a negligible effect on  $r_k$  for non-matches. They also found an increase in the separation between starting point and match boundary from .037 to .063. Ratcliff’s (1978) earlier findings were similar, but because he used the Monsell task rather than the Sv task, they are not relevant. In contrast to these findings, Donkin and Nosofsky (2012) found, in fitting a version of the parallel self-terminating LBA model to their data from three subjects in an Sv task (Exp. Donk12), that  $n_{pos}$  influenced the  $e_j$  as well as the  $r_k$ .

<sup>77</sup>Applying the diffusion model to an experiment in which the relative proportion of two responses was varied, Ratcliff and McKoon (2008, p. 899) concluded that “[a] difference in starting point accounted for most of the proportion effect” and that fitting an effect of proportion on the drift rate as well “increased the chi square goodness of fit value by only 1%”.



**Figure 6.** Mean RTs from an experiment in which  $Pr\{pos\}$  (between subjects) and  $n_{pos}$  (within subjects) were varied. Broken lines are least-squares fits to the functions relating  $RT_{neg}$  (Panel A) and  $RT_{pos}$  (Panel B) to  $n_{pos}$  for the three groups of 12 subjects. Equations of the fitted lines for N-trials (Panel A) are  $437 + 30.1n_{pos}$ ,  $408 + 35.8n_{pos}$ , and  $380 + 33.7n_{pos}$ , for  $Pr\{neg\} = 0.25, 0.50,$  and  $0.75$ , respectively. Equations of the fitted lines for P-trials (Panel B) are  $342 + 32.2n_{pos}$ ,  $358 + 36.0n_{pos}$ , and  $402 + 27.2n_{pos}$ , for  $Pr\{pos\} = 0.25, 0.50,$  and  $0.75$ , respectively. The six unbroken lines have equal fitted slopes ( $32.5$  ms/item) and individually fitted intercepts. Standard errors are based on residuals in ANOVAs.

In an experiment (Exp. Stern69b.4) to examine the question of how the effects of  $n_{pos}$  and  $Pr\{pos\}$  combine, three groups of 12 subjects each were run in the Sf task with digits, each subject with  $n_{pos}$  values of 1, 2, and 4, and each group with a different relative frequency of positive and negative responses: .25/.75, .50/.50, and .75/.25. For each subject and each

of the three  $n_{pos}$  values, 40 practice trials were followed by 100 test trials, in blocks of 20.<sup>78</sup> The stimulus ensemble consisted of the ten digits; the N-set was the complement of the P-set. Accuracy was high in all conditions; the mean error rate was 1.4%.<sup>79</sup> Mean RTs are plotted in Figure 6; summary data are shown in Table 7.

<sup>78</sup>In retrospect, a full practice session should have been provided, to reduce variability. Because of the importance of knowing how the effects of these two factors combine, a better experiment, perhaps with both factors varied within subjects and with more than three values of  $n_{pos}$ , should be run.

<sup>79</sup>Mean error rates were similar for P-trials and N-trials and differed little across  $n_{pos}$  values: 1.3%, 1.5%, and 1.4% for values 1, 2, and 4, respectively. However, they did vary with response probability: For low, medium, and high probabilities, mean error rates were 2.4%, 1.2%, and 0.6%, respectively—not surprisingly, subjects tended to make the high-probability response when the low-probability response was called for, more than the reverse. Means of  $var(RT)$  differed little between P-trials and N-trials: 7,021 and 6,995  $ms^2$ , respectively. Also, they were influenced little by response probability: 6,907, 6,903, and 7,214  $ms^2$  for low-, medium-, and high-

Given the predictions above, these data invite several comparisons: First, consider the effects on  $\overline{RT}$ , in the top part of Table 7. As expected, increasing the proportion of trials that call for a response causes the  $\overline{RT}$  for that response to shrink. The effects, which are almost identical for P-trials and N-trials, are substantial, showing that subjects are sensitive to the  $Pr\{pos\}$  manipulation.<sup>80</sup> Next, consider whether and how response probability modulates the effects of  $n_{pos}$ , as shown by  $\beta_{pos}$  and  $\beta_{neg}$  in the bottom part of the table. As  $Pr\{pos\}$  is increased from 0.25 to 0.75,  $\beta_{pos}$  does not shrink, as predicted; the increase, by 5.0 ms/item, is not statistically significant, but rules out a large decrease. Similarly, as  $Pr\{neg\}$  decreases from 0.75 to 0.25,  $\beta_{neg}$  does not grow, as predicted; the decrease, by 3.6 ms/item, is not statistically significant, but rules out a large increase. Finally, consider the within-subject difference between  $\beta_{neg}$  and  $\beta_{pos}$ . As  $Pr\{pos\}$  increases from .25 to .75, this difference does not grow, as predicted; the decrease, by 8.6 ms/item, is not statistically significant, but rules out a large increase.

What slope differences would we expect from an evidence-accumulation model? Consider this question first for  $\bar{\beta}_{pos}$ , in relation to a model in the spirit of the LBA, in which evidence grows linearly (faster for smaller  $n_{pos}$ ), until it reaches a threshold (lower for higher-probability responses). Ignore the data for intermediate conditions  $Pr\{pos\}/Pr\{neg\} = .50/.50$  and  $n_{pos} = 2$ . Writing  $\overline{RT} = \overline{RT}(Pr\{pos\}, n_{pos})$ , the relevant means are  $\overline{RT}(.75, 1) = 366.7$  ms,  $\overline{RT}(.75, 4) = 467.4$  ms,  $\overline{RT}(.25, 1) = 431.1$  ms, and  $\overline{RT}(.25, 4) = 511.8$  ms.

Given mean effects of  $n_{pos}$  and  $Pr\{pos\}$  on  $\overline{RT}$ , together with the mean duration  $\bar{\tau}$  of the decision process, it can be shown that the predicted effect

on  $\beta_{pos}$  of the change of conditions from .25/.75 to .75/.25 is:

$$\Delta\beta_{pos} = \frac{\bar{\beta} \times \bar{\pi}}{\bar{\tau}}, \tag{4}$$

where  $\bar{\pi}$  is the mean effect of probability on  $\overline{RT}_{pos}$ , between .75 and .25 (54.4 ms),  $\bar{\beta}$  is the mean of the slopes of the  $\overline{RT}$  functions on P-trials (30.2 ms/item), and  $\bar{\tau}$  is the mean duration of the decision process averaged over the four conditions.<sup>81</sup>

To determine  $\bar{\tau}$ , we need to know how much of  $\overline{RT}$  is consumed by stimulus-encoding and response-output stages. In the diffusion model, the sum of their durations is the value of  $T_{er}$ . (The smaller  $T_{er}$ , hence the larger  $\bar{\tau}$ , the smaller the predicted slope difference.) In their review of 23 applications of the diffusion model, Matzke and Wagenmakers (2009) found the smallest estimate of  $T_{er}$  to be 206 ms. In his analysis of Hockley's (1984) Sv data, Ratcliff (1988) obtained mean  $T_{er} = 343$  ms. Let us use  $T_{er} = 206$  ms, the conservative choice. Averaging over  $n_{pos}$  values of 1 and 4,  $\overline{RT} = 444.2$  ms. Thus,  $\bar{\tau} = 444 - 206 = 238$  ms. Using Equation (4) we find the predicted value of  $\beta_{.75} - \beta_{.25}$  to be  $-6.9$  ms/item, different from the value observed for these reduced data, of  $+5.0$  ms/item.

Given the design of this experiment, the effect of  $Pr\{pos\}$  in the above analysis is a between-subjects effect, so that the precision of the difference between the predicted and observed values is low. An alternative is available if we note that the observed equality of the effects of  $Pr\{response\}$  and  $n_{pos}$  on P-trials and N-trials suggests that the same mechanism is at work on the two kinds of trial:

probability responses, respectively. However, they were influenced by  $n_{pos}$ , similarly for P-trials and N-trials: 5,607, 6,587, and 8,830 ms<sup>2</sup>, for  $n_{pos}$  values of 1, 2, and 4, respectively.

<sup>80</sup>As discussed in Sternberg (1969b), another way of expressing the equality of these effects on P-trials and N-trials is that the effects of response probability and response type (positive or negative) are additive, consistent with their selectively influencing distinct processes arranged in stages.

<sup>81</sup>To show this, let  $\tau$  = duration of the decision process,  $\gamma$  = half of the mean effect of  $n_{pos}$  (from 1 to 4) on  $\tau$ ,  $\pi$  = half of the mean effect of  $Pr\{response\}$  (from .75 to .25) on  $\tau$ , and  $\bar{\tau}$  = the mean of  $\tau$  over the four conditions, and write  $\tau$  as  $\tau(Pr\{response\}, n_{pos})$ . Then  $\tau(.75, 1) = (\bar{\tau} - \gamma)(\bar{\tau} - \pi)/\bar{\tau}$ ;  $\tau(.75, 4) = (\bar{\tau} + \gamma)(\bar{\tau} - \pi)/\bar{\tau}$ ;  $\tau(.25, 1) = (\bar{\tau} - \gamma)(\bar{\tau} + \pi)/\bar{\tau}$ ; and  $\tau(.25, 4) = (\bar{\tau} + \gamma)(\bar{\tau} + \pi)/\bar{\tau}$ . Next, combine these to determine the effects of  $n_{pos}$  on  $\tau$  (in this case, three times the slope of the  $\overline{RT}$  function) under the two conditions of response probability:  $\tau(.75, 4) - \tau(.75, 1)$  and  $\tau(.25, 4) - \tau(.25, 1)$ .

Table 7. Results of Exp. Stern69b.4, in which  $Pr\{pos\}$  and  $n_{pos}$  were varied factorially.

Condition	$Pr\{pos\}/Pr\{neg\}$			Effect
	.25/.75	.50/.50	.75/.25	
Measure				
$\overline{RT}_{pos}$	465 ± 13	442 ± 11	417 ± 13	-48 ± 18 ( $p = .02$ )
$\overline{RT}_{neg}$	458 ± 10	491 ± 14	507 ± 17	+49 ± 20 ( $p = .02$ )
$\overline{RT}_{neg} - \overline{RT}_{pos}$	-7.0 ± 6.6	49.7 ± 7.0	89.6 ± 7.3	+97 ± 10 ( $p < .001$ )
$\bar{\beta}_{pos}$	27.2 ± 3.0	36.0 ± 4.9	32.2 ± 4.5	+5.0 ± 5.4 qpred
$\bar{\beta}_{neg}$	33.8 ± 3.2	35.8 ± 3.0	30.1 ± 5.4	-3.6 ± 6.1
$\bar{\beta}_{neg} - \bar{\beta}_{pos}$	6.5 ± 2.3	-0.2 ± 3.1	-2.1 ± 4.3	-8.6 ± 5.3 qpred

Note. Measures are mean ± SE. An “effect” on a measure is its value for .75/.25 minus its value for .25/.75. The  $p$  values are based on the Welch two-sample  $t$  test. Values in the third and sixth rows are means of within-subject differences. The values for which quantitative predictions are derived from the evidence-accumulation model are marked “qpred”.

That is, for both responses, the probability effect is due to a difference in the response threshold, and the  $n_{pos}$  effect is due to a difference in the rate of evidence accumulation. Assuming that this is true, we can derive within-subject estimates of both the predicted and the observed values. Using Equation (4), and letting  $\beta_{.25}$  and  $\beta_{.75}$  be the slopes of the  $\overline{RT}$  functions for response probabilities of .25 and .75, respectively, averaged over P-trials and N-trials, note first that the predicted effect of the change of conditions from .25/.75 to .75/.25 on the quantity in the last row of Table 7 is

$$\Delta[\beta_{neg} - \beta_{pos}] = 2(\beta_{.25} - \beta_{.75}) = \frac{2\bar{\beta} \times \bar{\pi}}{\bar{\pi}} \quad (5)$$

Now, writing  $\overline{RT} = \overline{RT}(Pr\{response\}, n_{pos})$ , and averaging over RTs for P-trials and N-trials for corresponding (equal) response probabilities, we have  $\overline{RT}(.75,1) = 391.1$  ms,  $\overline{RT}(.75,4) = 491.5$  ms,  $\overline{RT}(.25,1) = 447.9$  ms, and  $\overline{RT}(.25,4) = 534.0$  ms.

