Psychological Research (2004) 69: 77–105
DOI 10.1007/s00426-003-0170-5

ORIGINAL ARTICLE

Denis Cousineau · Serge Larochelle

Visual-memory search: An integrative perspective

Received: 10 January 2003 / Accepted: 15 December 2003 / Published online: 19 February 2004

Abstract A large, single-frame, visual-memory search experiment is reported in which memory and display loads of 1, 2, and 4 alphanumeric characters were factorially combined. In addition to the usual Consistent Mapping and Varied Mapping conditions, the experiment also involved a Categorical Varied Mapping condition in which different sets of stimuli switched roles as targets and distractors over trials. The stimuli used in these various mapping conditions were either digits, letters, or digits and letters. Analyses of the response time means obtained early and late in training suggest that the presence of categorical distinctions among the stimuli is the most important determinant of search efficiency. Comparison of the load effects on the response time means and on their standard deviations revealed a fairly constant ratio throughout the experimental conditions, which suggests that similar search processes may have been involved. A feature-based comparison model is indeed shown to account for the response time means obtained after extensive training under just about all training conditions, as well as for the ratios of load effects on means and standard deviations. According to the model, improvement in search efficiency results from a reduction in the number of features considered. The model's performance questions the necessity to postulate qualitative differences between controlled and automatic processing, while the experiment forces a reassessment of the importance of the consistent mapping that underlies dual-process theories.

Introduction

The present study focuses on the classical visual-memory search paradigm, the question addressed concerning the factors that are responsible for the automatization of performance. Although different factors have been proposed, they have never been tested under equivalent conditions. One purpose of the large experiment reported here was to fill this void. The results obtained and the model proposed differ in critical respects from Schneider and Shiffrin's (1977; Shiffrin & Schneider, 1977) early account.

D. Cousineau (✉) · S. Larochelle
Département de Psychologie,
Université de Montréal,
succ. Centre-ville, C. P. 6128,
Montréal, Québec, H3C 3J7, Canada
E-mail: Denis.Cousineau@UMontreal.ca
Fax: +1-514-3432285

Stimulus-response mapping

Readers will recall that Schneider and Shiffrin's seminal study involved two different search tasks. In the single-frame task, participants had to commit to memory a set of one or more items and to search for the presence of any of these potential targets in a single visual display containing one or more elements, some of which serve as distractors. In such a task, the number of items in memory (called the memory set and labeled *M* hereafter) is usually small so that when the display (called the display set and labeled *D* hereafter) remains visible long enough, error rates in the task are low and response times serve as the main measure of performance. In the multiple-frame task, participants have to detect the presence (or location) of one or more targets in a series of multi-element displays presented in rapid succession, the measure of performance being the percentage of correct responses.

In Schneider and Shiffrin's (1977) Experiments 1 and 2, some participants had to search for one or more digits among a set of consonants. So, whenever a digit was present in the display, it was also a member of the memory set and the correct response was "yes." For these participants, digits were consistently mapped onto positive responses over trials, as were consonants for the participants who had to search for consonants among digits. The basic finding was that participants came to detect the target very rapidly and accurately. Moreover, performance was found to be relatively independent of the number of items in the *M* and *D* sets. In other conditions, the stimuli were either all digits or all consonants, depending on the participants. The targets and distractors were chosen randomly over trials, so that a given stimulus was associated with a positive response in some trials and with a negative response in others. Compared with the condition described previously, detection accuracy in the multiple-frame task was smaller and response times in the single-frame task were larger. Performance also remained dependent on both the number of potential targets (i.e., M size), and the number of characters on the display (i.e., D size).

Schneider and Shiffrin (1977) attributed the diverging results across conditions to differences in the mapping of stimuli to responses. They argued that training under varied mapping conditions (labeled VM) did not yield any improvement in performance, and always relied on limited-capacity search. By contrast, training under consistent mapping conditions (labeled CM) eventually led to automatic detection of the target. Subsequent research has led to a better specification of the consistency principle. For instance, Schneider and Fisk (1982) showed that the effects of mapping are not all-or-none. They varied the proportion of times that stimuli served as targets from 100% (perfectly consistent mapping) to 33% and found graded effects on the accuracy of target localization in a multiple-frame visual search task. Fisk and Schneider (1984) delineated the consistent mapping effect more precisely by showing that variations in the physical responses associated with target-present and target-absent decisions did not prevent automatization in the multiple-frame task. Logan (1978) reported similar findings, using response time in a single-frame task as the measure of performance.[1]

Categorical distinctions

There are many factors other than mapping that could have contributed to the differences observed in Schneider and Shiffrin's (1977) experiments, one of which is the presence vs. absence of categorical distinctions among the stimuli. Indeed, the targets and distractors used for any given participant in the CM conditions belonged to two distinct and well-known categories (digits and consonants), whereas the stimuli used in the VM condition belonged to only one category (digits or consonants). So, the advantage of the CM condition can be attributed to this categorical distinction (which we will refer to as categorical heterogeneity among the stimuli) as much as it can be attributed to consistent mapping. Schneider and Shiffrin were aware of this confound and countered the argument by presenting the results of an earlier study by Briggs and Johnsen (1973). In this study, the stimuli were homogeneous, i.e., they were all letters. In one condition, the VM condition, all stimuli served equally often as targets and as distractors whereas in two other conditions, the CM conditions, the targets and distractors were chosen from two separate sets of letters. The results obtained by Briggs and Johnsen in the VM condition showed the same interaction pattern found in Schneider and Shiffrin's experiment. The combined effects of memory and display size were much reduced but not eliminated from the CM conditions. In order to account for this small load effect, Schneider and Shiffrin argued that Briggs and Johnsen's participants did not have enough practice to fully automatize search.

Shiffrin and Schneider (1977) also performed an experiment using homogeneous stimuli, i.e., only consonants, in a CM condition. Like Schneider and Shiffrin's (1977) Experiment 1, Shiffrin and Schneider's experiment involved a multiple-frame task. However, memory size was fixed at 4 (labeled M4) and display size at 2 (labeled D2) for the entire experiment. Performance in terms of hit rates (82%) failed to reach the level obtained with heterogeneous stimuli (92%) in the M4D2 condition of Schneider and Shiffrin's Experiment 1. These results suggest that the presence vs. absence of a pre-established categorical distinction between targets and distractors does have an effect on performance. Although the amount of training was quite comparable in the M4D2 conditions of the two experiments, the total amount of practice was greater in Schneider and Shiffrin's Experiment 1 since it involved other memory and display sizes. So, the differences observed across experiments could again be due to a differential amount of practice.

In the present study, we compare the results obtained with homogeneous vs. heterogeneous stimuli under equal amounts of practice, in order to better specify the effects of categorical distinctions on the performance obtained in CM as well as in other mapping conditions. Our analyses are not only of the results obtained after extensive training, we also consider performance obtained early in training, the rationale being that the effects of pre-established categorical distinctions should be present at the very beginning of the experiment whereas the effects of mapping should develop with training.

Set composition

Another factor that could possibly account for the differences between CM and VM conditions obtained in Schneider and Shiffrin's (1977) experiments, and in many other studies since then, is the following. Even when stimuli are homogeneous, they nonetheless form two disjointed sets in CM conditions: A set of targets and a set of distractors. This is not the case in VM conditions. Shiffrin and Schneider (1977) were aware of this possible confound and they designed a critical experimental condition to investigate it. In this condition, called categorical varied mapping (labeled CVM), the stimuli are divided into two sets that switch roles over trials. If in a given trial the potential targets are

[1] The expression *stimulus-response mapping*, often heard in the context of automatic attention attraction theory, is therefore somewhat misleading since it is *stimulus-decision mapping* that appears to be critical.

chosen from one set, the distractors are picked from the other, and vice versa. So, the stimuli are consistently mapped onto different sets, as in CM conditions, but the sets are not consistently mapped onto the target-present vs. target-absent decisions. Finally, the decisions are consistently mapped on the physical responses, as in CM conditions. CVM conditions are therefore very different from the conditions involved in the experiments of Fisk and Schneider (1984) and Logan (1978), described previously, where the mapping of the stimulus sets to the target-present vs. target-absent decisions was consistent, but the mapping of the decisions to the physical responses was not.

Shiffrin and Schneider (1977, Experiment 3) tested a CVM condition with homogeneous stimuli (i.e., consonants) in a multiple-frame task. Performance was found to be much better in the CVM condition than in the VM condition. Moreover, with practice, the effect of memory set size disappeared suggesting that, in the CVM condition, participants proceeded by determining to which set the display items belonged instead of comparing the display items with those in memory. Unfortunately, display size was not varied in the experiment. Hence, it is not known whether visual search of the display can proceed automatically in CVM conditions. Nonetheless, Shiffrin and Schneider (1977; see also Schneider & Shiffrin, 1985) argued that only consistent mapping could produce automaticity. Their conclusion was based on the rapid improvement in performance observed when participants were transferred to a CM condition after extensive initial training in VM and CVM conditions. Note that such improvement in performance could as well be due to the fact that the number of stimuli serving as targets was reduced by half when participants were transferred to the CM condition. Cheng (1985) also criticized Shiffrin and Schneider's conclusion on the grounds that an absolute level of performance cannot serve as a criterion of automaticity when it is not theoretically motivated. Other arguments against Shiffrin and Schneider's conclusion can be found in Logan and Stadler (1991, Footnote 1).

The only single-frame experiment approximating a CVM condition is that of Dumais (1979, Experiment 3; also described in Shiffrin & Dumais, 1981, and in Shiffrin, Dumais & Schneider, 1980). The experiment involved many sets of stimuli (labeled A, B, and C), each comprising a mixture of characters: Roman, Greek, and Hebrew letters, as well as Arabic digits. In one condition, the stimuli in set A served as targets and the stimuli in set B served as distractors. In another condition, the B stimuli were targets and the C stimuli distractors. In a third condition, the C stimuli served as targets to be found among the A stimuli. In the first part of the experiment, the different conditions were run in separate blocks of trials so that stimulus-response mapping was consistent within each block but varied over blocks. In a later part, the three conditions occurred within blocks, but the set distinctions were preserved. Similar but mixed results were obtained in both parts of the experiment. In the best condition, performance was comparable with that obtained in a control CM condition. At worst, performance was comparable with that obtained in a control VM condition. Moreover, the best and worst conditions were not the same across participants.

Available evidence concerning CVM performance is therefore quite ambiguous. It is also incomplete: We do not know what level of performance would be reached in CVM if the two sets of stimuli were taken from different categories such as digits and letters. The experiment reported here involved such a condition (labeled CVM-HETERO).

Target-distractor discriminability

So far, we have argued that the consistent mapping effects originally reported by Schneider and Shiffrin (1977) could have been at least partly confounded with two other factors: The presence vs. absence of pre-established categorical distinctions and the set composition of the stimuli. The effects of these variables could in turn be mediated by a third factor, namely the discriminability of targets and distractors. Treisman and Gelade (1980) argued that the CM effects obtained by Schneider and Shiffrin (1977) with categorically distinct stimuli could be due to perceptual differences between digits and letters, digits tending "to be narrower, asymmetrical, open to the left, and to have shorter contour than letters" (p. 120). Evidence that the categorical effect in visual search could possibly be due to perceptual differences between letters and digits can be found in Duncan (1983), Krueger (1984), and Cardosi (1986).

The set composition of the stimuli can also have consequences on the relative discriminability of targets and distractors. Treisman and Gelade (1980) speculated that, even when there is no categorical distinction among the stimuli, participants might learn which features best discriminate each target from the distractors. Given an equal number of stimuli in CM and VM conditions, search will be easier in CM than in VM conditions since participants in CM conditions have to discriminate only a subset of stimuli serving as targets from another subset serving as distractors whereas in VM conditions participants must learn to discriminate each stimulus from all the others.

Fisher (1986) proposed a model of visual search that provides a discriminability-based account of mapping effects. The model (also described by Czerwinski, Lightfoot, & Shiffrin, 1992) assumes that the feature tests required to locate a target are gradually reordered so that, after training, the feature that discriminates the target from the largest number of potential distractors is tested first. The display locations that do not contain this feature are excluded from further consideration. Then the second most discriminative feature is tested and so on. However, as the number of feature tests required to locate the target increases, the number of channels available to analyze the stimuli decreases so that fewer and fewer display items can be tested in parallel. So, the size of the display load effects observed depends primarily on the amount of featural overlap between targets and distractors. With a given stimulus ensemble, amount of overlap will be maximal when all stimuli serve equally often as targets and distractors, as is often the case in VM conditions. By contrast, an optimal sequence of feature tests can be found in CM conditions since the same subset of stimuli always serves as targets, thereby reducing the effects of display load on performance. Fisher's featural overlap model was shown to account for the differences in hit rates obtained in the CM and VM conditions of a multiple-frame, visual search task, using upper-case letters as stimuli.

In the present experiment, we did not manipulate target-distractor discriminability. Instead, we attempted to control it in a manner similar to Krueger (1984). However, the proposed model suggests how target-distractor discriminability (as well as the similarity among distractors, see Duncan & Humphreys, 1989; Humphreys, Quinlan, & Riddoch, 1989), interacts with set composition and categorical distinctions in visual-memory search.

Experiment

The Experiment involved the three mapping conditions discussed in the Introduction: CM, CVM, and VM. As noted by Cheng (1985), an orthogonal comparison of the influence of sets vs. mapping would require a fourth condition involving consistent mapping but no set distinction between the targets and distractors. Unfortunately, such a mapping condition is impossible. The stimuli used in each mapping condition were homogeneous (labeled HOMO), i.e., composed of digits or consonants, or heterogeneous (labeled HETERO), i.e., composed of both digits and consonants.

Method

Participants

There were four different participants in each of the six conditions resulting from the crossing of mapping (CM, CVM, and VM) and stimulus factors (HETERO and HOMO). All participants were undergraduate students. Each received $CAN 100 for completing ten experimental sessions of approximately 1.5 h each. Duplication of one experimental condition involved four additional participants, each of whom could earn an additional $CAN 2.50 per session if mean response time was smaller than that of a matched participant (and if errors did not exceed 5%). In order to prevent discouragement (which would annihilate the effects of the extra monetary incentive), participants were only informed of the result of this competition at the end of the experiment.

Stimuli

The digits used were chosen in such a way that the stimuli serving as targets and as distractors in the CM condition could not be reliably distinguished on the basis of a simple perceptual feature such as angular vs. curved shape. Other constraints were that the set of potential targets and the set of potential distractors in the CM condition contained the same number of even and odd digits, and of large and small digits (i.e., above and below 4.5). The two resulting sets of digits were {2, 3, 6, 7} and {1, 4, 5, 8}. The corresponding sets of consonants were {Z, B, G, F} and {L, H, R, S}. Selection of the consonants in each set was done by choosing a consonant that resembled each digit in the corresponding set. This procedure was followed to ensure that there would be no gross difference in the degree of perceptual distinctiveness between the consonants and the digits used. It is important to note, however, that this type of control does not guarantee an absolutely equivalent degree of discriminability within and across categories.

Design

Table 1 shows the assignment of the stimuli to the various participants in the CM conditions. Among the 4 participants in the CM-HOMO condition, 2 saw only digits and 2 saw only consonants. Each of the four sets of stimuli previously described served as targets for one of these participants, the other set of the same category serving as distractors. Each of these four sets of stimuli also served once as targets and once as distractors over participants in the CM-HETERO condition. Assignment of the stimuli to the participants in the HETERO condition was as follows: If the targets were taken from one set of digits, the distractors consisted of the consonants resembling the other set of digits (i.e., resembling the set of digits used as distractors for the corresponding participant in the HOMO condition). This procedure was followed to minimize differences in target-distractor discriminability across the HOMO and HETERO conditions.

Assignment of the stimuli to the various participants in the CVM and VM conditions was the same as described for the CM conditions. However, in the CVM conditions, the two sets of stimuli assigned to a given participant switched roles randomly over trials, serving sometimes as targets and sometimes as distractors. In the VM condition, the stimuli serving as targets and distractors in each trial were chosen from all eight stimuli assigned to the participant, without any regard to their set of origin.

Some of the results obtained in the CVM-HOMO condition led us to test four additional participants under a slightly modified condition. In contrast to participants in the original CVM-HOMO condition, participants in the new condition were informed of the existence of two different sets of stimuli at the onset of the experiment and they were reminded at the beginning of every session thereafter. Instead of being told about the specific composition of each set, they were asked to recall the members of the two sets after each session. Most participants were able to recall correctly the four characters of each set by the end of the first session, and all of them by the end of the second session. Since the goal of this duplication was to determine the best performance achievable under CVM-HOMO conditions, these participants were given the extra monetary incentive described previously. Unfortunately, this

Table 1 Assignment of stimuli to participants in the CM conditions of the experiment

Participant	Stimulus condition	
	Targets	Distractors
	Homogeneous	
S1	2, 3, 6, 7	1, 4, 5, 8
S2	1, 4, 5, 8	2, 3, 6, 7
S3	L, H, R, S	Z, B, G, F
S4	Z, B, G, F	L, H, R, S
	Heterogeneous	
S5	2, 3, 6, 7	L, H, R, S
S6	L, H, R, S	2, 3, 6, 7
S7	1, 4, 5, 8	Z, B, G, F
S8	Z, B, G, F	1, 4, 5, 8

emphasis on performance prevents unambiguous interpretation of potential differences in results between the two CVM-HOMO conditions since these conditions differed in two ways: Amount of information and amount of money.

Display (D) and memory (M) set size varied within subjects, as in Briggs and Johnsen's (1973) and Schneider and Shiffrin's (1977) studies. However, in Briggs and Johnsen's experiment, M size was blocked over consecutive trials and in Schneider and Shiffrin's experiment, both M size and D size were blocked. By contrast, both M size and D size varied randomly over trials in the present experiment. This reduces the likelihood that participants develop and rely on specific strategies for specific set sizes. The possible values that M size and D size could take were 1, 2, and 4. The specific stimuli forming the memory and the display sets also varied over trials except in CM conditions where a memory set size of four exhausted all available targets and where a display size of four also required all distractors in negative trials.

Procedure

Participants in the CVM and VM conditions spent 8 of the 10 experimental sessions practicing the search task. The last two sessions were devoted to a transfer task. For the participants in the CM conditions, only the first six sessions were devoted to learning, and the last four served as transfer tasks. The transfer tasks performed are of secondary interest with respect to the issues addressed here. For purpose of brevity and focus, these tasks will not be described nor discussed here, our presentation being limited to the training results.

Each training session comprised 864 trials. Although participants could take a break after every 108 trials, each session was composed of two equivalent blocks of 432 trials, half of which were positive and half negative. All combinations of M size and D size were equally represented within each block. Moreover, every possible target was used equally often in positive trials in each combination of M and D size. Since there were twice as many potential targets in the CVM and VM conditions, participants in these conditions had half as much practice per target as those in the CM conditions (which explains why there were more training sessions in CVM and VM conditions).

In each learning trial, an asterisk was presented in the center of a VGA screen for 500 ms before and after the presentation of the memory set. Members of the M set were presented for 1 s in random order, on a line centered on the position of the asterisk. Following Schneider and Shiffrin's (1977) procedure, the display set was presented in a square formation around the center of the screen. Members of the D set occupied a randomly chosen corner of the square and a dot was centered in each position left unoccupied when display size was smaller than four. The visual angle of the D set was about 2° horizontally and vertically, as in Schneider and Shiffrin's study. However, in contrast with Schneider and Shiffrin, who masked the display set after 160 ms in their single-frame experiment, our display set remained visible until the participant responded or until 3 s had elapsed. Stimulus presentation was synchronized with the display refresh rate.

Participants responded by pressing the "1" or "2" key on the numeric keypad of the computer keyboard to indicate a "yes" or "no" response respectively. They were instructed to respond as fast as possible while maintaining an error rate lower than 5%. Participants got feedback on their response and response time after every trial, and recapitulative feedback before every rest break (i.e., after every 108 trials).

Results and discussion

Response time mean

Figures 1, 2, and 3 show the response times (RTs) obtained in the first block of the first session of the experiment along with those obtained in the second block of the sixth session for the six original Mapping × Stimulus conditions. Each figure contains four panels. The top two panels present the average RTs obtained in positive (left panel) and negative trials (right panel) as a function of M size and D size in the first training block. Since there was no practice prior to the experiment and no warm-up at the beginning of any of the blocks, the results summarize performance obtained in the first 432 trials. The bottom two panels present, likewise, the average RTs obtained in the twelfth training block. The scale used in the bottom panels spans 1,000 ms, as in the top panels. However, the origin has been lowered by 200 ms to accommodate the smaller RTs obtained in Block 12. Each participant had performed over 4,500 training trials prior to Block 12. The data of one participant are excluded from the results in Fig. 3 because this participant's performance was very erratic due to drowsiness caused by medication. Therefore, the results of this participant were excluded from all analyses reported. Unfortunately, this participant had been assigned to a fairly critical condition, namely CVM-HETERO.