To estimate the sampling error of the difference between observed and predicted values of  $2(\beta_{.25} - \beta_{.75})$ , without having to adjust for the effect of response type (positive or negative), I created 12

pseudo-subjects by averaging the RTs for  $Pr\{response\} = .25$  from each of the 12 subjects who provided .25/.75 data with the RTs for each of the corresponding subjects who provided .75/.25 data, and did the same for the  $Pr\{response\} = .75$  data for the two groups.<sup>82</sup> I assumed that  $T_{er} = 206$  ms for each pseudo-subject, which could then provide both an observed value and a prediction. The mean observed value is -9.5 ms/item, the mean predicted value is +11.9 ms/item, and the difference between them is  $21.4 \pm 5.2$  ms/item; a  $t$  test of this difference (with  $df = 11$ ) gives  $p = .002$ . If we assume that  $T_{er} = 0$ , so that the decision process occupies the full RT (a highly conservative assumption), the result is a lower bound for the mean predicted value of 6.6 ms/item, and a difference between predicted and observed values of  $16.2 \pm 5.1$  ms/item; in this case the  $t$  test gives  $p < .01$ , which is still statistically significant.

It is possible that the goodness of fit of the evidence-accumulation model to these data is influenced adversely by anomalous values when  $n_{pos} = 1$  (see below). For this reason, I repeated the above analysis using the data for  $n_{pos} = 2$  and 4.

<sup>82</sup>Because the two groups are independent, the assignment to create the pairs is arbitrary, but the result might depend on the assignment. For this reason I created 1000 random assignments and determined, for each, the observed value, the predicted value, and the difference between them. The values reported are the means of the thousand differences determined in this way. It turns out that the results were not especially sensitive to the assignment: While the mean difference between predicted and obtained values was 21.4 ms, the range of this difference was small (20.7, 22.1 ms).

In this analysis the mean observed value is  $-3.2$  ms/item, the mean predicted value is  $9.4$  ms/item, and the difference between them is  $12.6 \pm 5.6$  ms/item; a  $t$  test of this difference gives  $p = .05$ .

In conclusion, while subjects are sensitive to both  $n_{\text{pos}}$  and  $Pr\{pos\}$ , there is no support for the magnitude of the interaction of these two factors that we expect from this evidence-accumulation model and plausible interpretations of its parameters. Instead, the effects of the two factors are approximately additive, consistent with the idea that they influence distinct processing stages, possibly one for “memory scanning”, sensitive to  $n_{\text{pos}}$  but not to  $Pr\{pos\}$ , and one or more for decision and response-organization sensitive to the latter but not the former. However, more precise data are needed for a persuasive test of additivity: The interaction contrast of  $Pr\{pos\}$  and  $n_{\text{pos}}$  is  $14.3 \pm 7.4$  (based on  $n_{\text{pos}} = 1$  and 4) and  $3.2 \pm 6.2$  (based on  $n_{\text{pos}} = 2$  and 4). Neither is significant, but the first, at least, is uncomfortably large.

*Impossibility of selective influence.* Suppose we are confident of the additivity of the effects of two factors  $F_j$  and  $G_k$  on  $\overline{RT}$ , each factor at two levels ( $j = 1, 2; k = 1, 2$ ). Consider an evidence-accumulation model with two parameters:  $e_j > 0$ , the amount of evidence required, and  $r_k > 0$ , the rate of evidence accumulation. Then it is easy to show that neither factor can selectively influence just one of the parameters: In such a model,  $RT_{jk} = RT_0 + e_j/r_k$ . Assuming non-zero effects ( $e_2 \neq e_1$  and  $r_2 \neq r_1$ ), the additivity implies that  $(e_2 - e_1)/r_1 = (e_2 - e_1)/r_2$ , which requires  $r_2 = r_1$ , a contradiction. For effects of factors on  $\overline{RT}$  to be additive in such a model, any factor that influences one of the parameters must influence both.

Atkinson and Juola's (1974) hybrid model (Section 2.1) leads us to expect the same kind of interaction of the effects of  $n_{\text{pos}}$  and  $Pr\{pos\}$ , as follows: As  $Pr\{pos\}$  increases, we expect the familiarity criteria to be adjusted so that scanning occurs on a smaller proportion of P-trials and a larger proportion of N-trials. The result of these changes in the mixture distributions would be an increase in  $\bar{\beta}_{\text{neg}} - \bar{\beta}_{\text{pos}}$ , which is not what is observed.

## 10.1. The set-size one anomaly

Data from both Sv and Sf procedures occasionally show a deviation from a linear  $\overline{RT}$  function in which the  $\overline{RT}$  for P-trials with  $n_{\text{pos}} = 1$  falls below a fitted linear function (responses faster than they “should be”). One example is in the Sv data of Exp. Stern66.1 (Figure 3B). Why was this deviation absent from the Sf data of Exp. Stern66.2? In that experiment,  $Pr\{pos\}$  was low (0.27), which provides a clue. Results from the experiment described above provide a possible answer, showing that the deviation is sensitive to relative response frequency. To examine the deviation, I extrapolated linear functions fitted to the data for  $n_{\text{pos}} = 2$  and 4 to  $n_{\text{pos}} = 1$  for each of the 12 subjects in each response-frequency group, separately for P-trials and N-trials, and determined the deviation by subtracting the extrapolated value from the corresponding observed value for each subject. For data containing the anomaly, the deviation should be negative.

For responses on N-trials, there was no effect of response frequency: For the .25/.75, .50/.50, and .75/.25 groups, the numbers of subjects out of 12 who showed negative deviations were 6, 5, and 6, respectively. However, for responses on P-trials, the corresponding numbers were 4, 10, and 9, and the differences between the mean deviations for the .25/.75 and .75/.25 groups, 7.3 and  $-28.0$  ms, were large enough, despite considerable variability, so that a  $t$  test produced  $p < .05$ .

A possible explanation for the anomaly is that when a particular probe occurs on half or more of the trials, the benefits of preparing selectively for that probe outweigh the costs, so that subjects tend to prepare in that way. The present experiment was perhaps too insensitive to reveal any possible costs on N-trials.

## 11. SEQUENTIAL COMPARISON VERSUS SEQUENTIAL ACTIVATION

In this section I ask whether the effect of  $n_{\text{pos}}$  on  $\overline{RT}$  in the Sv and Sf tasks results from a process of sequential activation of members of the P-set,

rather than sequential comparison of those members to the probe, as originally proposed (Sternberg, 1963, 1966), a possibility also discussed by Corballis (1979) and Corballis and Miller (1973).

Consider the LIJ model (Section 3), which describes a continuous process of activation and reactivation of the P-set. Although the sequence of operations driven by the gamma oscillation is said to include comparison of a representation of the probe to a representation of each member of the P-set, when needed, no details are provided about how the comparisons are carried out. For example, it is not clear whether or how the similarity of an N-probe to one or more members of the P-set could have an effect on RT. If the effect of  $n_{\text{pos}}$  on  $\overline{RT}$  is due to its effect on the duration of an activation cycle, then comparisons of a probe to P-set members must occur either (a) at the same time as each member is activated, or (b), after the cycle is completed, simultaneously. If (a), the doubling of the slope of the  $\overline{RT}$  function described in Section 8 when retrieval from active and inactive memory are compared then becomes even more mysterious: Why should an activation process have to occur twice? However, if (b), so that the effect of  $n_{\text{pos}}$  on  $\overline{RT}$  reflects just an activation process, and the comparison process occurs afterwards, then we have an alternative explanation for  $\beta_{\text{pos}} \approx \beta_{\text{neg}}$ , the property used to argue that a hypothesized comparison process is exhaustive, hence seemingly inefficient.<sup>83</sup>

Whereas the time per item associated with an activation process can be influenced by the kind of items comprising the P-set, it should not be influenced by the number of its members to which a N-probe is similar. On the other hand, if a comparison process is sequential, and the duration of the comparison of the N-probe to a member of the P-set is increased when they are similar, the LIJ model would seem to require serious modification, because, as it stands, the rate

of the serial process in that model reflects the complexity of items in the P-set, and not the relation of the probe to those items.

Suppose a process of sequential comparison: If we recognize that each comparison is a stage, and call two of these stages **A** and **B**, with durations  $t_a$  and  $t_b$ , then we can apply the additive factor method (Sternberg, 1969b, 1998, 2001) by varying N-probe similarity so as to increase the difficulty of neither, or of **A**, or of **B**, or of both, and consider the structure of the four mean RTs that result. Additivity of the effects on  $t_a$  and  $t_b$  would favour the hypothesis that the process includes sequential comparison, not merely sequential activation.<sup>84</sup> Furthermore, if  $n_{\text{pos}}$  is varied, with  $n_{\text{pos}} > 2$ , the additivity of effects of similarity on  $t_a$  and  $t_b$  should still obtain, and, in addition, these effects should also be additive with the effect of  $n_{\text{pos}} > 2$ . And if we vary the number of set members  $n_{\text{sim}} \leq n_{\text{pos}}$  to which the N-probe is similar in the same way, then we should find that for a given  $n_{\text{pos}}$ ,  $\overline{RT}$  should increase linearly with  $n_{\text{sim}}$ .

An experiment to perform such tests must satisfy at least four requirements:

1. The task should be Sf or Sv.
2. The structure of the P-sets should be independent of the similarity manipulation. For example, the sets in which two members are similar to the probe should not have members that are more similar to each other than sets in which one or no members are similar to the probe.
3. At least two different P-set sizes (and preferably three) should be used, to permit determining whether the  $\overline{RT}_{\text{pos}}$  and  $\overline{RT}_{\text{neg}}$  functions for fixed  $n_{\text{sim}}$  are parallel, and whether the magnitude of the effect of  $n_{\text{pos}}$  on  $\overline{RT}$  also is consistent with the “standard” phenomenon for the stimulus ensemble being used. (With three P-

<sup>83</sup>The puzzling “translation effect”, investigated thoughtfully by Clifton and his associates (e.g., Clifton, Sorce, & Cruse, 1977; see Sternberg 1975, Section 7.1) would perhaps be less mysterious if it reflects activation of both the P-set and its translations, rather than comparison to both.

<sup>84</sup>If the process is indeed exhaustive, and we consider time to respond to P-probes as the similarity of the probe to two non-matching set members is varied in the same way as described for N-probes, then a similar data pattern, with effects of the same size, would be expected.

set sizes, linearity of the  $\overline{RT}$  function can also be assessed.)

4. The overall  $\overline{RT}$  and error rate should be consistent with the “standard” phenomenon for the types of item being used.