Considering the size of the experiment, the RTs obtained in each mapping condition were analyzed separately. Each of the three ANOVAs performed involved one between-participants factor: Stimulus (S: HOMO, HETERO), and four within-participants factors: Block (B: 1, 12), Memory size (M: 1, 2, 4), Display size (D: 1, 2, 4), and correct Response (R: Positive, Negative). The data entered in these analyses were the participants' mean RT per condition rather than their median RT (see Miller, 1988). Means were preferred because of their intrinsic relationship with standard deviations, which will be considered later. All correct RTs were included in the computation of the participants' means, RTs longer than 3 s having already been excluded at the data collection stage.

Although they are not presented at length here, the error rates were affected by the various factors in the experiment. For instance, the average error rates were smaller in Block 12 (3.7%) than in Block 1 (4.7%). They were smaller in CM (3.3%) than in CVM (4.6%) and in VM (4.7%) conditions. The error rates obtained with heterogeneous stimuli (3.7%) were also smaller, on average, than those obtained with homogeneous stimuli (4.7%). Correlations computed between the mean error rates and the mean response times of the six groups of participants did not reveal any significant speed accuracy trade-off over $M \times S$ conditions in either Block 1 or Block 12. There was a tendency for error rates to increase with both M size and D size. The error rates were also higher for positive (4.9%) than for negative trials (3.5%), as is usually found in search experiments (Zenger & Fahle, 1997). In order to test for the presence of trade-offs across the various $M \times D \times R$ conditions, separate correlations between mean error rates and response times were computed for each of the six groups of participants in Block 1 and in Block 12. None of the correlations computed was significantly smaller than zero.

Varied mapping

Let us start our review of the RT results with the least controversial conditions, namely the VM conditions. As shown in Fig. 1, VM performance exhibited the type of fanning, $M \times D$ interaction, characteristic of limited-capacity search, $F(4, 24) = 18.2$, $p < .0001$, $MS_e = 6{,}057$. Comparison of the right half with the left half of Fig. 1 shows the $M \times D$ interaction to be more pronounced in negative trials than in positive trials, which resulted in a significant $M \times D \times R$ interaction, $F(4, 24) = 6.66$, $p < .001$, $MS_e = 3{,}850$. Comparison of the top and bottom halves of Fig. 1 shows that VM performance was characterized by the same type of $M \times D \times R$ interaction in Block 12 as it was in Block 1. However, the magnitude of the interaction was different, resulting in a significant interaction also involving Block, $F(4, 24) = 3.62$, $p < .02$, $MS_e = 1{,}097$. To get an estimate of load effects on search time, we computed the difference between conditions M4D4 and M1D1. For positive trials, this difference was 17% smaller in Block 12 than in Block 1. For negative trials, the difference was 6% smaller in Block 12 than in Block 1. The difference between M4D4 and M1D1 allows a search rate of about 22 ms/item for positive trials in Block 12 to be calculated, which is about half the size of that for negative trials (40 ms/item).

Schneider and Shiffrin (1977) attributed such a pattern of results to a serial, self-terminating comparison process. They argued further that VM performance should be relatively insensitive to training since it is assumed to always rely on the same search process. This appears to have been the case in our experiment, at least in negative trials. Finally, Schneider and Shiffrin's original theory made no provision for categorical distinctions among the stimuli. Accordingly, Fig. 1 shows VM performance to be quite insensitive to the type of stimuli used, at least in Block 12 where the results of the HOMO and the HETERO groups are almost indistin-

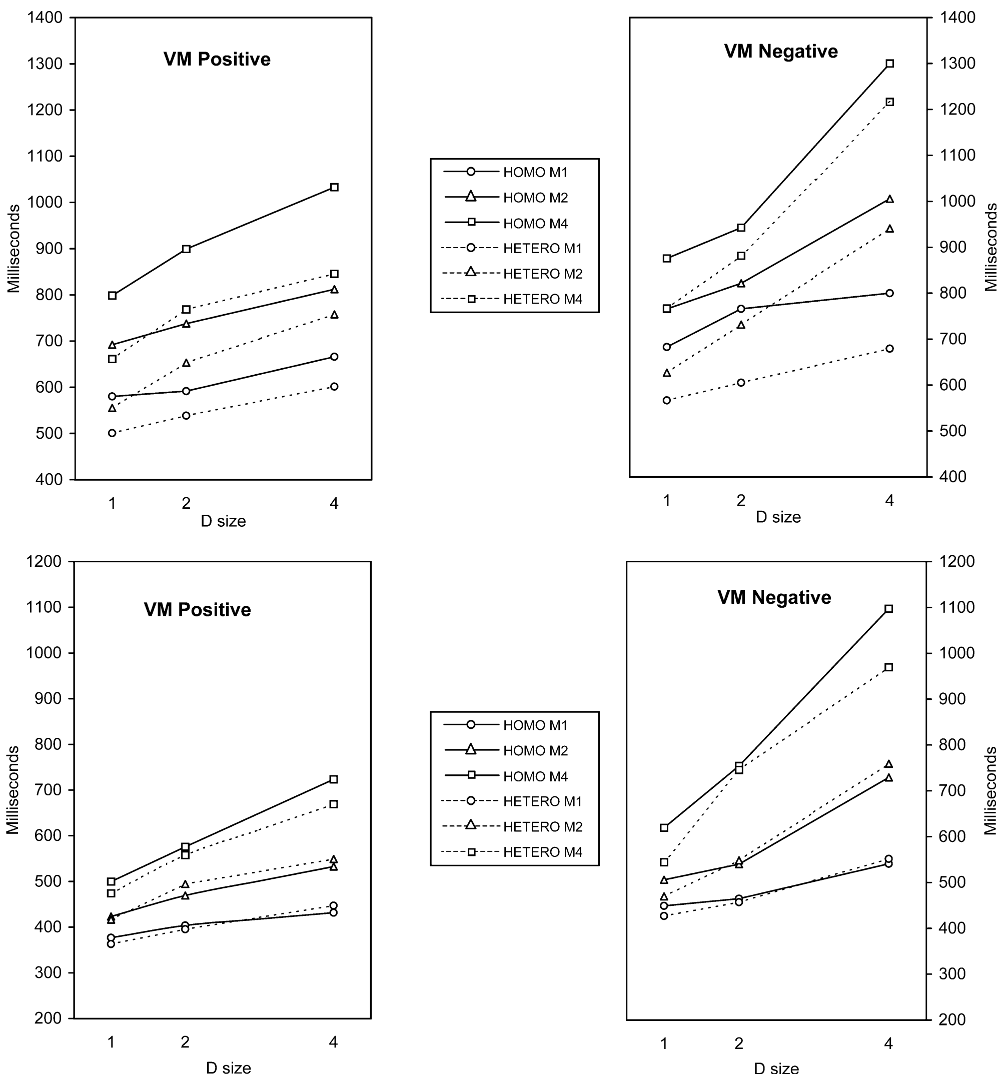


Fig. 1 Mean response times obtained in the first block (*top row*) and the twelfth block (*bottom row*) of training under the varied mapping (*VM*) conditions of the experiment. Results are presented separately for positive (*left panels*) and negative (*right panels*) trials, as a function of stimulus composition homogeneous vs. heterogeneous (HOMO vs. HETERO), memory (*M*), and display (*D*) size

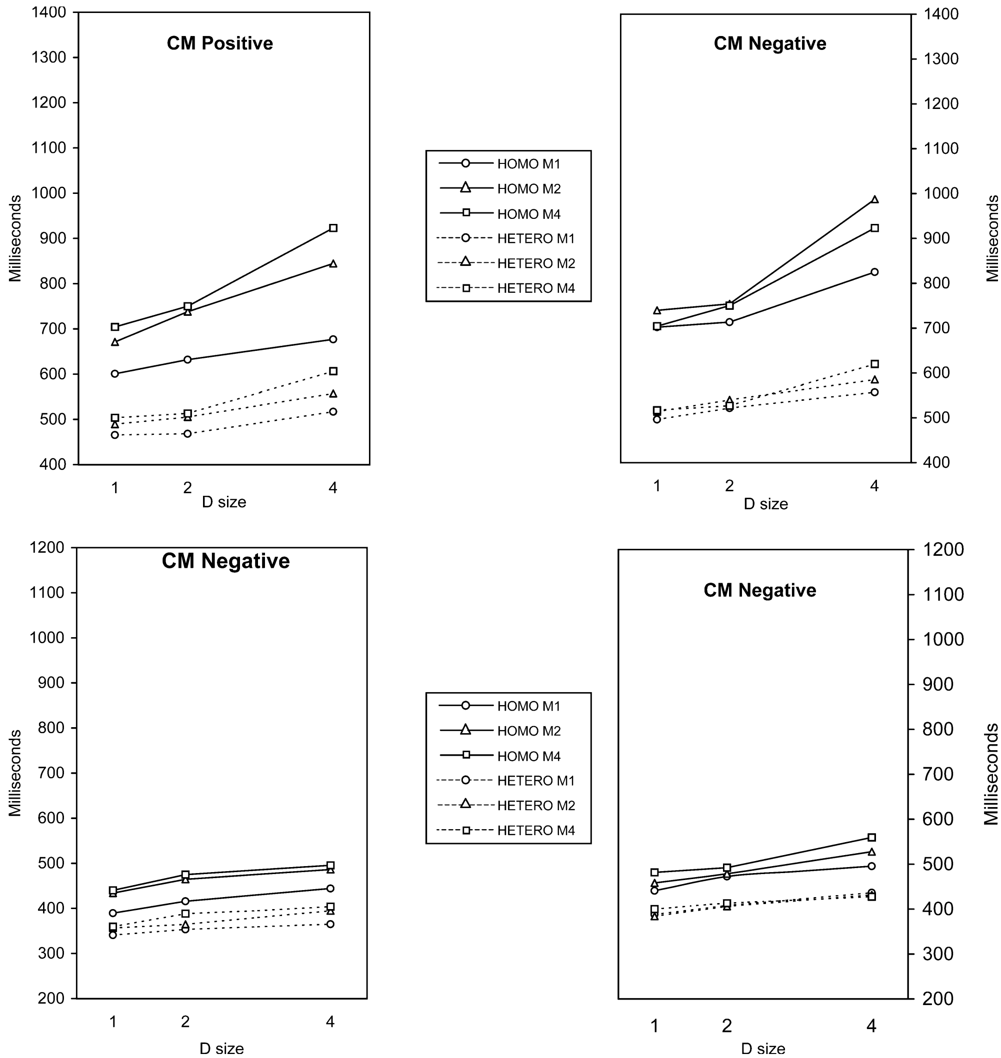


Fig. 2 Mean response times obtained in the first block (*top row*) and the twelfth block (*bottom row*) of training under the consistent mapping (*CM*) conditions of the experiment. Results are presented separately for positive (*left panels*) and negative (*right panels*) trials, as a function of stimulus composition (HOMO vs. HETERO), memory (*M*), and display (*D*) size

guishable. However, in Block 1, performance of the HETERO group exhibited a $M \times R$ interaction, $F(2,12) = 3.50$, $p < .07$, $MS_e = 7{,}143$, which was more pronounced than that found in the performance of the HOMO group, $F(2,12) = 0.09$, $p > .90$, $MS_e = 7{,}143$, resulting in a significant $S \times B \times M \times R$ interaction, $F(2,12) = 5.51$, $p < .02$, $MS_e = 1{,}872$. No other effect or interaction involving the composition of the stimuli reached statistical significance (all $p > .30$).

Consistent mapping

By contrast with the VM results just described, the CM results presented in Fig. 2 show a quite different pattern across groups. This was confirmed by the ANOVA, which revealed many reliable interactions involving Stimulus conditions. Decomposition of these interactions along the Stimulus dimension gave the following results for the HOMO condition. The fanning $M \times D \times R$ interaction was significant for this group, $F(4, 24) = 3.10$, $p < .04$, $MS_e = 755$, but the interaction also involving Block was not, $F(4, 24) = 1.50$, $p > .20$, $MS_e = 890$. The largest interaction with Block only involved M size and D size, $F(4, 24) = 9.96$, $p < .001$, $MS_e = 1{,}131$. Decomposing this interaction along the Block dimension revealed the $M \times D$ interaction to be significant in Block 1, $F(4, 24) = 13.0$, $p < .0001$, $MS_e = 2{,}005$, but not in Block 12, $F(4, 24) = .52$, $p > .70$, $MS_e = 329$. The M size effect remained significant in Block 12, $F(2, 12) = 48.7$, $p < .0001$, $MS_e = 290$, but its magnitude was larger for positive than negative trials, the difference between M4 and M1 being 54 ms in the former case and 41 ms in the latter, which resulted in a significant $M \times R$ interaction, $F(2, 12) = 6.85$, $p < .02$, $MS_e = 155$. The D size effect was also significant in Block 12, $F(2, 12) = 51.1$, $p < .0001$, $MS_e = 437$, the difference between D4 and D1 being smaller in positive trials (54 ms) than in negative trials (67 ms), but not significantly so, $F(2, 12) = 2.33$, $p > .10$, $MS_e = 336$. Overall, search time, as indexed by the difference between M4D4 and M1D1 decreased by over 70% from Block 1 to Block 12.

The pattern of results just described is exactly what theories that postulate qualitative differences between

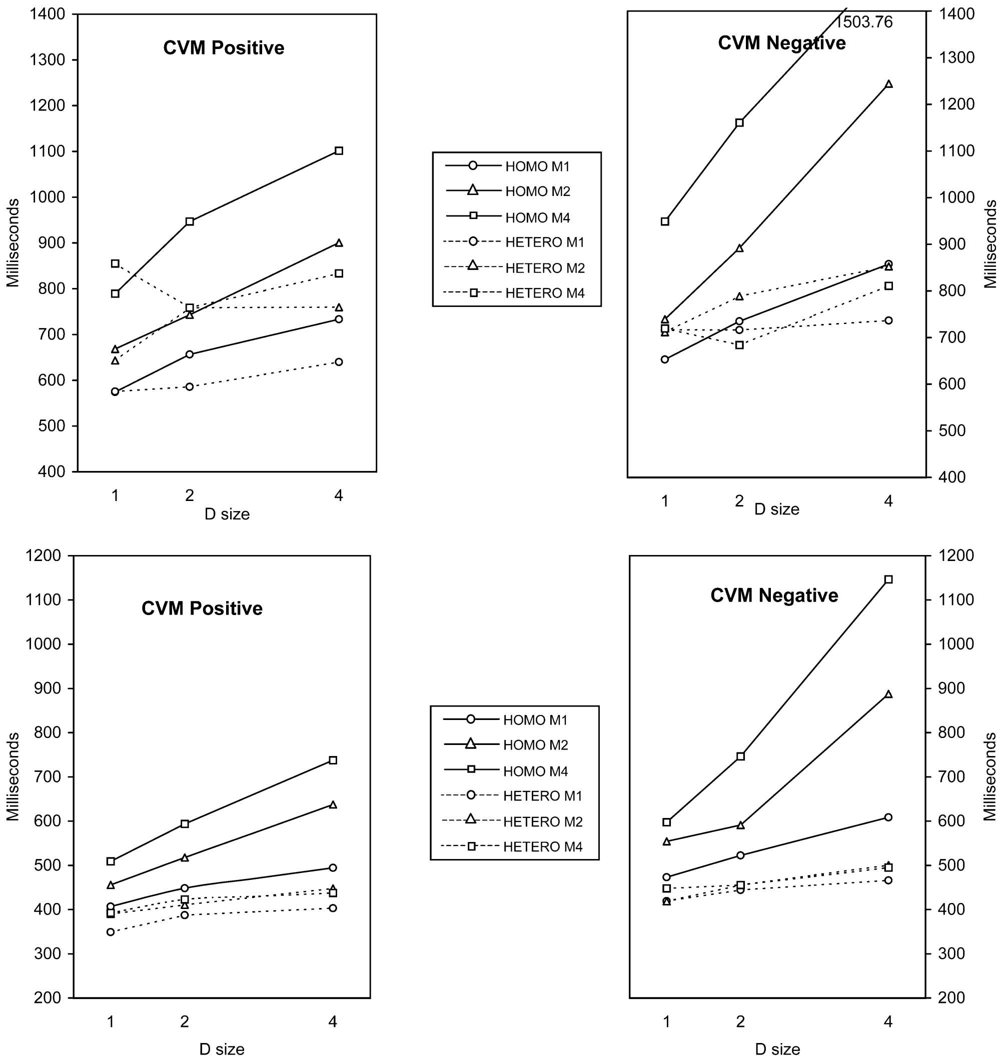


Fig. 3 Mean response times obtained in the first block (*top row*) and the twelfth block (*bottom row*) of training under the categorical varied mapping (*CVM*) conditions of the experiment. Results are presented separately for positive (*left panels*) and negative (*right panels*) trials, as a function of stimulus composition (HOMO vs. HETERO), memory (*M*), and display (*D*) size

controlled and automatic processing would predict. Indeed, the large effects of memory and display load obtained early in training are best characterized as multiplicative, whereas the much smaller effects obtained after extensive practice appear additive. However, both the memory and display load effects remained significantly different from 0 after more than 4,500 training trials. The memory search rates estimated from the differences between M4 and M1 are 18 ms/item for positive trials and 14 ms/item for negative trials. Analogous calculations give display search rates of 18 ms/item in positive trials and 22 ms/item in negative trials.

If automaticity can only be achieved through extended practice under consistent mapping conditions, performance in the HETERO condition should also exhibit large multiplicative load effects early in training and smaller additive effects later on. In reality, we found little evidence of an $M \times D \times R$ interaction, $F(4, 24) = .29$, $p > .80$, $MS_e = 755$, or of an $M \times D$ interaction in the performance of the HETERO group, $F(4, 24) = .62$, $p > .60$, $MS_e = 1{,}202$, even at the beginning of training. The only interaction with Block to approach statistical significance involved *D* size, $F(2, 12) = 2.84$, $p < .10$, $MS_e = 2{,}176$, the difference between D4 and D1 being about twice as large in Block 1 as it was in Block 12. The *D* size effect remained significant in Block 12, $F(2, 12) = 20.3$, $p < .001$, $MS_e = 437$, and of similar magnitude for positive and negative trials, $F(2, 12) = .15$, $p > .80$, $MS_e = 336$, the difference between D4 and D1 being 35 ms in the former case and 41 ms in the latter (for estimated search rates of 12 ms/item and 14 ms/item respectively). The failure of *M* size to interact with Block, $F(2, 12) = .74$, $p > .40$, $MS_e = 3{,}131$, suggests that memory search was affected little by training. However, it was affected by the nature of the response. In Block 12, the difference between M4 and M1 was 31 ms in positive trials (search rate of 10 ms/item) but only 3 ms in negative trials (search rate of 1 ms/item), resulting in a significant $M \times R$ interaction, $F(2, 12) = 8.58$, $p < .005$, $MS_e = 155$.

The qualitative change in performance that would have been expected on the basis of a dual-process view clearly failed to materialize: The combined effects of memory and display load were already additive in the first 432 training trials.[2] So, whatever strategy participants in the CM-HETERO condition relied on appears to have already been established at the start of the Experiment. The efficiency of display search did improve with training but load effects remained after extensive training. So, if the criterion of automaticity is taken as a total absence of load effect, it has to be concluded that visual search was not fully automatic, even in the CM-HETERO condition.[3] However, the display rates obtained approximate the less stringent 10 ms/item figure often taken as indicative of automatic processing (Wolfe, 1998). The absence of a memory load effect in negative trials suggests that the participants came to ignore the memory set when there was no target on the display. Participants could also have bypassed memory search in positive trials because consistent mapping insures that any member of the target category found on the display was also necessarily part of the memory set. The significant M size effect obtained in positive trials shows that participants did not ignore the memory set altogether, even after extensive practice (see Kramer, Strayer, & Buckley, 1990).

Categorical varied mapping

As Fig. 3 shows, the two original CVM groups differed even more from each other than the two CM groups. Decomposition of the analyses along the Stimulus dimension revealed no significant interaction between M size and D size, $F(4, 20) = 1.42$ $p > .20$, $MS_e = 4{,}575$, or between these two factors and Response, $F(4, 20) = .74$ $p > .50$, $MS_e = 1{,}985$, in the CVM-HETERO condition, either early or late in training, $B \times M \times D \times R$: $F(4, 20) = .47$ $p > .70$, $MS_e = 3{,}469$. For this condition, the only interactions with Block to approach statistical significance involved M size, either singly, $F(2, 10) = 3.13$ $p < .09$, $MS_e = 5{,}659$, or in combination with Response, $F(2, 10) = 3.97$ $p < .06$, $MS_e = 5{,}436$. By Block 12, memory search rates, as estimated from the difference between M4 and M1, were 12 ms/item in positive trials and 8 ms/item in negative trials. Display search rates estimated from the difference between D4 and D1 were 17 ms/item in positive trials and 19 ms/item in negative trials.