Although not a requirement, an additional desirable property is:

5. The number of items in the P-set,  $n_{\text{sim}} \leq n_{\text{pos}}$ , to which the probe is similar *in the same way* (e.g., with respect to the same dimension) should be varied over three levels (possible with  $n_{\text{pos}} \geq 2$ ), to test linearity (additivity) of the effects on  $\overline{RT}$ .<sup>85</sup>

These requirements are not difficult to achieve. Suppose items that can be characterized by their values,  $j = 1, 2, \dots$  and  $k = 1, 2, \dots$ , on two dimensions,  $d_1$  and  $d_2$ , so that an item can be described as  $(j, k)$ . Let two items of the P-set be  $(1, 1)$  and  $(2, 2)$ . Then, among possible N-probes,  $(3, 3)$  shares no values with set members, which we can symbolize by  $[0, 0]$ ;  $(1, 3)$  and  $(2, 3)$  each share a  $d_1$  value with (are similar to) one set member, symbolized  $[1, 0]$ ;  $(3, 1)$  and  $(3, 2)$  each share a  $d_2$  value with one set member, symbolized  $[0, 1]$ ; while  $(1, 2)$  and  $(2, 1)$  each share a value with

(are similar to) two set members, symbolized  $[1, 1]$ . The set can be increased in size by adding  $[0, 0]$  items such as  $(4, 4)$  and  $(5, 5)$ . Now, consider the interaction contrast  $I_x = (\overline{RT}_{[1,1]} - \overline{RT}_{[0,1]}) - (\overline{RT}_{[1,0]} - \overline{RT}_{[0,0]})$ . Townsend and Fific (2004) recognize the importance of this interaction contrast; they point out that we expect the two kinds of similarity to be additive ( $I_x = 0$ ) if comparisons are sequential, but that if comparisons occur in parallel we expect  $I_x < 0$  (a “negative interaction”).<sup>86,87,88</sup>

Surprisingly, there appears to be only one published experiment that uses the Sv or Sf task and satisfies requirements (1), (2), (3), and (4) above: the condensation task of Checkosky (1971).<sup>89</sup> In Checkosky’s experiment, the items were coloured geometric forms with two dimensions: colour and shape, such as red circle and blue square. The experiment satisfied requirements (2), (3), and (4) above, but not property (5). Averaging over the two test days, and over levels of probe-set similarity for negative responses, the slopes of the  $\overline{RT}$  functions for positive and negative responses were 29 and 32 ms, respectively, close to parallel, and within the range of values often found in Sv and Sf tasks. However, the function relating  $\overline{RT}$  for negative responses to the number of items with a feature that matched

<sup>85</sup>The prediction of additivity could fail if the similarity of a probe to a member of the set along a particular dimension changes the way in which that dimension is processed in subsequent comparisons.

<sup>86</sup>Note that  $I_x = 0$  if and only if the effect of the mean number of shared features is linear, because it is equivalent to  $\overline{RT}_{[1,1]} - (\overline{RT}_{[1,0]} + \overline{RT}_{[0,1]})/2 = (\overline{RT}_{[1,0]} + \overline{RT}_{[0,1]})/2 - \overline{RT}_{[0,0]}$ .

<sup>87</sup>Townsend and Fific (2004) also show that with a negative interaction ( $I_x < 0$ ), a more elaborate analysis based on the RT distributions can distinguish among alternative parallel processes.

<sup>88</sup>To test whether  $\overline{RT}$  is linear with the number of set members to which an N-probe is similar in the *same way* (i.e., with respect to the same dimension), at least two members of the P-set must have the same value on one of the dimensions, and for both dimensions to be relevant to the decision there should be additional members. An example of such a P-set is  $(1, 1)$ ,  $(1, 2)$ ,  $(3, 3)$ ,  $(4, 4)$ . Then the N-probes  $(5, 5)$ ,  $(3, 5)$ , and  $(1, 5)$ , symbolized  $[0, 0]$ ,  $[1, 0]$ , and  $[2, 0]$ , are similar with respect to the same dimension to 0, 1, and 2 members, respectively.

<sup>89</sup>The interesting Sv experiment by Townsend and Fific (2004), which used an ensemble of Serbian pseudowords with Serbian subjects and showed strong effects of similarity, satisfies neither requirement (2) nor (3). However, it was revealed by Yang, Fific, and Townsend (2014) that their experiment also included a condition with  $n_{\text{pos}} = 4$ , so that requirement (3) could be satisfied, given an adequate analysis. However, assuming these violations to be unimportant, the analysis shows that whether the comparison process is serial or parallel varies from subject to subject (one consistently serial, one consistently parallel) and the probe delay (the remaining three subjects serial with a delay of 0.7 s, parallel with a delay of 2 s). The experiment by Huesmann and Wocher (1976) used the novel-negatives task, in which N-probe words were presented only once during the experiment, which invites the use of strength discrimination. Chase and Calfee’s (1969) experiments did not satisfy requirement (2). Dick and Hochhaus’s (1975) subjects were extremely slow and inaccurate. Also, the attempt by Hockley and Corballis (1982, Exp. 2) satisfied neither requirement (1) (using a variant of the Sf task with an R–S interval of only 0.5 s) nor (2).

one of the probe's features, though roughly additive with  $n_{\text{pos}}$ , was distinctly non-linear, indicating a substantial positive interaction.<sup>90</sup>

A possibility consistent with Checkosky's (1971) conclusions is that rather than engaging in a sequence of comparisons of the probe with representations of whole items in the P-set, subjects search separately for the presence, among the features in the P-set, of each of the two features of the probe; only when both features (one on each of the two dimensions) are discovered to be present would the subject have to proceed further, to determine whether they are associated with the same item. Such a strategy might have been encouraged by the use of dimensions that were highly separable (Garner & Felfoldy, 1970), and by the subjects serving in a mixture of two tasks: the condensation task and another task that invited searching for single features.<sup>91</sup>

Thus, an adequate experiment has yet to be performed that uses variation in perceptual similarity to manipulate the time taken to reject an N-probe in an Sv or Sf task. Another approach would be to use variation in categorical similarity, as was done by Darley (1973; described by Atkinson et al., 1974, pp. 226–231), for a

different purpose. Darley's was a factorial experiment with P-sets containing digits and letters, in which one factor was the number of letters (1, 2, or 3) in the P-set, the other was the number of digits (1, 2, or 3) in the P-set, and the probe could be a digit or a letter.<sup>92</sup> The available data<sup>93</sup> are RT means over subjects, P-trials and N-trials, and letters and digits. Let  $n_{\text{pos}} = n_s + n_d$ , where  $n_s$  is the number of P-set members in the same category as the probe, and  $n_d$  is the number in the other category. The effects of  $n_s$  and  $n_d$  on  $\overline{RT}$  were both linear; the mean data are well described by  $\overline{RT} = \alpha + \beta_s n_s + \beta_d n_d$ . If the process is sequential comparison, and it takes longer to decide on a mismatch if probe and P-set member are in the same category, then we would expect  $\beta_s > \beta_d$ . The estimates are  $\hat{\beta}_s = 37.0$  and  $\hat{\beta}_d = 33.1$ ; the difference,  $\hat{\beta}_s - \hat{\beta}_d = 3.9 \pm 5.0$  ms/item, is consistent with serial comparison, but too small relative to variability to be conclusive. The adjusted  $R^2$  for a model in which  $\beta_s = \beta_d$  is 0.963, to be compared with the adjusted  $R^2$  of 0.961 for a model in which  $\beta_s$  and  $\beta_d$  are not constrained to be equal, again indicating that while suggestive, these data provide no persuasive evidence for sequential comparison.<sup>94</sup>

<sup>90</sup> $\overline{RT}$  for  $n_{\text{sim}} = 0, 1,$  and  $2$  was 510, 534, and 625 ms, respectively; the corresponding error percentages were 0.3%, 2.5%, and 7.8%.

<sup>91</sup>The ingenious experiments by Mewhort and Johns (2000) and Johns and Mewhort (2002, 2003) fail to satisfy requirements (2) or (3). With ensembles of coloured shapes, their subjects are substantially slower than those of Checkosky (1971), who received no feedback, perhaps because they received feedback only on accuracy. With ensembles of words, and  $n_{\text{pos}} = 4$ , their subjects may also be slower and less accurate than others. For example, Juola and Atkinson (1971), with  $n_{\text{pos}} = 4$  words, obtained  $\overline{RT} = 712$  ms and 0.3% errors; in their accuracy condition, with  $n_{\text{pos}} = 4$  words, Banks and Atkinson (1974) obtained  $\overline{RT} = 828$  ms and 1.3% errors (averaged over  $n_{\text{pos}}$  values of 2, 3, 4, 5, and 6). In contrast, averaging over Exps. 5 and 6 in Mewhort and Johns (2000), and Exp. 4 in Johns and Mewhort (2002), all using P-sets containing four words, the  $\overline{RT}$  was 931 ms and the error rate 3.4%. However, under their conditions, Johns and Mewhort show persuasively that subjects search for probe features in the P-set rather than comparing the probe as a whole to P-set members, just as Clifton and Gutschera (1971) showed that subjects sometimes engage in such "hierarchical search" when the stimulus ensemble consists of two-digit numbers and the "features" are the tens digit and the units digit.

<sup>92</sup>Also included in his experiment were "pure" P-sets, containing just letters or just digits. On trials with such P-sets, the category of the probe was the same as that of the P-set, so that subjects knew the category of the probe before it was presented, unlike trials with mixed P-sets, on which the category of the probe was uncertain. Data from pure trials are thus omitted from the present analysis. Also, it should be mentioned that Darley's experiment is best described as an Ashby task, as members of the P-set on each trial were displayed simultaneously, with the letters and digits in different columns.

<sup>93</sup>Values read from plot in Atkinson et al., 1974, Figure 21.

<sup>94</sup>Darley's data also provide three contrasts to answer the question more directly. Let  $\overline{RT}_{s,d}$  be the mean RT when  $n_s = s$  and  $n_d = d$ . Then, for  $n_{\text{pos}} = 3$ ,  $\overline{RT}_{2,1} = 602$  ms and  $\overline{RT}_{1,2} = 602$  ms; for  $n_{\text{pos}} = 4$ ,  $\overline{RT}_{3,1} = 638$  ms and  $\overline{RT}_{1,3} = 633$  ms; and for



Thus, an adequate experiment that uses N-probe similarity to discriminate between sequential activation and sequential comparison as responsible for the effect of  $n_{\text{pos}}$  on  $\overline{RT}$  has yet to be performed. It should be noted, however, that some of the challenges for the LIJ model (footnote 29) are also issues for the sequential activation account.

## 12. SOME OPEN QUESTIONS

Much could be gained by improving on some of the experiments discussed above, providing more practice, using within-subject comparisons for some questions, and retaining P-set and probe identities in the record of each trial. Some issues that seem especially intriguing are mentioned below.

### 12.1. Effect of positive set size on mean RT: comparison or activation?

What is perhaps the most important open question is described in Section 11, along with one way to answer it.

### 12.2. Do Sv and Sf tasks elicit the same process?

Although I have claimed that the same scanning process is elicited by the Sv and Sf tasks, based on similarities of their RT patterns, a within-subject task comparison is needed, which would also provide an opportunity to determine whether the difference in variability suggested by the values in Table 4 is real. If so, an experiment using the Sv task in which the same P-set is used in more than one separated group of, say, five successive trials, sometimes with the same probe, would permit determining the extent to which any variance difference is due to differences among P-sets of the same size, and the extent to which it is due to interference

from sets presented on the previous one or two trials.

### 12.3. Scanning rate, stimulus ensemble, and mixed sets

Cavanagh's discovery and its replications have made it clear that the scanning (or activation) rate depends systematically on the items in the stimulus ensemble. Presumably, each "complex" item (such as a colour name) takes longer to activate or compare than each "simple" item (such as a digit). Corresponding to this, and given the LIJ model, a more complex item is associated with a longer gamma period. How might we test this idea further? Consider how long it would take to scan a mixed list: Scanning time for a P-set of fixed size should increase by the same increment for each simple item that is replaced by a complex one, and, replacing, for example, two simple items by complex ones in P-sets of different size should have the same effect. In other words, the number of complex items should have a linear effect on RT that is additive with the effect of  $n_{\text{pos}}$ .

The design and analysis of such an experiment must take into account the possibility of selective search of just the items in the category of the probe (e.g., Naus, 1974; Naus et al., 1972). One way that may avoid this possibility is to order the items in each P-set so as to avoid either all the simple items or all the complex items appearing consecutively, and to insure that the subject maintains the P-set in the order in which it was presented. Results of the experiment by Darley (1973), described in Section 11), using "duplex target sets" that changed from trial to trial are promising: Even though the sets, consisting of subsets of letters and digits, were presented with the subsets separated (visually, so it was a variant of an Ashby task), subjects did not appear to search selectively. Clifton and Brewer (1976) found that even with an Sf task, there are conditions under which search of mixed lists is not selective.

---

$n_{\text{pos}} = 5$ ,  $\overline{RT}_{3,2} = 667$  ms and  $\overline{RT}_{2,3} = 653$  ms. The differences, whose mean implies that  $\beta_s - \beta_d = 3.2 \pm 2.0$  ms, provide evidence for sequential comparison that is again suggestive, but not conclusive.