These results, in conjunction with those obtained in the CM-HETERO condition, strongly suggest that the existence of pre-established categorical distinctions among the stimuli allows for additivity of the memory and display load effects, provided that these distinctions are mapped onto different sets of stimuli, as was the case here. Swapping the digits and consonants as targets in the CVM-HETERO condition probably generated some confusion about the composition of the memory set. As a result, participants may have been tempted to verify the content of the memory set, even in negative trials. This would explain the difference in memory search rates in negative trials between the CVM-HETERO and CM-HETERO groups. For positive trials, memory search rate was comparable for the two groups. Visual search rates obtained in the CVM-HETERO condition are larger than those reported for the CM-HETERO condition, but they have about the same magnitude as those obtained in the CM-HOMO condition. The overall average response time obtained in CVM-HETERO (430 ms) in Block 12 was also between that obtained in CM-HETERO (390 ms) and CM-HOMO (470 ms).

These results are problematic for theories proposing that consistent stimulus-response mapping is necessary and sufficient for the automatization of visual-memory search. If we consider that the participants in the CM-HOMO condition were on the path to automaticity, we must also consider that the participants in the CVM-HETERO condition had achieved an even higher degree of automaticity. Since neither the individual stimuli nor the stimulus sets were consistently mapped onto decisions or responses in the latter condition, we have to conclude that consistent mapping is not necessary for automatization. Alternatively, if we wish to argue that only participants in the CM-HETERO condition achieved automaticity, we can maintain that consistency of mapping is necessary for automatization. However, we have to concede that it is not sufficient since CM-HOMO performance did not reach the level obtained in the CM-HETERO condition.

If the performance of the CVM-HETERO group resembled that obtained in the CM conditions, CVM-HOMO performance was more akin to that obtained in the VM conditions. The sub-analysis performed on the results of CVM-HOMO confirmed the reliability of the $M \times D$, $F\,(4, 20) = 17.0$, $p < .0001$, $MS_e = 4{,}575$, and the $M \times D \times R$ interactions, $F(4, 20) = 8.18$, $p < .0001$, $MS_e = 1{,}985$), which are manifest in Fig. 3. Neither of these interactions was significantly different across Blocks: $B \times M \times D$, $F\,(4, 20) = .70$, $p > .50$, $MS_e = 2{,}653$; $B \times M \times D \times R$, $F(4, 20) = .79$, $p > .50$, $MS_e = 3{,}469$. However, Block interacted significantly with M size, $F(2,10) = 11.4$, $p < .01$, $MS_e = 5{,}659$, and to a less reliable extent, with D size, $F(2, 10) = 3.79$, $p > .06$, $MS_e = 4{,}878$.

Shiffrin and Schneider (1977) reported that the CVM-HOMO performance obtained in early versions of their Experiment 3 failed to improve. If we exclude that there are fundamental differences between the multiple-frame task used by Shiffrin and Schneider and the single-frame task used here, then we can hypothesize that the improvements observed later by Shiffrin and Schneider were due to the participants' awareness of the set composition of the stimuli. By contrast, our participants clearly failed to realize the existence of such a distinction among stimuli, as indicated by informal questioning at the end of the experiment.

[2] Since trials were randomized within blocks, one block constituted the smallest subset of trials within which all factors were balanced, thereby preventing us from considering only the first 72 or 144 trials, for instance.

[3] It may be worth remembering that Schneider and Shiffrin did not report any statistical tests on their effects.

We therefore duplicated condition CVM-HOMO in order to determine the level of performance that can be achieved when participants are aware of the set structure of the stimuli and motivated (recall the added monetary incentive). As shown in Fig. 4, the RTs obtained in the informed CVM-HOMO condition, labeled CVM-HOMO-I, are smaller than those obtained in the original condition, relabeled CVM-HOMO-O for contrastive purposes. This difference is not attributable to a speed-accuracy trade-off since the error rates were also smaller in CVM-HOMO-I (3%) than in CVM-HOMO-O (5%). Search times measured by the difference between M4D4 and M1D1 were smaller in the informed condition than in the original one, even in Block 1. In CVM-HOMO-I, search times decreased by about 50% from Block 1 to Block 12, compared with 30% in CVM-HOMO-O. Focusing on Block 12 data, the analyses confirmed the presence of a reliable $M \times D$ interaction, $F(4, 24) = 16.1$, $p < .0001$, $MS_e = 2{,}649$. The interaction was not of the same magnitude across groups, $F(4, 24) = 3.02$, $p < .04$, $MS_e = 2{,}649$, but sub-analyses confirmed that it was still significant for group CVM-HOMO-I, $F(4, 24) = 3.05$, $p < .04$, $MS_e = 2{,}649$. The overall $M \times D \times R$ interaction was significant, $F(4, 24) = 4.67$, $p < .01$, $MS_e = 1{,}339$, and so was the interaction also involving Groups, $F(4, 24) = 3.80$, $p < .02$, $MS_e = 1{,}339$. In this case, however, the sub-analyses showed the interaction not to be significant for the CVM-HOMO-I group, $F(4, 24) = .25$, $p > .90$, $MS_e = 1{,}339$.

These results show that, despite some improvement over the CVM-HOMO-O group, performance of the CVM-HOMO-I group remained characterized by a multiplicative pattern of memory and display load effects, even in Block 12. However, the (total) lack of interaction involving Response type in the results of the CVM-HOMO-I group suggests that performance was little affected by the nature of the response. For the informed group, the search times estimated from the difference between M4D4 and M1D1 in Block 12 yield search rates of about 17 ms/item in positive trials and 23 ms/item in negative trials, compared with 22 ms/item and 44 ms/item for positive and negative trials respectively, in the original condition. The 2:1 ratio of negative to positive trial search rates obtained in CVM-HOMO-O, as well as in the VM conditions, is congruent with the hypothesis of self-terminating search in positive trials

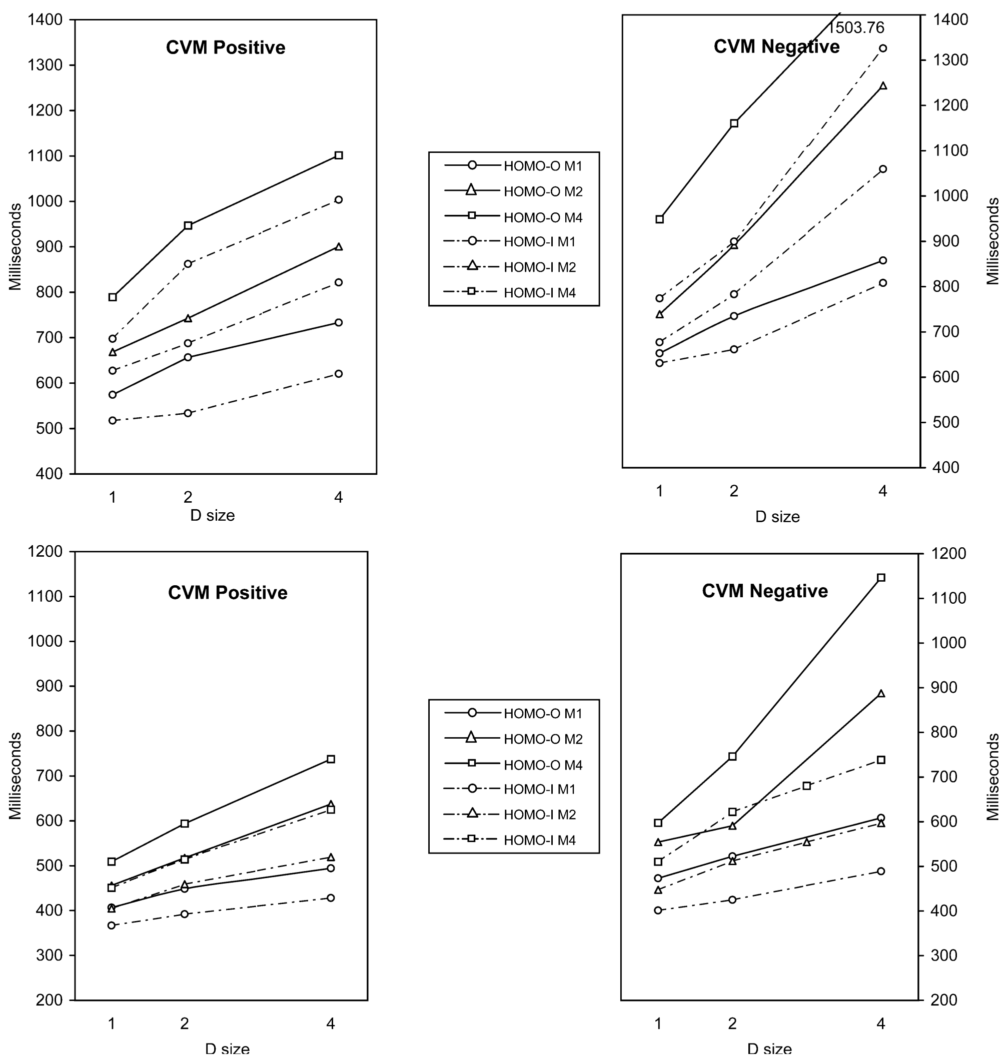


Fig. 4 Mean response times obtained in the first block (*top row*) and the twelfth block (*bottom row*) of training under the CVM-HOMO conditions of the experiment. Results are presented separately for positive (*left panels*) and negative (*right panels*) trials, as a function of replication (HOMO-O vs. HOMO-I), memory (M), and display (D) size

and exhaustive search in negative trials (see Townsend & Colonius, 1997). In this context, the similarity of the positive trial rates over groups suggests that search was also self-terminating for the CVM-HOMO-I group in positive trials. The difference concerning the negative trial rates across groups suggests that participants in the CVM-HOMO-I group may have terminated search early in negative trials. We will return to the issue of the termination rule later.

Comparison of the bottom panels of Figs. 2 and 4 shows that performance in the CVM-HOMO-I condition did not reach the level of proficiency achieved in the CM-HOMO condition. One interesting question is whether the differences between the CVM-HOMO-I and CM-HOMO conditions are similar to those obtained between the CVM-HETERO and CM-HETERO conditions. Then, according to the logic of stage additivity (Sternberg, 1966), we would be led to conclude that the effects of mapping and of categorical distinctions originate from different levels of processing. Alternatively, the effects of Mapping and Stimulus type could interact, thereby suggesting that these two factors affect the same processes. An ANOVA comparing the four groups mentioned showed the differences in RT to be significant over Stimulus conditions, $F(1, 11) = 13.0$, $p < .005$, $MS_e = 26{,}552$, but not over Mapping conditions, $F(1, 11) = 2.66$, $p < .14$, $MS_e = 26{,}552$. The Mapping × Stimulus interaction was far from significant, $F(1, 11) = .14$, $p > .70$, $MS_e = 26{,}552$. However, Mapping and Stimulus composition were involved in many interactions with the within-group factors, the largest of which also included M size and D size, $F(4, 44) = 3.70$, $p < .02$, $MS_e = 493$. Decomposing this interaction along the Stimulus dimension confirmed what our earlier analyses led us to expect. In the HOMO conditions, Mapping interacted with M size, $F(2, 22) = 29.5$, $p < .0001$, $MS_e = 1{,}306$, with D size, $F(2,22) = 25.1$, $p < .0001$, $MS_e = 684$, and with both, $F(4, 44) = 6.85$, $p < .001$, $MS_e = 493$, confirming the presence of a significant $M \times D$ interaction in the CVM condition, $F(4, 44) = 16.4$, $p < .0001$, $MS_e = 493$, but not in the CM condition, $F(4, 44) = .35$, $p > .80$, $MS_e = 493$. The $M \times D \times R$ interaction was not significant, $F(4, 44) = .59$, $p > .60$, $MS_e = 307$, nor was the interaction also involving groups, $F(4, 44) = 1.67$, $p < .18$, $MS_e = 307$. By contrast, in the HETERO conditions, Mapping failed to interact with M size, $F(2, 22) = .69$, $p > .50$, $MS_e = 1{,}306$, with D size, $F(2, 22) = 1.09$, $p > .35$, $MS_e = 684$, and with both, $F(4, 44) = .26$, $p > .80$, $MS_e = 493$. In fact, there was not a single reliable interaction involving Mapping in the performance of the HETERO groups (all $p > .35$).

These results show that Mapping and Stimulus composition do not have independent effects. Consistency of mapping had little, if any effect on performance when there was a pre-existing categorical distinction separating the targets and the distractors. In such HETERO conditions, the CVM and the CM groups produced very similar results. However, under HOMO conditions, the CVM and CM groups behaved quite differently. The efficiency and the additivity of memory and display search observed in condition CM-HOMO suggest that consistency of mapping does contribute to the automatization of performance when there is no pre-existing categorical distinction among the stimuli. In the absence of consistent mapping and of pre-established categorical distinctions (CVM-HOMO-I), performance remained poor and characterized by interacting Memory and Display size effects.[4]

Standard deviations

Most prior studies of visual-memory search have been exclusively concerned with mean performance, which is regrettable because standard deviations can provide important constraints on models of performance (e.g., Ashby, 1982). Schneider and Shiffrin (1977) have proposed an account of standard deviations but only in the VM-HOMO condition. In the present section, we analyze the standard deviations obtained in all conditions of the experiment, along with their relationship to the corresponding means. Fig. 5 presents the standard deviations obtained in the six original Mapping × Stimulus conditions in Block 12 of the experiment. The organization of Fig. 5 is similar to that of previous figures but the scale has been reduced by half (i.e., 500 ms) and the origin has been lowered to 0. The variability illustrated is the means of the within-participants standard deviations for every $M \times D \times R$ condition.

Comparison of Fig. 5 with the bottom panels of Figs. 1, 2, and 3 shows that the standard deviations (SD) behaved in a manner quite similar to the means (MN). The load effects on SDs appear small and additive in the two CM conditions as well as in the CVM-HETERO condition. By contrast, the SDs obtained in the two VM conditions and in the original CVM-HOMO condition appear to increase as a function of the product of M size and D size. Moreover, the increase is more pronounced in negative trials than in positive trials. This is a surprising result given traditional accounts of VM performance. The fact that load effects on MNs are smaller in positive than in negative

[4] It could be argued that the comparisons involved in the ANOVA reported were unfair to participants in the CVM groups since there were twice as many different stimuli serving as targets in CVM conditions as in CM conditions. Consequently, the participants in CVM conditions received half as much practice per target as those in CM conditions. Such a difference may not be negligible. According to exemplar theories of learning (e.g., Logan, 1988, 1992; Nosofsky Palmeri, 1997) and memory (Hintzman, 1986), amount of training per stimulus is much more critical to performance than total amount of practice. In order to equate amount of practice per target, we compared performance obtained on the 8th training block in each of the two CM conditions (HOMO, HETERO) to that obtained on the 16th training block in each of the two CVM conditions (HOMO-I and HETERO). The results of these analyses support the same conclusions as those described in the text: The $M \times D$ interaction was reliably smaller, though not yet absent, in condition CM-HOMO in Block 8 than it was in CVM-HOMO-I in Block 16. No such interaction was present under HETERO conditions.

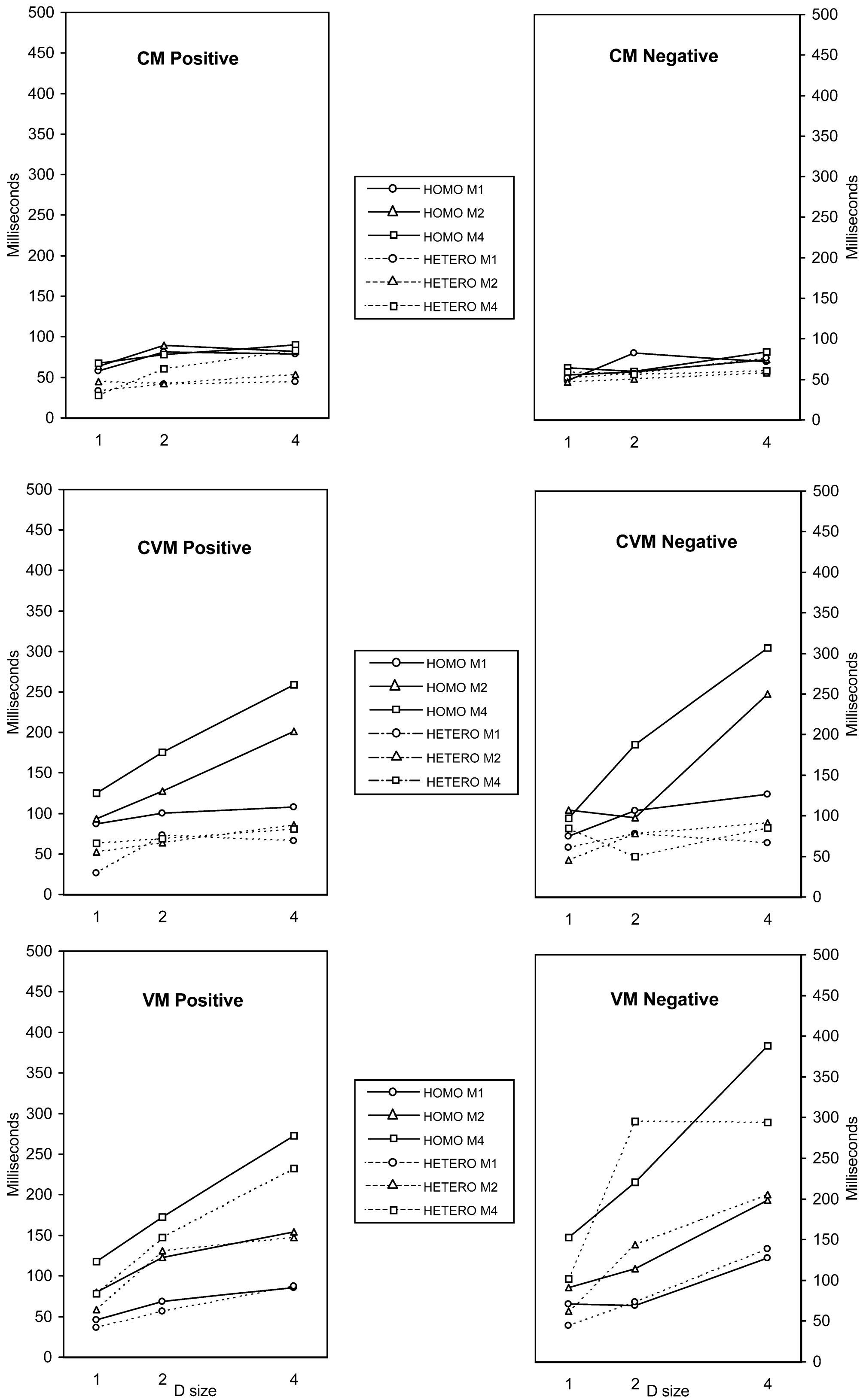


Fig. 5 Standard deviations in response times obtained in the 12th block of training under the six original Mapping × Stimulus conditions of the experiment. Results are presented separately for positive (*left panels*) and negative (*right panels*) trials, for CM (*top*), CVM (*middle*), and VM (*bottom*) mapping, as a function of stimulus composition (HOMO vs. HETERO), memory (*M*), and display (*D*) size

trials is usually attributed to the fact that search is self-terminating in positive trials and exhaustive in negative trials. For the very same reason, serial self-terminating earch (SSTS) models predict that load effects on SDs should be more pronounced in positive trials than in negative trials, as explained below.

In SSTS models, the expected RT for a given load condition, E (L), can be conceived as being equal to

$E(N)\,E(\mathrm{T}) + E(\mathrm{I})$, where N is the number of items to compare, T is the time needed for a single comparison and I, the time related to the intercept processes. When search is exhaustive, the total number of comparisons required is N, the number of items. When search is self-terminating, $E(N)$ corresponds to $(N + 1)/2$. Traditional SSTS models, therefore, predict negative trial means (MN−) to increase about twice as fast as positive trial means (MN+), as originally noted by Sternberg (1966). In SSTS models, the variance of the RTs in a given load condition, Var (L), can be conceived (Townsend and Ashby, 1983) as being equal to $E(N)$ Var (T) + Var (N) E^2 (T) + Var (I). When search is exhaustive, Var (N) = 0 within each load condition. The variances should therefore increase linearly with load in negative trials. By contrast, when search is self-terminating, Var (N) equals $(N + 1)(N - 1)/12$ within each load condition. Therefore, the variances of positive RTs should increase faster than linearly with load, eventually exceeding the variances obtained in negative trials, as was previously observed by Schneider and Shiffrin (1977, Experiment 2).