#### 12.4. Is choice of process obligatory or optional?

To what extent do the details of a task constrain the process used by subjects to perform it? For example, when the choice of response can be based either on (a) discrimination of the probe's strength in memory, or on (b) serial comparison of the probe to the P-set, can the subject choose which process to use? Consider the Monsell task: Suppose that the conjecture by Monsell (1978, p. 496) and Diener (1988, p. 375) that forming a representation of the P-set that can be scanned takes more time than is permitted by that task. If so, we might conclude that strength discrimination is obligatory. One way to test such a conjecture is to compare task performance after different kinds of training. One could train either in the Sv task (experimental group) or the Monsell or novel negatives task (control group), and then test using the Monsell task with a high-accuracy requirement, comparing  $\overline{RT}$  functions, serial-position curves, and sizes of the N-probe recency effect.

With respect to the Ashby task, I suggested (Section 2.4) that a simultaneous visual display of members of the P-set invites but does not require the subject to scan a visual image representation, rather than the kind of representation used in the Sv task. Again, one could train either in the Sv task (experimental group) or the Ashby task (control group), and then test using the Ashby task, looking for effects on the  $\beta_{\text{neg}}/\beta_{\text{pos}}$  ratio, and on the serial-position curves.

A similar question can be asked about the Atkinson and Juola (1974) hybrid model (Section 2.1), in which a component of all RTs is the duration  $T_{\text{strength}}$  of an early stage in which the strength of the probe's representation in memory is evaluated. If a procedure is used in which strength discrimination is unlikely to be helpful, will the  $T_{\text{strength}}$  component be eliminated? In one such procedure, some N-probes could be made very familiar, as in Monsell (1978, Exp. 2), and high accuracy would be required, rendering strength discrimination unlikely to be helpful. If

this eliminates the strength-evaluation process, then there should be no N-probe recency effect. If so, then whether or not the zero intercepts of the  $\overline{RT}_{\text{pos}}$  and  $\overline{RT}_{\text{neg}}$  functions decreased would be evidence about whether the strength-evaluation process preceded the scanning process, or occurred in parallel with it.

#### 12.5. Predicted additivity of effects of $Pr\{pos\}$ and stimulus quality

The four-stage model for memory scanning described by Sternberg (1969b, Figure 6) was based on finding four factors: (1) stimulus quality, (2)  $n_{\text{pos}}$ , (3) response type, and (4)  $Pr\{pos\}$ , with approximately additive effects on  $\overline{RT}$  of five of the six pairs. Additivity of the effects of factors (1) and (4) has never been tested, but is required by the model. One way to manipulate  $Pr\{pos\}$  while controlling probe frequencies would be to use the Sv procedure.

#### 12.6. Retrieval from active versus inactive memory

Why should these two procedures, discussed in Section 8, be associated with such precise doubling of slope? It seems highly implausible that the same process occurs twice in the LTM memory condition. Are they different processes, but driven by the same oscillation? If so, in what way are they different? (One possibility is that the two processes in the LTM condition are activation and comparison.)

Given that the scanning rate depends on the stimulus ensemble, one test would ask whether the 2:1 slope ratio is maintained when we increase the AM slope by using more complex stimuli. If so, and if the AM slope is found to reflect comparison rather than activation (as discussed in Section 11), and the LTM slope reflects comparison preceded by activation, then it would be interesting to ask about the effects of varying the similarity of N-probes to P-set members, which should influence only the comparison process.<sup>95</sup>

<sup>95</sup>If we are able to influence only one of the two processes, and thus disrupt the 2:1 ratio, then having found such an exact 2:1 ratio becomes even more mysterious.

### 12.7. Tests of the LIJ model

Phenobarbital, a barbiturate that works by increasing the activity of the neurotransmitter GABA, slows the gamma oscillation (Insel, Patron, Hoang, Nematollahi, & Barnes, 2012; Whittington, Traub, & Jefferys, 1995). This suggests asking whether phenobarbital decreases the scanning rate, ideally while also measuring gamma oscillations either at the scalp, or in patients with implanted electrodes. The possibility that there is a “skip” process also influenced by the drug would need to be considered. Also, given that victims of MS display steeper  $\overline{RT}$  functions, it would be useful to know whether gamma frequencies in the relevant brain regions are lower for them.

Jensen and Lisman (1998) explain the exhaustiveness of the search by suggesting that responses can occur only at the trough of the theta oscillation, after all members of the P-set have been activated and compared. They do not argue that no responses of any kind can occur during other parts of the theta period. However, if the inhibition does apply to other responding, then their idea could be tested, by using an Sf task and introducing with low probability a special signal that occurs at a random time. Responses to this signal should then be delayed by some fraction of the theta period; if the process is adapting theta, then this period will depend on  $n_{\text{pos}}$ .

## 13. CONCLUSION

In hindsight, it is easy to recognize that the way the human mind represents information may differ, depending on what the information is, how it was presented, and how much time has elapsed since then, and that such representation differences might influence how that information is interrogated. Also, it should surely not be surprising that results may be influenced by availability of cues (such as level of activation of the internal representation of the probe) that permit more than one strategy to be used to perform a laboratory task.

However, following up the early work on high-speed memory scanning, and tests of the SES model using Sv and Sf tasks, some investigators overlooked these possibilities: They used experimental procedures (e.g., Monsell task, or Ashby task) that produced some results that were qualitatively consistent with SES (approximately linear and sometimes approximately parallel  $\overline{RT}$  functions), but others that were not (e.g., pronounced effects of recency, non-parallel  $\overline{RT}$  functions). Findings that these  $\overline{RT}$  functions and the obtained error rates differed *quantitatively* from those produced by Sv and Sf tasks seems to have been ignored. These oversights generated scepticism about SES and the conditions under which it occurs, and unnecessary controversy, which, I hope, this paper will help resolve.

Some investigators have offered alternative theories that explain the discrepant facts they uncovered, but in seeking evidence relevant to these alternatives they ignored published findings that would have been troublesome, such as the similarity of results from Sv and Sf tasks (Sections 2.2, 2.4, 5.2, 6.2), the absence of an effect of  $n_{\text{neg}}$  (Section 9), and the behaviour of the RT variance (Section 5) and of the RT minimum (Section 6).<sup>96</sup>

Another reason for scepticism about SES may have been beliefs about brain function in the 1960s and 1970s, which favoured slow and parallel processes rather than fast and sequential ones. These beliefs have now changed with recognition of the importance of brain oscillations at a range of rates, some even faster than what would be required to underlie high-speed memory scanning, and with the development of neurophysiological models of SES based on such oscillations (Sections 3, 4).

Misunderstandings about SES led to repeated claims that its predictions (about the minimum RT and the RT variance) were violated, claims supported by data that were questionable or that came from procedures other than the Sv or Sf task. These predictions, which were not recognized as

<sup>96</sup>Selective attention to evidence that favours the author's position is perhaps to be expected, given the considerable self-discipline required to follow Chamberlin's (1890) method of multiple working hypotheses.

depending on the validity of extensions of the SES model, and whose testing requires some care, turn out to be supported (Sections 5.2, 6.2, 6.3).

The two alternative explanations of greatest interest for the phenomena under discussion are (a) strength discrimination, in which judgement about a probe's membership in the P-set is based on the strength (or level of activation) of its directly accessible representation in memory, and (b) parallel comparisons embodied in evidence-accumulation models such as the diffusion model. Evidence against (a) is presented in Sections 7, 8, and 9, with evidence against a version of (b) in Section 10.

Remarkably, many questions remain about how people decide whether an item is contained in their active memory (Section 12).

## REFERENCES

- Amit, D. J., Sagi, D., & Usher, M. (1990). Architecture of attractor neural networks performing cognitive fast scanning. *Network: Computation in Neural Systems*, 1, 189–216.
- Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review*, 80, 417–438.
- Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, 84, 413–451.
- Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (in press). The discovery of processing stages: Extension of Sternberg's method. *Psychological Review*.
- Archibald, C. J., & Fisk, J. D. (2000). Information processing efficiency in patients with multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*, 22, 686–701.
- Ashby, F. G., Tein, J.-Y., & Balakrishnan, J. D. (1993). Response time distributions in memory scanning. *Journal of Mathematical Psychology*, 37, 526–555.
- Atallah, B. V., & Scanziani, M. (2009). Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. *Neuron*, 62, 566–577.
- Atkinson, R. C., Herrmann, D. J., & Wescourt, K. T. (1974). Search processes in recognition memory. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 193–238). Potomac, Md.: Erlbaum Assoc.
- Atkinson, R. C., Holmgren, J. E., & Juola, J. F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics*, 6, 321–326.
- Atkinson, R. D., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. Krantz, R. Atkinson, R. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (pp. 243–293). San Francisco: W. H. Freeman, 1974.
- Axmacher, N., Henseler, M. M., Jensen, O., Weinrich, I., Elger, C. E., & Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proceedings of the National Academy of Sciences*, 107, 3228–3233.
- Baddeley, A. D. (1990). *Human memory: Theory and practice*. Boston: Allyn & Bacon.
- Baddeley, A. D., & Ecob, R. J. (1973). Reaction time and short-term memory: Implications of repetition effects for the high-speed exhaustive scan hypothesis. *Quarterly Journal of Experimental Psychology*, 25, 229–240.
- Bahramisharif, A., Jensen, O., Jacobs, J., & Lisman, J. (2016). Serial representation of items during the maintenance of working memory at content-specific cortical sites. Manuscript submitted for publication.
- Banks, W. P., & Atkinson, R. C. (1974). Accuracy and speed strategies in scanning active memory. *Memory & Cognition*, 2, 629–636.
- Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, 13, 90–102.
- Bertelson, P., & Renkin, A. (1966). Reaction times to new versus repeated signals in a serial task as a function of response-signal time interval. *Acta Psychologica*, 25, 132–136.
- Biederman, I., & Stacy, E. W. Jr. (1974). Stimulus probability and stimulus set size in memory scanning. *Journal of Experimental Psychology*, 102, 1100–1107.
- Briggs, G. E., & Johnsen, A. M. (1972). On the nature of central processing in choice reactions. *Memory & Cognition*, 1, 91–100.
- Brown, H. L., & Kirsner, K. (1980). A within-subjects analysis of the relationship between memory span and processing rate in short-term memory. *Cognitive Psychology*, 12, 177–187.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57, 153–178.