The results obtained in our VM and CVM-HOMO-O conditions fail to match these predictions in many respects. First of all, it is the standard deviations rather than the variances that seem to increase linearly with load. Secondly, the increase looks linear not only for negative trials, but also for positive trials. Thirdly, the increase is more pronounced in negative than in positive trials. It may be worth noting that the results obtained by Ward and McClelland (1989) from a conjunctive, visual search task also failed to match the predictions of SSTS models. They found variances to increase faster than linearly with load in both positive and negative trials, the latter exhibiting a larger increase than the former.

Adopting a different perspective on SSTS models can accommodate the pattern of load effects on positive trial SDs obtained in our experiment. This change of perspective consists in postulating that the variance of individual item comparison times, Var (T), is negligible. We will use the label SSTS* to distinguish models based on this assumption from more traditional SSTS models. Given the postulate of zero variability in individual comparison times, then Var (L) = Var (N) E^2 (T) + Var (I). If we further assume that Var (I) is negligible, SD(L) becomes equal to SD (N) E (T). On positive trials:

$$\mathrm{SD}+(\boldsymbol{N}) = \sqrt{((N+1)n-1)/12}$$

which is closely approximated by

$$\frac{N}{\sqrt{12}}$$

Thus, the load effects on

$$\mathrm{SD}+(L) = \frac{N}{\sqrt{12}}E(\mathbf{T})$$

and on

$$E+(L) = \frac{N+1}{2}E(\mathbf{T})$$

become linear functions of N. Furthermore, this SSTS* model predicts the ratio of load effects on SD+ and MN+ to be about .60, since

$$2/\sqrt{12} = 0,577$$

To estimate the ratio of load effects on positive trial means and standard deviations, we computed linear regressions of SD+ on MN+ over the nine load conditions in the experiment (from M1D1 to M4D4). This approach bypasses the issue of the additive vs. multiplicative nature of memory and display load effects on either variable, thereby allowing easy comparisons across experimental conditions. Separate regressions were performed on the results of each individual participant, including those in condition CVM-HOMO-I. Individual r^2 obtained for positive trials ranged from .22 to .99. For 25 of the 27 participants, r^2 exceeded the .444 value required for the linear trend to be considered significant at the .05 level (d.f. = 7). Table 2 shows the mean intercepts and slopes of these individual regressions for each group of participants, along with the mean squared regression coefficients (r^2). Table 2 also gives the results of similar regressions performed on negative trial means (MN−) and standard deviations (SD−). In this case, individual r^2 ranged from .11 to .99, 24 of which were statistically significant. Finally, Table 2 summarizes the results of regressions done on MN− as a function of MN+ for all participants. For MNs, 25 of

Table 2 Regressions of the positive and negative MNs and SDs obtained in Block 12 of the experiment averaged across participants

Condition	Mean intercept	Mean slope	Mean r^2
ΔSD+:ΔMN+			
CM-HETERO	−173.7	.62	.701
CM-HOMO	−81.5	.49	.588
CVM-HETERO	−154.4	.55	.783
CVM-HOMO-I	−203.9	.65	.956
CVM-HOMO-O	−161.2	.56	.930
VM-HETERO	−214.2	.68	.965
VM-HOMO	−202.5	.67	.954
ΔSD−:ΔMN−			
CM-HETERO	−89.9	.43	.509
CM-HOMO	−81.2	.36	.584
CVM-HETERO	−151.1	.50	.725
CVM-HOMO-I	−104.2	.41	.808
CVM-HOMO-O	−109.1	.38	.918
VM-HETERO	−149.5	.45	.903
VM-HOMO	−192.0	.53	.933
ΔMN−:ΔMN+			
CM-HETERO	237.9	.46	.505
CM-HOMO	124.9	.82	.803
CVM-HETERO	156.1	.74	.746
CVM-HOMO-I	−76.8	1.32	.958
CVM-HOMO-O	−334.0	1.88	.933
VM-HETERO	−258.7	1.80	.953
VM-HOMO	−225.1	1.75	.945

the 27 individual r^2 were significant, ranging from .15 to .98. Regressions of SD− on SD+ have not been performed, the relationship between these two variables being redundant on the other three relationships considered.

As shown in the top part of Table 2, the mean slopes of the regressions of SD+ on MN+ are all within about ± .11 of the .60 value predicted by SSTS* models. Recall that this prediction was based on the assumption that the variability of the intercept processes, Var (I), was null, an assumption that is certainly wrong. After comparing the results of two experiments, Hopkins and Kristofferson (1980) concluded that the variance introduced by a key press contributed 50 ms^2 to Var (I). This corresponds to a SD of about 7 ms, which is about 10 times smaller than the SDs observed here. Of course, the motor response is only one of the processes contributing to the intercept. Using various values, we informally estimated that the .60 prediction may only be accurate within a range of ± 10%. This extended range would also allow the mean slope of group CM-HOMO to fit. In short, the ratios of means and standard deviations obtained suggest that the same serial, self-terminating search processes may have been involved in positive trials in all the Stimulus and Mapping conditions of the experiment.

The middle part of Table 2 shows the results of the regressions of SD− on MN−. The average slopes obtained range from .36 to .53, which again suggests a similarity of processing over all conditions of the experiment. In traditional SSTS models, search is assumed to be exhaustive in negative trials. Consequently, the number of comparisons required in any load condition is constant and equal to the total number of items to compare. So, with no variability in individual comparison times, exhaustive search leads SSTS* models to predict no load effect on SD−, which is clearly not supported by the data. The problem can be remedied by abandoning the postulate of exhaustive processing. Stopping search early in negative trials (i.e., before all items are considered) produces rapid "no" responses that reduce the load effects observed on MN− and increase the effects observed on SD−. As will be shown shortly, we can provide a fair account of the slopes of SD− on MN− by assuming that search is not always exhaustive. Stopping search early in positive trials (i.e., before the target is located) produces misses. Since errors are usually excluded from RT analyses, early termination of processing does not affect the predictions of SSTS* models concerning the load effects on MN+ and SD+.

Another problem encountered by SSTS models is that they cannot explain the differences in the load effects on MN+ and MN− across mapping and stimulus conditions (shown in Figs. 1, 2, 3, and 4), which underlie the ratios given in the bottom part of Table 2. The ratios of two VM groups and group CVM-HOMO-O are close to 2:0, in agreement with the postulate of self-terminating vs. exhaustive processing that characterizes SSTS models. This problem is resolved in the model, described next, by letting the amount of processing vary across conditions.

Model

Our model, called the Sufficient Features Model or SFM, can be roughly characterized as having an SSTS* architecture augmented with information reduction mechanisms. SFM differs from traditional SSTS models in that such models assume that whole characters form the basic units of processing. Because information reduction is difficult to achieve on such a level, all comparisons in SFM occur at the level of elementary features. Note that the postulate of negligible variability in individual comparison time, which characterizes SSTS* models, has much more plausibility when the comparisons are thought to involve elementary features instead of whole characters, whose complexity can vary. SFM also differs from dual-process theories of automatization in that the same processes are involved in all Mapping and Stimulus conditions. Moreover, these processes are assumed to remain constant throughout training. Only the number of feature tests performed by SFM varies.

Basically, SFM continuously attempts to reduce to a minimum the number of feature tests performed before reaching a decision concerning the presence or absence of a target on the display. Due to its dynamic character, the model does not lend itself easily to analytical predictions. A computer simulation was therefore used to test the validity of the principles involved. The ideas behind SFM being rather general, we made various choices in order to implement the model in a computer program. Some of these choices were motivated by our results and by other findings in the field of visual attention. Others were made for ease of computation. So, in presenting the computer implementation of SFM, we will distinguish the theoretically important from the technically convenient. After describing the model and its implementation, we will turn to the results of the simulations, which will be presented in the same order as the empirical results of the previous sections.

Description of the model

Featural representation

In order for SFM to reproduce the empirical results obtained, a featural representation for the characters used in the Experiment had to be adopted. Although a lot is now known about the features that guide visual search (see Wolfe, 1994 for a review), we do not know which of the candidate set of elementary features people actually use when dealing with alphanumeric characters, not to mention that different persons may parse such characters differently. We avoided the issue of the "true" featural representation by using a schematized representation similar to that of McClelland and Rumelhart's (1981, Rumelhart & McClelland, 1982) word recognition model. The feature set used to represent our stimuli is shown in Fig. 6. It is composed of 12 horizontal, vertical or diagonal line segments, each occupying a specific position within the character matrix. The schematic representation of the various subsets of stimuli used in the Experiment is also shown in Fig. 6. Each character is represented by a conjunction of six features or less. Some features are shared by many characters, but there is always at least one feature that allows the characters to be discriminated from each other. In other words, no character is fully included in another one. Therefore, SFM does not have to use the absence of some feature to identify the stimuli (as would be the case for O among Qs, for instance). Although consistent with Treisman and Gormican's (1988) findings, this constraint was adopted mainly to simplify the comparison algorithm. Because of the small number of features involved, some characters look strange. A larger feature set would have allowed the design of more realistic-looking characters but again, the goal of SFM was to test the feasibility of an information reduction approach to the automatization of search rather than to test the validity of a specific featural representation.