- Browning, P. G. F., Baxter, M. G., & Gaffan, D. (2013). Prefrontal-temporal disconnection impairs recognition memory but not familiarity discrimination. *Journal of Neuroscience*, *33*, 9667–9674.
- Bunge, S. A., Ochsner, K. N., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (2001). Prefrontal regions involved in keeping information in and out of mind. *Brain*, *124*, 2074–2086.
- Burle, B., & Bonnet, M. (2000). High-speed memory scanning: A behavioral argument for a serial oscillatory model. *Cognitive Brain Research*, *9*, 327–337.
- Burrows, D., & Okada, R. (1971). Serial position effects in high-speed memory search. *Perception & Psychophysics*, *10*, 305–308.
- Burrows, D., & Okada, R. (1975). Memory retrieval from long and short lists. *Science*, *188*, 1031–1033.
- Buzsaki, G. (2006). *Rhythms of the brain*. New York: Oxford University Press.
- Casement, M. D., Broussard, J. L., Mullington, J. M., & Press, D. Z. (2006). The contribution of sleep to improvements in working memory scanning speed: A study of prolonged sleep restriction. *Biological Psychology*, *72*, 208–212.
- Cavanagh, J. P. (1972). Relation between the immediate memory span and the memory search rate. *Psychological Review*, *79*, 525–530.
- Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*, *15*, 92–96.
- Chase, W. G., & Calfee, R. C. (1969). Modality and similarity effects in short-term recognition memory. *Journal of Experimental Psychology*, *81*, 510–514.
- Checkosky, S. F. (1971). Speeded classification of multidimensional stimuli. *Journal of Experimental Psychology*, *87*, 383–388.
- Clifton, C., Jr. & Birenbaum, S. (1970). Effects of serial position and delay of probe in a memory scan task. *Journal of Experimental Psychology*, *86*, 69–76.
- Clifton, C. Jr., & Brewer, E. (1976). Partially selective search of memory for letters and digits. *Memory & Cognition*, *4*, 616–626.
- Clifton, C. Jr., & Gutschera, K. D. (1971). Hierarchical search of two-digit numbers in a recognition memory task. *Journal of Verbal Learning and Verbal Behavior*, *10*, 528–541.
- Clifton, C. Jr., Sorce, P., & Cruse, D. (1977). The translation effect in memory search. *Cognitive Psychology*, *9*, 1–30.
- Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General*, *123*, 354–373.
- Corballis, M. C. (1979). Memory retrieval and the problem of scanning. *Psychological Review*, *86*, 157–160.
- Corballis, M. C., Kirby, J., & Miller, A. (1972). Access to elements of a memorized list. *Journal of Experimental Psychology*, *94*, 185–190.
- Corballis, M. C., & Miller, A. (1973). Scanning and decision processes in recognition memory. *Journal of Experimental Psychology*, *98*, 379–386.
- Corbin, L., & Marquer, J. (2008). Effect of a simple experimental control: The recall constraint in Sternberg's memory scanning task. *European Journal of Cognitive Psychology*, *20*, 913–935.
- Corbin, L., & Marquer, J. (2009). Individual differences in Sternberg's memory scanning task. *Acta Psychologica*, *131*, 153–162.
- Corbin, L., & Marquer, J. (2013). Is Sternberg's memory scanning task really a short-term memory task? *Swiss Journal of Psychology*, *72*, 181–196.
- Cover, K. S., Vrenken, H., Geurts, J. J. G., van Oosten, B. W., Jelles, B., Polman, C. H., ... van Dijk, B. W. (2006). Multiple sclerosis patients show a highly significant decrease in alpha band interhemispheric synchronization measured using MEG. *NeuroImage*, *29*, 783–788.
- Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., ... Hitch, G. J. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General*, *132*, 113–132.
- Cowan, N., Wood, N. L., Wood, P. K., Keller, T. A., Nugent, L. D., & Keller, C. V. (1998). Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology: General*, *127*, 141–160.
- Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, *2*, 263–275.
- Darley, C. F. (1973). *Effects of memory load and its organization on the processing of information in short term memory* (Doctoral dissertation). Stanford University.
- Dick, R. B., & Hochhaus, L. (1975). Memory search as a function of phonological context. *Bulletin of the Psychonomic Society*, *5*, 256–258.
- Diener, D. (1988). Absence of the set-size effect in memory-search tasks in the absence of a preprobe delay. *Memory & Cognition*, *16*, 367–376.
- Donkin, C., Brown, S., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the linear ballistic

- accumulator. *Journal of Mathematical Psychology*, 55, 140–151.
- Donkin, C., & Nosofsky, R. M. (2012). The structure of short-term memory scanning: An investigation using response time distribution models. *Psychonomic Bulletin & Review*, 19, 363–394.
- Dosher, B. A., & McElree, B. (1992). Memory search. In L. R. Squire (Ed.), *Encyclopedia of learning and memory* (pp. 398–406). New York: Macmillan.
- Dosher, B. A., & Sperling, G. (1998). A century of human information-processing theory: Vision, attention, and memory. In J. Hochberg (Ed.), *Perception and cognition at century's end* (pp. 199–252). San Diego: Academic.
- Drew, M. A., Starkey, N. J., & Isler, R. B. (2009). Examining the link between information processing speed and executive functioning in multiple sclerosis. *Archives of Clinical Neuropsychology*, 24, 47–58.
- Ellis, S. H., & Chase, W. G. (1971). Parallel processing in item recognition. *Perception & Psychophysics*, 10, 379–384.
- Ells, J. G., & Gotts, G. H. (1977). Serial reaction time as a function of the nature of repeated events. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 234–242.
- Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, 6, 205–254.
- Forrin, B., & Cunningham, K. (1973). Recognition time and serial position of probed item in short-term memory. *Journal of Experimental Psychology*, 99, 272–279.
- Forstmann, B. U., Brown, S., Dutilh, G., Neumann, J., & Wagenmakers, E. J. (2010). The neural substrate of prior information in perceptual decision making: A model-based analysis. *Frontiers in Human Neuroscience*, 4. doi:10.3389/fnhum.2010.00040
- Fougnie, D., Zughni, S., Godwin, D., & Marois, R. (2015). Working memory storage is intrinsically domain specific. *Journal of Experimental Psychology: General*, 144, 30–47.
- Franklin, P. E., & Okada, R. (1983). Effect of reaction-time feedback on subject performance in the item-recognition task. *American Journal of Psychology*, 96, 323–336.
- Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., & Duzel, E. (2010). Theta-coupled periodic replay in working memory. *Current Biology*, 20, 606–612.
- Gaffan, D. (1977). Exhaustive memory-scanning and familiarity discrimination: Separate mechanisms in recognition memory tasks. *Quarterly Journal of Experimental Psychology*, 29, 451–460.
- Garner, W., & Felfoldy, G. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1, 225–241.
- Geisler, C., Dipa, K., Pastalkova, E., Mizuseki, K., Royer, S., & Buzsáki, G. (2010). Temporal delays among place cells determine the frequency of population theta oscillations in the hippocampus. *Proceedings of the National Academy of Sciences*, 107, 7957–7962.
- Hacker, M. J. (1980). Speed and accuracy of recency judgments for events in short-term memory. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 651–675.
- Hale, D. J. (1967). Sequential effects in a two-choice serial reaction task. *Quarterly Journal of Experimental Psychology*, 19, 133–141.
- Hawkins, H. L., & Hosking, K. (1969). Stimulus probability as a determinant of discrete choice reaction time. *Journal of Experimental Psychology*, 82, 435–440.
- Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience*, 9, 467–479.
- Helfrich, R. F., Schneider, T. R., Rach, S., Trautmann-Lengsfeld, S. A., Engel, A. K., & Herrmann, C. S. (2014). Entrainment of brain oscillations by transcranial alternating current stimulation. *Current Biology*, 24, 333–339.
- Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *The Quarterly Journal of Experimental Psychology Section A*, 56, 1307–1334.
- Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 598–615.
- Hockley, W. E. (2008). Memory search: A matter of time. In H. L. Roediger III (Ed.), *Cognitive psychology of memory*. Vol 2 of *Learning and Memory: A Comprehensive Reference*, (J. Byrne, Ed.). Oxford: Elsevier. pp. 417–444.
- Hockley, W. E., & Corballis, M. C. (1982). Tests of serial scanning in item recognition. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 36, 189–212.

- Hockley, W. E., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review*, *94*, 341–358.
- Horn, D., & Usher, M. (1992). Oscillatory model of short term memory. In J. E. Moody, S. J. Hanson, & R. P. Lippmann, (Eds.), *Advances in neural information processing systems 4* (pp. 125–132). San Mateo, CA: Morgan Kaufman.
- Huesmann, L. R., & Woosler, F. D. (1976). Probe similarity and recognition of set membership: A parallel-processing serial-feature-matching model. *Cognitive Psychology*, *8*, 124–162.
- Hulme, C., Newton, P., Cowan, N., Stuart, G., & Brown, G. (1999). Think before you speak: Pauses, memory search, and trace reintegration processes in verbal memory span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 447–463.
- Hyndman, R. J., & Fan, Y. (1996). Sample quantiles in statistical packages. *American Statistician*, *50*, 361–365.
- Insel, N., Patron, L. A., Hoang, L. T., Nematollahi, S., & Barnes, C. A. (2012). Reduced gamma frequency in the medial frontal cortex of aged rats during behavior and rest: Implications for age-related behavioral slowing. *Journal of Neuroscience*, *32*, 16331–16344.
- Jacobs, J., Hwang, G., Curran, T., & Kahana, M. J. (2006). EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision making. *NeuroImage*, *32*, 978–987.
- Jacobs, J., & Kahana, M. J. (2009). Neural representations of individual stimuli in humans revealed by gamma-band electrocorticographic activity. *Journal of Neuroscience*, *29*, 10203–10214.
- Jensen, O., & Lisman, J. E. (1996a). Novel lists of  $7 \pm 2$  known items can be reliably stored in an oscillatory short-term memory network: Interaction with long-term memory. *Learning & Memory*, *3*, 257–263.
- Jensen, O., & Lisman, J. E. (1996b). Theta/gamma networks with slow NMDA channels learn sequences and encode episodic memory: Role of NMDA channels in recall. *Learning & Memory*, *3*, 264–278.
- Jensen, O., & Lisman, J. E. (1996c). *An oscillatory model for the Cavanagh constancy of short-term memory*. Unpublished manuscript.
- Jensen, O., & Lisman, J. E. (1998). An oscillatory short-term memory buffer model can account for data on the Sternberg task. *Journal of Neuroscience*, *18*, 10688–10699.
- Jensen, O., & Lisman, J. E. (2005). Hippocampal sequence encoding driven by a cortical multi-item working memory buffer. *Trends in Neurosciences*, *28*, 67–72.
- Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *European Journal of Neuroscience*, *15*, 1395–1399.
- Johns, E. E., & Mewhort, D. J. K. (2002). What information underlies correct rejections in short-term recognition memory? *Memory & Cognition*, *30*, 46–59.
- Johns, E. E., & Mewhort, D. J. K. (2003). The effect of feature frequency on short-term recognition memory. *Memory & Cognition*, *31*, 285–296.
- Johns, E. F., & Mewhort, J. K. (2011). Serial-position effects for lures in short-term recognition memory. *Psychonomic Bulletin & Review*, *18*, 1126–1132.
- Johnson, N. L., & Rogers, C. A. (1951). The moment problem for unimodal distributions. *Annals of Mathematical Statistics*, *22*, 433–439.
- Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*, 193–224.
- Jou, J. (2014). Task-switching cost and repetition priming: Two overlooked confounds in the fixed-set procedure of the Sternberg paradigm and how they affect memory set-size effects. *Quarterly Journal of Experimental Psychology*, *67*, 1871–1894.
- Juola, J. F., & Atkinson, R. C. (1971). Memory scanning for words versus categories. *Journal of Verbal Learning and Verbal Behavior*, *10*, 522–527.
- Kaminski, J., Brzezicka, A., & Wrobel, A. (2011). Short-term memory capacity ( $7 \pm 2$ ) predicted by theta to gamma cycle length ratio. *Neurobiology of Learning and Memory*, *95*, 19–23.
- Kawasaki, M., Kitajo, K., & Yamaguchi, Y. (2014). Fronto-parietal and fronto-temporal theta phase synchronization for visual and auditory-verbal working memory. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.00200
- King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*, 203–210.
- Kirsner, K. (1972). Naming latency facilitation: An analysis of the encoding component in recognition reaction time. *Journal of Experimental Psychology*, *95*, 171–176.
- Klatzky, R. L., Juola, J. F., & Atkinson, R. C. (1971). Test stimulus representation and experimental

- context effects in memory scanning. *Journal of Experimental Psychology*, 87, 281–288.
- Klatzky, R. L., & Smith, E. E. (1972). Stimulus expectancy and retrieval from short-term memory. *Journal of Experimental Psychology*, 94, 101–107.
- Kornblum, S. (1969). Sequential determinants of information processing in serial and discrete choice reaction time. *Psychological Review*, 76, 113–131.
- Kornblum, S., Stevens, G. T., Whipple, A., & Requin, J. (1999). The effects of irrelevant stimuli: 1. The time course of stimulus-stimulus and stimulus-response consistency effects with Stroop-like stimuli, Simon-like tasks, and their factorial combinations. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 688–714.
- Kristofferson, M. W. (1972a). When item recognition and visual search functions are similar. *Perception & Psychophysics*, 12, 379–384.
- Kristofferson, M. W. (1972b). Effects of practice on character-classification performance. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 26, 54–60.
- Landauer, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills*, 15, 646.
- Lass, U., Lüer, G., Becker, D., Fang, Y., & Chen, G. (2004). Encoding and retrieval components affecting memory span: Articulation rate, memory search, and trace reintegration. In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition* (pp. 349–370). Mahwah, NJ: Erlbaum.
- Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision Making*, 6, 651–687.
- Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the non-existent problem of decay. *Psychological Review*, 122, 674–699.
- Lisman, J. (2010). Working memory: The importance of theta and gamma oscillations. *Current Biology*, 20, R490–R492.
- Lisman, J. E., & Idiart, M. A. P. (1995). Storage of  $7 \pm 2$  short-term memories in oscillatory subcycles. *Science*, 267, 1512–1515.
- Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron*, 77, 1002–1016.
- Lively, B. L. (1972). Speed/accuracy trade off and practice as determinants of stage durations in a memory-search task. *Journal of Experimental Psychology*, 96, 97–103.
- Lively, B. L., & Sanford, B. J. (1972). The use of category information in a memory-search task. *Journal of Experimental Psychology*, 93, 379–385.
- Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, 2, 174–207.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- Lüer, G., Lass, U., Becker, D., Fang, Y., Chen, G., & Wang, Z. (1998). Zum Einfluss von Belohnung auf die Geschwindigkeit von Suchprozessen im Kurzzeitgedächtnis [Effect of reward on the speed of memory scanning]. In K. C. Klauer & H. Westmeyer (Eds.), *Psychologische Methoden und soziale Prozesse* [Psychological methods and social variables] (pp. 352–371). Lengerich, Germany: Pabst Science Publishers.
- Mallows, C. I. (1956). Note on the moment problem for unimodal distributions when one or both terminals are known. *Biometrika*, 43, 224–227.
- Matzke, D., & Wagenmakers, J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, 16, 798–817.
- McElree, B., & Doshier, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, 118, 346–373.
- McElree, B., & Doshier, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General*, 122, 291–315.
- McNicol, D., & Stewart, G. W. (1980). Reaction time and the study of memory. In A. T. Welford (Ed.), *Reaction times* (pp. 253–307). London: Academic Press.
- Mewhort, D. J. K., & Johns, E. E. (2000). The extralist-feature effect: Evidence against item matching in short-term recognition memory. *Journal of Experimental Psychology: General*, 129, 262–284.
- Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed-accuracy decomposition. *Psychological Review*, 95, 183–237.
- Miller, J. O., & Pachella, R. G. (1973). Locus of the stimulus probability effect. *Journal of Experimental Psychology*, 101, 227–231.
- Miller, J. O., & Pachella, R. G. (1976). Encoding processes in memory scanning tasks. *Memory & Cognition*, 4, 501–506.



- Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, *10*, 465–501.
- Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience*, *32*, 2335–2343.
- Naus, M. J. (1974). Memory search of categorized lists: A consideration of alternative self-terminating search strategies. *Journal of Experimental Psychology*, *102*, 992–1000.
- Naus, M. J., Glucksberg, S., & Ornstein, P. A. (1972). Taxonomic word categories and memory search. *Cognitive Psychology*, *3*, 643–654.
- Nickerson, R. S. (1966). Response times with a memory-dependent decision task. *Journal of Experimental Psychology*, *72*, 761–769.
- Nistico, R., Mango, D., Mandolesi, G., Piccinin, S., Berretta, N., Pignatelli, M., ... Centonze, D. (2013). Inflammation subverts hippocampal synaptic plasticity in experimental multiple sclerosis. *PLoS One*, *8*, e54666. doi:10.1371/journal.pone.0054666
- Nosofsky, R. M., & Alfonso-Reese, L. A. (1999). Effects of similarity and practice on speeded classification response times and accuracies: Further tests of an exemplar-retrieval model. *Memory & Cognition*, *27*, 78–93.
- Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*, 280–315.
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.
- Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 948–957.
- Ollman, R. (1966). Fast guesses in choice reaction time. *Psychonomic Science*, *6*, 155–156.
- Pashler, H., & Bayliss, G. (1991). Procedural learning: 2. Intertrial repetition effects in speeded-choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 33–48.
- Puckett, J. M., & Kausler, D. H. (1984). Individual differences and models of memory span: A role for memory search rate? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 72–82.
- Puffe, M. (1990). Quantized speed-capacity relations in short-term memory. In H.-G. Geissler (Ed.), *Psychophysical explorations of mental structures* (pp. 290–302). Lewiston, NY: Hogrefe & Huber.
- Ramsay, B., Bitschnau, W., Schmidhuber-Eiler, B., Berger, W., Karamat, E., Poewe, W., & Kemmler, G. W. (1990). Slowing of high-speed memory scanning in Parkinson's disease is related to the severity of parkinsonian motor symptoms. *Journal of Neural Transmission - Parkinson's Disease and Dementia Section*, *2*, 265–275.
- Rao, S. M., St. Aubin-Faubert, P., & Leo, G. J. (1989). Information processing speed in patients with multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, *11*, 471–477.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
- Ratcliff, R. (1988). A note on mimicking additive reaction time models. *Journal of Mathematical Psychology*, *32*, 192–204.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.
- Remington, R. J. (1969). Analysis of sequential effects on choice reaction times. *Journal of Experimental Psychology*, *82*, 250–257.
- Rizzuto, D. S., Madsen, J. R., Bromfield, E. B., Schulze-Bonhage, A., Seelig, D., Aschenbrenner-Scheibe, R., & Kahana, M. J. (2003). Reset of human neocortical oscillations during a working memory task. *Proceedings of the National Academy of Sciences*, *100*, 7931–7936.
- Roeber, U., & Kaernbach, C. (2004). Memory scanning beyond the limit – If there is one. In C. Kaernbach, E. Schr[o]lger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition* (pp. 371–388). Mahwah, NJ: Erlbaum.
- Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: Alpha-gamma versus theta-gamma codes for distinct WM information? *Trends in Cognitive Sciences*, *18*, 16–25.
- Scarborough, D. L. (1972). Memory for brief visual displays of symbols. *Cognitive Psychology*, *3*, 408–429.
- Schall, J. D. (2003). Neural correlates of decision processes: Neural and mental chronometry. *Current Opinion in Neurobiology*, *13*, 182–186.
- Schall, J. D., Purcell, B. A., Heitz, R. P., Logan, G. D., & Palmeri, T. J. (2011). Neural mechanisms of saccade target selection: Gated accumulator model of the visual-motor cascade. *European Journal of Neuroscience*, *33*, 1991–2002.

- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1–66.
- Schneider-Garces, N. J., Bordon, B. A., Brumback-Peltz, C. R., Shin, E., Lee, Y., Sutton, B. P., ... Fabiani, M. (2009). Span, CRUNCH, and beyond: Working memory capacity and the aging brain. *Journal of Cognitive Neuroscience*, 22, 655–669.
- Schon, K., Newmark, R. E., Ross, R. S., & Stern, C. E. (2016). A working memory buffer in parahippocampal regions: Evidence from a load effect during the delay period. *Cerebral Cortex*, 26, 1965–1974.
- Schweickert, R., Fisher, D. L., & Sung, K. (2012). *Discovering cognitive architecture by selectively influencing mental processes*. Singapore: World Scientific.
- Sederberg, P. B., Kahana, M. J., Howard, M. W., Donner, E. J., & Madsen, J. R. (2003). Theta and gamma oscillations during encoding predict subsequent recall. *Journal of Neuroscience*, 23, 10809–10814.
- Siegel, M., Warden, M. R., & Miller, E. K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proceedings of the National Academy of Sciences*, 106, 21341–21346.
- Sigman, M., & Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during dual-task performance. *Journal of Neuroscience*, 28, 7585–7598.
- Smith, E. E. (1967). Effects of familiarity on stimulus recognition and categorization. *Journal of Experimental Psychology*, 74, 324–332.
- Smith, M. C. (1968). Repetition effect and short-term memory. *Journal of Experimental Psychology*, 77, 435–439.
- Smulders, S. F. A., Notebaert, W., Meijera, M., Crone, E. A., van der Molen, M. W., & Soetens, E. (2005). Sequential effects on speeded information processing: A developmental study. *Journal of Experimental Child Psychology*, 90, 208–234.
- Soetens, E. (1998). Localizing sequential effects in serial choice reaction time with the information reduction procedure. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 547–568.
- Stadler, M. A., & Logan, G. D. (1989). Is there a search in fixed-set memory search? *Memory & Cognition*, 17, 723–728.
- Sternberg, S. (1963). *Retrieval from recent memory: Some reaction time experiments and a search theory*. Paper presented at the meeting of the Psychonomic Society, Bryn Mawr, PA., August 1963.
- Sternberg, S. (1964). *Estimating the distribution of additive reaction-time components*. Paper presented at the meeting of the Psychometric Society, Niagara Falls, Canada, October 1964.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153, 652–654.
- Sternberg, S. (1967a). *Scanning a persisting visual image versus a memorized list*. Paper presented at the Eastern Psychological Association meeting, April. (Bell Laboratories Technical Memorandum 67-1221-3).
- Sternberg, S. (1967b). Retrieval of contextual information from memory. *Psychonomic Science*, 8, 55–56.
- Sternberg, S. (1967c). Two operations in character-recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics*, 2, 45–53.
- Sternberg, S. (1969a). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 57, 421–457.
- Sternberg, S. (1969b). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II*. *Acta Psychologica*, 30, 276–315.
- Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, 27, 1–32.
- Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science: Methods, models, and conceptual issues* (pp. 703–863). Cambridge, MA: MIT Press.
- Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica*, 106, 147–246.
- Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica*, 45, 175–197.
- Sternberg, S., Knoll, R. L., & Nasto, B. (1969). *Retrieval from long-term vs. active memory*. Paper presented at the annual meeting of the Psychonomic Society, St. Louis, Mo., November 1969. (Bell Laboratories Technical Memorandum MM 69-1221-20).
- Strayer, D. L., & Kramer, A. F. (1994). Strategies and automaticity: I. Basic findings and conceptual framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 318–341.
- Theios, J., Smith, P. G., Haviland, S. E., Traupmann, J., & Moy, M. C. (1973). Memory scanning as a serial self-terminating process. *Journal of Experimental Psychology*, 97, 323–336.
- Theios, J., & Walter, D. G. (1974). Stimulus and response frequency and sequential effects in memory

- scanning reaction times. *Journal of Experimental Psychology*, *102*, 1092–1099.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- Townsend, J. T., & Fific, M. (2004). Parallel versus serial processing and individual differences in high-speed search in human memory. *Perception & Psychophysics*, *66*, 953–962.
- Townsend, J. T., & Roos, R. N. (1973). Search reaction time for single targets in multiletter stimuli with brief visual displays. *Memory & Cognition*, *1*, 319–332.
- Treisman, M., Faulkner, A., Naish, P. L. N., & Brogan, D. (1990). The internal clock: Evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception*, *19*, 705–743.
- Van Vugt, M. K., Beulen, M. A., & Taatgen, N. A. (2016). Is there neural evidence for an evidence accumulation process in memory decisions? *Frontiers in Human Neuroscience*, *10*. doi:10.3389/fnhum.2016.00093
- Van Zandt, T., & Townsend, J. T. (1993). Self-terminating versus exhaustive processes in rapid visual and memory search: An evaluative review. *Perception & Psychophysics*, *53*, 563–580.
- Vergauwe, E., & Cowan, N. (2014). A common short-term memory retrieval rate may describe many cognitive procedures. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00126
- Vergauwe, E., & Cowan, N. (2015). Attending to items in working memory: Evidence that refreshing and memory search are closely related. *Psychonomic Bulletin & Review*, *22*, 1001–1006.
- Verguese, D. (1966). Trente-huit millisecondes pour se souvenir. *Le Monde*, #6765, October 13.
- Vosskuhl, J., Huster, R. J., & Herrmann, C. S. (2015). Increase in short-term memory capacity induced by down-regulating individual theta frequency via transcranial alternating current stimulation. *Frontiers in Human Neuroscience*, *10*. doi:10.3389/fnhum.2015.00257
- Whittington, M. A., Traub, R. D., & Jefferys, J. G. R. (1995). Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature*, *373*, 612–615.
- Wickens, D. D., Moody, M. J., & Dow, R. (1981). The nature and timing of the retrieval process and of interference effects. *Journal of Experimental Psychology: General*, *110*, 1–20.
- Wickens, D. D., Moody, M. J., & Vidulich, M. (1985). Retrieval time as a function of memory set size, type of probes, and interference in recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *11*, 154–164.
- Yang, H., Fific, M., & Townsend, J. T. (2014). Survivor interaction contrast wobble predictions of parallel and serial models for an arbitrary number of processes. *Journal of Mathematical Psychology*, *58*, 21–32.
- Yellott, J. I. Jr. (1971). Correction for fast guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology*, *8*, 159–199.
- Zarahn, E., Rakitin, B. C., Abela, D., Flynn, J., & Stern, Y. (2006). Distinct spatial patterns of brain activity associated with memory storage and search. *NeuroImage*, *33*, 794–804.
- Zysset, S., & Pollmann, S. (1999). Retrieval from secondary memory: Detailed analysis of process components. *European Journal of Cognitive Psychology*, *11*, 87–104.

APPENDIX A

Experiments: Abbreviations, sources, and sections where mentioned

<i>Abbrev</i>	<i>Source</i>	<i>Sections</i>					
Ashby93	Ashby et al. (1993)	2.1	2.4	3.2	F		
Donk12	Donkin and Nosofsky (2012)	2.2	7	10			
Frank83	Franklin and Okada (1983) Experiments 1 & 2	2.1	2.4	F			
Hock84	Hockley (1984), Exp. 1, Memory search task	2.2	5.2	6.2	10	F	
Jacbs06	Jacobs, Hwang, Curran, and Kahana (2006)	2.2					
Knoll69	Sternberg, Knoll, and Nasto (1969)	8					
McEl89.p	McElree and Doshier (1989), Pilot experiment	2.1	2.2	2.3	7		
McEl89.2	McElree and Doshier (1989), Exp. 2	2.1	2.2	2.3	7		
Monsl78.1	Monsell (1978), Exp. 1	2.1	2.2	7	8.4	F	
Monsl78.2	Monsell (1978), Exp. 2	2.1	2.2	7	F		
Nosof11	Nosofsky et al. (2011), Exp. 2	2.1	2.2	7			
Rcliff78	Ratcliff (1978), Exp. 2	2.1	2.2	F			
Schnei77	Schneider and Shiffrin (1977), Exp. 2, Varied Mapping	2.1	2.4	F			
Stern66.1	Sternberg (1966), Exp. 1	2.2	2.4	5.2	6.2	10.1	E F
Stern66.2	Sternberg (1966), Exp. 2	2.2	2.4	5.2	10.1	D	E F
Stern67	Sternberg (1967c), intact & degraded probes	D					
Stern67int	Sternberg (1967c), intact probes	2.2	2.4	5.2	E	F	
Stern67int1	First of two sessions in Stern67int	3.2	3.4	5.2	6.2	6.3	
Stern69b.4	Sternberg (1969b), Exp. IV, all subjects	10	D	F			
Stern69b.4eq <sup>a</sup>	Sternberg (1969b), Exp. IV, $Pr\{pos\} = Pr\{neg\}$ group	2.2	2.4	5.2	6.2	6.3	F
Stern69a	Sternberg (1969a), Section 10	8	8.2				
Stern75b	Sternberg (1975), Figure 3b	5.2	9	D	E		
Stern75a	Sternberg (1975), Figure 3a	5.2	9	D			

<sup>a</sup>Except in Section 10, where  $Pr\{pos\}$  is of interest, I have used data from the 12 subjects with  $Pr\{pos\} = Pr\{neg\}$ , in order to have approximately equal sample sizes for P-trials and N-trials.

## APPENDIX B

### How to investigate high-speed scanning

#### *The importance of subject motivation*

Three studies indicate how important it can be to provide performance feedback and performance-based incentives for Sv, Sf, and Ashby tasks:

Using a variant of the Ashby task, Franklin and Okada (1983) found in two experiments with a total of 24 subjects, with digit stimuli and  $n_{\text{pos}} = 2, 3, 4, 5$ , that providing RT feedback reduced the mean zero intercept from 694 ms to 461 ms, and the mean slope from 52 to 41 ms, while the mean error rate increased only slightly, from 2.1% to 3.1%.

In a study using the Sf task with digit stimuli and  $n_{\text{pos}} = 1, 3, 5$ , Casement, Broussard, Mullington, and Press (2006) found that the mean slope for the 10 subjects with no feedback was  $63.2 \pm 12.4$  ms/item, while the mean slope for the 12 subjects with RT feedback was  $41.8 \pm 6.3$  ms/item.

In a large study using the Sv task with  $n_{\text{pos}} = 1, 2, \dots, 6$ , four different stimulus ensembles, and about 200 subjects, conducted in Germany and China, Lüer et al. (1998) provided all subjects with RT and accuracy feedback and examined the effect of augmenting this with monetary incentives for half of their subjects. While the effect on the Chinese subjects was small—mean  $\overline{RT}$  functions  $413 + 34.8n_{\text{pos}}$  ms ( $\overline{RT} = 534$  ms) without, and  $431 + 31.5n_{\text{pos}}$  ms ( $\overline{RT} = 545$  ms) with, the incentive—the corresponding functions for the Germans showed a large effect:  $459 + 53.2n_{\text{pos}}$  ms ( $\overline{RT} = 644$  ms) without and  $427 + 40.3n_{\text{pos}}$  ms ( $\overline{RT} = 534$  ms) with the incentive.

#### *Feedback and incentives*

My preference is to use a score that reflects both speed and accuracy, weighting the latter heavily, and to provide a monetary bonus that depends on the score or on its improvement. Without such incentives, some subjects, especially with enthusiastic experimenters who interact with them frequently, may approach optimal performance, but others may not, making the interpretation of their data difficult. In a typical experiment, the score for a block of perhaps 30 trials would be  $\overline{RT}$  in units of .01 s plus 10 points per error. Subjects would be instructed to minimize their scores. In a first session, subjects might be told that if their average score places them among the best half of the subjects, they will earn a (specified) cash bonus. In a later session with the same task, the bonus could depend on

## APPENDIX C

### Analysis of the Atallah–Scanziani Hippocampus Recordings

The analysis started with from 2 to 5 min of data from each of four rats. I then implemented the following steps:

whether or how much their scores improved relative to the previous session. (One goal is to avoid creating large differences in incentive across subjects.) Short blocks, containing 20–40 trials, provide opportunities for feedback and periods of rest.

It is not clear how much to trust the data from studies that do not provide feedback and performance-based incentives, such as studies in which a course requirement forces students to serve as subjects.

#### *Manipulandum*

In Sv and Sf tasks, as I have used them, the manipulandum was a pair of levers, one operated by each hand, positioned so that the hands can relax when not responding, with fingertips and wrists resting on a flat surface, and such that the response is made by flexing all the fingers of the responding hand. With this arrangement, precision of force or location is required by neither the resting position nor the response.

#### *Warning signal*

Subjects benefit from a visual or acoustic warning (e.g., a brief tone or noise burst) about 0.7 s before the test stimulus.

#### *Practice*

Even if subjects have served in other reaction-time experiments, I now believe that to obtain reasonably stable data they should have at least a full session of practice at the task before providing test data. Furthermore, in my experience, the full beneficial effects of practice do not show themselves until the next day. Also, within a session, at least one initial trial block should be regarded as practice and discarded, and, within each block, the first two or three trials should be discarded.

#### *Choice of paradigm*

Unless there are special reasons for using the Sv task, I would suggest using some version of the Sf task: The data appear to be less variable, and it permits separating effects of set and probe. Also, because of the anomaly for  $n_{\text{pos}} = 1$ , discussed in Section 10.1, I would suggest using  $n_{\text{pos}}$  values of 2, 3, 4, and 5, and, where possible,  $n_{\text{neg}} > n_{\text{pos}}$ , which, together, require  $n_{\text{pos}} + n_{\text{neg}} > 10$ .

#### *Design and analysis*

Ideally, experiments should be designed to enable characterization of individuals and their differences; also, analyses should include a demonstration that sufficient practice was given so that the test data are relatively stable.

1. Band-pass using a second-order Butterworth filter, from 25 to 100 Hz.
2. Apply a smoothing spline to the result.<sup>97</sup>
3. Segment the resulting vector into segments of about 0.5 s.

<sup>97</sup>I realize that these first two steps may be far from the ideal analysis. A Hilbert transformation should probably be applied to the filter output, and the assumption that the results do not depend strongly on the filter band should be tested.

4. Find the peaks in each segment, using changes in sign of the first derivative. Define a peak in a segment as one whose magnitude is at least 15% of the magnitude of the highest peak in that segment.
5. Define a “high peak” as one whose magnitude is at least as great as the 60% point in the distribution of all magnitudes over all segments for that rat. To consider only those segments with high gamma power, select those that contain at least six high peaks. This eliminated about 36% of segments.
6. Determine the inter-peak intervals (IPIs) for all peaks in the selected segments. Call these IPIs  $IPI_1$ ,  $IPI_2$ , etc.
7. Select the first six IPIs in each segment.
8. Calculate the first three moments of (a) the  $IPI_1$  across segments, (b) the sum  $IPI_1 + IPI_2$  across segments, (c)  $IPI_1 + IPI_2 + IPI_3$ , etc. These quantities would correspond to the scanning time for  $n_{\text{pos}} = 1, 2, 3$ .

It is now possible to examine the effect of the number of concatenated successive gamma periods on the mean, the variance, etc., of the duration of that concatenation. Means of the first three moments from the four rats are shown in Figure 1. Some of the data for individual rats, including

## APPENDIX D

### Frequency and recency effects in the Sf task and variants

In thinking about the effects of probe frequency (Section 9), it may be important to distinguish relative frequency (RF) within a P-set or N-set from absolute frequency (AF). For example, in the experiments described in Section 9, in which  $n_{\text{neg}}$  was varied, the probes in the N-set were presented with (approximately) equal frequency. The result is that differences in  $n_{\text{neg}}$  were associated with differences in AF, but not in RF. Both AF and RF are likely to influence the memory strength of the probe, but only RF is likely to systematically influence which probe(s) the subject expects and hence perhaps prepares for.

#### *Effects with brief response–stimulus intervals*

In their variants of the Sf task, Theios and associates used R–S intervals that were sufficiently brief so as not to permit feedback or warning signals. In one variant (Theios et al., 1973), the R–S interval was short (0.5 s), with  $n_{\text{neg}} = n_{\text{pos}}$  and P-sets and N-sets that were nested (providing a consistent stimulus–response mapping) and whose members were presented with different frequencies. Substantial effects of stimulus frequency were found on both P-trials and N-trials.  $\overline{RT}_{\text{pos}}$  and  $\overline{RT}_{\text{neg}}$  functions are both S-shaped (a feature accommodated by Theios’s pushdown stack model, but not by SES), with the mean increases in  $\overline{RT}$  from  $n_{\text{pos}} = 2$  to 3, 3 to 4, and 4 to 5 of 30.3, 39.3, and 23.2 ms. In another variant, Theios and Walter (1974) used the

**Table C1.** *Details of the Atallah–Scanziani data*

<i>Rat</i>	1	2	3	4
Slope of mean (ms/period)	29.4	29.5	31.8	29.0
Zero intercept of mean (ms)	1.0	−0.5	−2.1	1.4
Variance of one period (ms <sup>2</sup> )	163	102	150	112
Slope of variance (ms <sup>2</sup> /period)	144	154	251	143
Covariance of successive periods (ms <sup>2</sup> )	−17	−4	20	15
Standard error of covariance (ms <sup>2</sup> )	15	13	11	23
Seconds of data	154	296	172	119
Sampling rate (Hz)	4167	4040	4040	2222
Number of segments	320	597	346	265
Number of selected segments	209	399	201	177

*Note.* The maximum possible covariance is the variance.

slopes of lines fitted by least squares to the mean and variance of the first six cumulated durations, are shown in Table C1.

design of Exp. Stern66.2, but with R–S intervals of 0.5 s or 2.0 s in different trial blocks. The resulting  $RT_{\text{pos}}$  and  $RT_{\text{neg}}$  functions were decelerating rather than linear. Also, there were effects of RF differences of N-set members on  $RT_{\text{neg}}$  values, and pronounced sequential effects on  $RT_{\text{pos}}$  and  $RT_{\text{neg}}$  of both previous probes and previous responses.

Miller and Pachella (1973, 1976) also found pronounced effects of probe frequency, with brief R–S intervals (1.25 and 1.5 s, respectively, in their two experiments),  $n_{\text{pos}} = n_{\text{neg}} = 4$ , and the P-set and N-set fixed for 800 trials, inviting consistent-mapping effects.

#### *Probe frequency effects in the Sf task*

Are there effects of probe frequency or recency in the Sf task, given an R–S interval of 4.5 s or more that contains feedback and a warning signal? One answer is provided by Biederman and Stacy (1974) using a variant of the design of Exp. Stern66.2: the same probe sequence could be used for  $n_{\text{pos}} = 1, 2$ , or 4 by changing the mapping of probes onto responses. By adjusting the relative frequencies among the stimuli that would comprise the P-sets of  $n_{\text{pos}} = 2$  and 4, effects of relative frequency on both positive and negative responses could be measured. These RF effects were negligible on N-trials, consistent with the absence of effects of  $n_{\text{neg}}$  (and AF) described in Section 9, but they were substantial on P trials. How can these effects on P-trials be reconciled with SES? Biederman and Stacy suggest that its locus is the response-selection stage. Because the error rates were high in this experiment, with a mean of 6.1% in the equal-frequency conditions (compared to 1.2% in Exp.

Stern66.2), it should be revisited, adding incentives for good performance.

Except for Exps. Stern66.2 and Stern67, the Sf experiments described above all used approximately equal frequencies among P-probes and different but equal frequencies among N-probes. For Exps. Stern75b and Stern75a, the AFs of N-probes varied inversely with  $n_{neg}$ . Thus, for Exp. Stern75b, with  $n_{neg} = 2, 4,$  and  $8,$  the corresponding probabilities of individual members of the N-set were  $0.250, 0.125,$  and  $0.063;$  for Exp. Stern75a, with  $n_{neg} = 4$  and  $6,$  the corresponding probabilities were  $0.125$  and  $0.083.$  Yet, as shown in Table 6, there was essentially no effect of  $n_{neg}$  on  $\overline{RT}_{neg}$  in either experiment. (If AF and RF had the same effects, then, based on observations in Table 1 of Theios et al. (1973), as  $n_{neg}$  was changed from 2 to 8, with a corresponding reduction in the frequency of each N-probe, we would expect an increase in  $\overline{RT}_{neg}$  of about  $+50$  ms, instead of the obtained non-significant decrease of  $4.9$  ms.)

Other evidence about probe frequency effects in the Sf task can be extracted from the results of Exp. Stern69b.4. Because  $n_{neg} = 10 - n_{pos},$  as  $n_{pos}$  took on values of  $1, 2,$  and  $4,$  and  $Pr\{pos\}$  ranged from  $0.25$  to  $0.75,$  N-probes were presented with AFs that ranged from  $0.027$  to  $0.125$  (a factor of  $4.6),$  while P-probes were presented with AFs that ranged from  $0.083$  to  $0.750$  (a factor of  $9.0).$  Linear models and associated ANOVAs showed that for each response type, the effects of numeric variates  $n_{pos},$  and  $Pr\{pos\}$  are additive. After fitting them and the factor *subjects,* the effect of probe AF (as a numeric variate) on the RT residuals was found to be non-significant for both response types.

For Exps. Stern66.2 and Stern67, the variance of probe frequencies among N-probes increased with  $n_{pos};$  for  $n_{pos} = 1, 2,$

and  $4,$  these variances were  $0.0016, 0.0093,$  and  $0.0113,$  respectively. If N-probe RF had an effect, we would therefore expect that, relative to the variance of the corresponding  $RT_{pos}$  values, the variance of  $RT_{neg}$  would increase for  $n_{pos} > 1.$  This was found in neither of Exps. Stern66.2 or Stern67int. We can conclude, tentatively, that in the Sf task, probe AF has no effect, and probe RF has an effect on P-trials but not on N-trials.

### *Sequential effects in the Sf task*

Given the existence of sequential effects in some RT experiments (such as the “repetition effect”—i.e., facilitation when the stimulus and response from the preceding trial are repeated), it is reasonable to ask why, as in the experiments discussed in Section 9, the Sf procedure does not appear to produce sequential effects, which would, for example, show up strongly on trials with  $\Delta trials = 1$  (which occur with probability  $0.23$  on N-trials in Exp. Stern75b). Most reports of repetition effects are from experiments with short R–S intervals (as short as  $50$  ms); the effects decrease as the R–S interval is lengthened.<sup>98</sup> The exceptions of which I am aware, when repetition and other sequential effects are found at longer R–S intervals, with warning signals and sometimes feedback signals, as in the Sf task, have almost always involved either a complex mapping of four stimuli onto two responses (Smith, 1968, Exp. 1), or four stimulus–response pairs (Kornblum, 1969, discrete experiment; Remington, 1969; Smith, 1968, Exp. 2). Unfortunately for this account, in the study mentioned above by Biederman and Stacy, repetition effects were found for both P-probes and N-probes, with a magnitude that increased with  $n_{pos}$ —another reason to revisit that study.

## APPENDIX E

### Effect of requiring recall in the Sv task

On each trial in Exp. Stern66.1 subjects were required to recall the P-set in order of presentation after making their speeded response. It is conceivable that such a requirement influences the way in which the P-set is represented and hence the way it is searched. Based on data from many subjects in two large experiments modelled on Exp. Stern66.1, but in which the recall requirement was varied between subjects, Corbin and Marquer (2008, 2009, 2013) have argued that it does have such an effect: Averaging over their two experiments (one described in their 2008 and 2009 papers, the other in their 2013 paper), requiring recall caused the mean RT to be  $75$  ms longer and the slope of the  $\overline{RT}$  function to be  $18$  ms greater than when recall was not required.

Both with and without recall, however, their subjects were extremely slow and inaccurate relative to those in Exp. Stern66.1, rendering their findings hard to interpret. In Exp. Stern66.1, the fitted  $\overline{RT}$  function was  $397 + 37.9n_{pos},$  with a mean RT of  $529$  ms. In contrast, averaging over their two experiments, Corbin and Marquer found that when recall was required, the  $\overline{RT}$  function was  $546 + 64.3n_{pos},$  with a mean RT of  $771$  ms, which is substantially steeper and  $242$  ms slower. Mean error percentages in Exp. Stern66.1 in the speeded response and in ordered recall were  $1.3\%$  and  $1.4\%,$  respectively; corresponding values when recall was required in the Corbin–Marquer experiments were  $3.4\%$  and  $12.9\%,$  respectively, substantially greater. The subjects in both experiments were undergraduates, American in Exp. Stern66.1, French in the Corbin–Marquer experiments.<sup>99</sup>

Despite these large differences in speed and accuracy, but because in both experiments, the  $\overline{RT}$  functions averaged over

<sup>98</sup>Such a decrease was found from  $50$  to  $500$  ms (Bertelson, 1961; Smulders et al., 2005), from  $50$  to  $1000$  ms (Bertelson & Renkin, 1966; Soetens, 1998), from  $100$  to  $2000$  ms (Hale, 1967), from  $500$  to  $2000$  ms (Theios & Walter, 1974), from  $250$  to  $750$  ms (Ells & Gotts, 1977), and from  $100$  to  $1000$  ms (Pashler & Baylis, 1991).

<sup>99</sup>After the speeded responses in the experiment by Darley (1973) discussed in Section 11, subjects were required to recall the subset of the P-set that had not been probed. Atkinson et al. (1974, p. 227) report that results were essentially the same in another, similar experiment in which such recall was not required.

subjects were approximately linear, with  $\beta_{\text{neg}} \approx \beta_{\text{pos}}$ , Corbin and Marquer describe them as successful replications of Exp. Stern66.1. (Minimizing the importance of such quantitative differences is common, unfortunately.) Thus, the intriguing conjecture that the recall requirement influences the search process remains to be adequately tested.

There are a number of possible reasons for the poor performance of the French subjects. One is that they appear to

have been given no feedback in either experiment, other than being informed about the correctness of their speeded response after each trial. Nor do they appear to have been told what defines good performance, or to have been given any incentive to perform well: There is no mention of the subjects being paid in the 2008 experiment; they are described as receiving course credit for participation in the 2013 experiment. (See comments in Appendix B about the importance of feedback and incentives.)

## APPENDIX F

### Notes for Tables 1, 2, and 4

#### Table 1

*Exp. Stern67int:* Values are means over P-trials and N-trials, weighted by their relative frequencies (4/15 and 11/15, respectively) and averaged over the two sessions.

*Exp. Stern66.1:* Serial recall of the P-set was required after each speeded response. For  $n_{\text{pos}}$  values of 1, 2, ..., 6, mean recall error percentages were 0, 0, 0, 1.0, 1.6, and 5.7, respectively, with overall error rates the same for P-trials and N-trials.

Entries are numerical values found in papers or data records except for the following, which were read from graphs:

*Exp. Hock84:* Error rates.

*Exp. Monsl78.1:* Slope and intercept based on lines fitted to data from  $n_{\text{pos}} = 2, 3, 4$ , because data from  $n_{\text{pos}} = 1$  fell below the fitted line.

*Exp. Monsl78.1, Monsl78.2:* Error rates.

*Exp. Rdiff78:* Error rates.

#### Table 2

*Exp. Stern66.1:* Because  $\overline{RT}_{\text{pos}}$  for  $n_{\text{pos}} = 1$  tends to fall below the least-squares line, unless  $Pr\{\text{pos}\}$  is low, slopes based on  $n_{\text{pos}} = 2, 3, 4, 5, 6$  are also provided. This problem does not arise for Exp. Stern66.2, probably because  $Pr\{\text{pos}\}$  was only 0.27 (see Section 10.1).

*Exp. Schnei77:* Four subjects were run, two with stimulus ensembles of consonants, two with digits. Slopes for individual subjects are provided, but we are not told which subjects used which stimulus sets.

*Exp. Ashby93:* Slopes were determined from measurements of Figure 1 of Ashby et al. (1993).

*Exp. Frank83:* Two experiments, each with 12 subjects. The table shows slopes of their  $\overline{RT}$  functions. Insufficient data are provided to estimate standard errors. Because the probe delay (500 ms) was short, these experiments differ from others that used the Ashby task. One feature of their results, consistent with the search being self-terminating, is that for P-trials, the slopes of the serial-position curves are substantially greater than the slopes of the  $\overline{RT}$  functions.

#### Table 4

*Exps. Stern66.2 and Stern67int:* For these experiments,  $Pr\{\text{pos}\}/Pr\{\text{neg}\} = 4/11$ . "All" refers to the weighted mean variances, weighted by 4/15 and 11/15.

*Exp. Stern67int:* Variances were determined for each of the two sessions, then averaged over sessions for each of the 12 subjects. The table contains means over the 12 subjects and slopes of linear functions fitted to these means, together with their standard errors based on between-subject differences.

*Exp. Stern69b.4eq:* Individual data for full distributions are no longer available. What are shown are 10% trimmed mean variances, and their slopes.

*Exp. Stern66.1:* The table contains means over the eight subjects, and slopes of linear functions fitted to these means, together with their standard errors based on between-subject differences.

*Exp. Hock84:* Variances were determined from the reported mean parameters of fitted ex-Gaussian distributions, using the fitted linear function for  $\tau$ , and measuring  $\sigma$  values from Figure 4 of Hockley (1984).