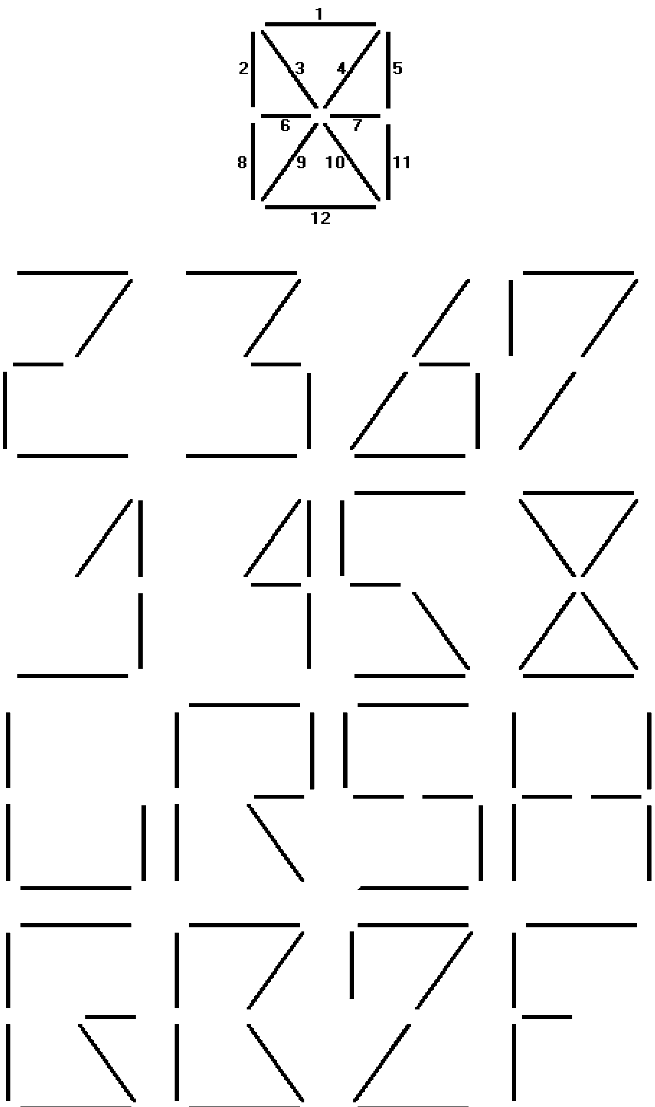


Fig. 6 Illustration of the features and characters used in the simulation

There are differences in the discriminability of the stimuli shown in Fig. 6. More importantly, discriminability is not equated across the HOMO and HETERO conditions (as it may not have been in the Experiment since we exerted only an approximate control over this factor). Given the schematic representation of the characters in Fig. 6, which was designed by a then naive assistant, each of the four stimuli used as targets in the CM and CVM HOMO conditions shares an average of 2.6 features with each of the four stimuli serving as distractors, whereas the number of features shared by the targets and distractors used in the corresponding HETERO conditions averages only 2.1. Because of this difference, the null hypothesis tested by SFM is that the different results obtained in HOMO and HETERO conditions are attributable to differences in featural similarity within and across the digit and letter categories, as hypothesized by Treisman and Gelade (1980), Krueger (1984), and others. As we will see, however, this difference in discriminability does not allow SFM to account for all the categorical effects obtained.

Learning processes

Throughout training, SFM continuously attempts to find the features that most efficiently discriminate the stimuli serving as targets from those serving as distractors. In its simplest implementation, SFM disposes of a master list containing the 12 features shown in Fig. 6. The features are arranged in a random order prior to training, which is tantamount to assuming that the relative diagnosticity of the various features is not known at the beginning of the experiment. This assumption is probably wrong since participants had many years of experience at discriminating letters and digits prior to the start of the experiment. However, we did not attempt to reproduce early performance.

Throughout training, the features are reordered by being raised one position at a time in the list. Feature promotion occurs in negative trials, a maximum of one feature being promoted in any given trial. The rule used for feature promotion is that the feature should be present in at least one character in the memory set, i.e., among the potential targets, and absent from all the distractors in the display set. If more than one feature satisfy this condition, the feature selected for promotion is the one that already occupies the highest position in the list. We do not attribute much psychological validity to the specifics of this procedure. Many other feature promotion schemes could have been used. For instance, we could have chosen to promote features in positive instead of negative trials, raising a feature present among the members of the memory set and absent from the distractors. Feature promotion could also have been restricted to the features of the target. What these choices affect is the rate of learning, not the end representation. Since we made no attempt at reproducing the exact shape of the learning curves, we also ignored other factors that almost certainly affect the learning rate in real participants. For instance, the feature promotion mechanism is error-free and SFM does not forget the relative diagnosticity of features over trials, sessions, or weekends.

The only critical aspect of the feature promotion procedure is that the features promoted should be diagnostic, i.e., present among some targets and absent from some distractors. It does not suffice to simply promote features that are shared by targets, irrespective of the featural composition of the distractors, or vice versa. However, any of the promotion schemes previously described will ultimately produce an ordering that not only reflects the dissimilarities between targets and distractors, but that is also sensitive to the similarity among distractors and to the similarity among targets, as Duncan & Humphreys (1989; see also Humphreys et al., 1989) suggested. Indeed, features that allow discriminating a given target from many distractors are more likely to exist when the distractors are similar than when the distractors share few features. Features that are frequently discriminative will be promoted more often and eventually reach a higher position than less discriminative ones. And the more similar the targets are, the more likely it is that the targets will share the same diagnostic features, which will therefore be repeatedly promoted to eventually dominate the feature list. Given the complexity of the stimuli that are typically used in visual-memory search tasks, it is perhaps unlikely that the feature list will come to be dominated by a single feature. However, it did happen in the simulation when the target set was composed of the digits <2, 3, 6, 7> and the distractor set of the letters <L, R, S, H>, the feature number 4 being present among all targets and absent from all distractors (see Fig. 6).

Information reduction is achieved in SFM by adjusting a boundary or threshold in the ordered feature lists so that below-threshold features can be ignored during search. Two kinds of features come to fall below threshold: Features that have little power to discriminate the targets from the distractors and features that are so strongly correlated with the diagnostic features as to be redundant. So, SFM allows for some improvement in performance, even when the same stimuli serve as targets and as distractors, as in VM conditions. However, there can be much greater improvement when the targets and the distractors come from different sets of characters, as in CM and CVM conditions and, within such conditions, the amount of information reduction possible further depends on the discriminability of the two sets of characters used.

Like the other processes in SFM, the threshold adjustment procedure is quite simple. The threshold is set at the bottom of the list prior to learning. It is raised by one position after every succession of 36 correct positive responses. Raising the threshold causes a reduction in the number of feature tests performed, thereby increasing search speed. However, it also leads to errors. At some point, the neglect of too many features will render some distractors indistinguishable from the targets, inevitably yielding false alarms. The risks of confusing a distractor with a target may increase with the number of items in the memory set and on the display, but false alarms do not depend on load per se. Load effects on the false-alarm rate are mediated by a correlated increase in the featural similarity of the characters in memory and on the display (see Zenger & Fahle, 1997). To maintain an acceptable level of false alarms, the threshold is lowered one position after every incorrect positive response (see Chun & Wolfe, 1996 for a similar staircase mechanism). This procedure maintains a false-alarm rate of about 3% (1 over 36).

Search and comparison processes

We assume that there is a delay, albeit small, before search can start after display onset. We think of this delay as reflecting some psychophysical process whereby visual information emerges from noise. Such a delay prior to search onset was observed by Hockley (1984, Fig. 5) and by Cousineau and Shiffrin (2003). It may also be at the origin of the residual search slopes obtained in visual search studies even when search is thought to be automatic. In SFM's current implementation, this stage contributes a small increase in search time with display size, thereby affecting SFM's predictions concerning load effects on the RT means. However, it does not affect SFM's predictions concerning the standard deviations since the duration of this stage is assumed to have negligible variability within each display size. Like all other processes in SFM, this stage remains the same throughout training under all conditions of the experiment.

Only the above-threshold features are considered by the search and comparison processes. This assumption is in agreement with Bundesen's (1990) suggestion that information found to be useful in the past can come to guide visual search. If only one display item contains any of the above-threshold features, we assume that attention is immediately drawn to the corresponding location. This is tantamount to assuming that feature search is parallel over the display, in agreement with the Feature Integration Theory (Treisman & Gelade, 1980). If many display items contain diagnostic features, then SFM picks one of the corresponding locations at random. So, the order of display search is not fully determined by the order of the features in the list. Feature ordering was initially proposed by Fisher (1984, 1986) as a method of optimizing the feature tests performed by a limited-capacity system. Order was critical in Fisher's model because there was no threshold to prevent testing irrelevant and/or redundant features. The presence of a diagnosticity threshold in SFM allows the model not to rely as heavily on the exact ordering of features.

Display search is not affected either by the number of diagnostic features present in any given display location. In other words, display items containing many above-threshold features are not given any priority over locations containing fewer diagnostic features. Visual attention in SFM is guided only by a disjunctive set of features (i.e., the set of above-threshold features), as initially proposed by Treisman and Gelade (1980). More recently, Treisman & Sato (1990) and Wolfe (1994, Wolfe, Cave, & Franzel, 1989) have shown that some feature conjunctions can also guide search. However, the feature conjunctions that were found to allow automatic detection are quite different from those considered here in that the former generally involve different attributes (e.g., color, orientation etc.). There is little evidence (Treisman & Gormican, 1988; but see Kyllingsbaeck, Schneider, & Bundesen, 2001) that conjunctions of line segments such as those that make up alphanumeric characters can attract attention.

Display search in SFM is fully automatic in that no time is allocated for focusing on a given display location. In this sense, SFM does involve automatic attention attraction, as was also proposed by Schneider and Shiffrin (1977, Shiffrin & Schneider, 1977). However, Schneider and Shiffrin assumed that automatic attention attraction could only occur after extensive training under the CM condition whereas in SFM the same mechanism is operant throughout training, even in VM conditions. Within SFM's framework, the only difference between CM and VM conditions is that many more features remain above threshold in VM conditions, so that many more display locations need to be considered, these locations being examined sequentially.

When a display location is selected, SFM attempts to determine whether the item in that location matches an item in the memory set. Since all memory set items were equally likely to be the target in any given trial of our experiment, SFM picks a candidate from these items at random. Comparison of the display and memory set item selected proceeds on a featural basis and it is restricted to the above-threshold features. If the feature that attracted attention to the current display location is shared by the memory item selected, comparisons continue with another of the above-threshold features of the display item under consideration. In other words, the comparison process attempts to determine whether the conjunction of diagnostic features present in a given display location can also be found in the selected memory set item. Processing of the selected memory set item stops as soon as one of the above-threshold features of the display item fails to match, in which case another memory item is randomly selected for comparison with the current display item. If no memory item is found to match all the above-threshold features of the current display item, processing starts anew with another display location, randomly selected from those containing above-threshold features. The comparison process is therefore serial over display items, as Treisman and Gelade (1980) proposed conjunctive search to be, as well as over memory items.[5]

In SFM, successive sampling of the memory and display items is done without replacement. Since the memory set is cycled through faster than the display set, this implies that memory items can be repeatedly tested but that each display item is considered only once. This is tantamount to assuming that the search process keeps track of the display items considered, in agreement with recent evidence (McCarley, Wang, Kramer, Irwing, & Peterson, 2003; see also Horowitz & Wolfe, 1998, and von Mühlenen, Müller, & Müller, 2003). The mechanism that prevents return to previously visited locations is not specified in SFM's current implementation.

The number of feature comparisons performed by SFM is the main determinant of the model's performance. By contrast with other models, such as the one proposed by Schneider and Shiffrin (1977) to account for VM performance, SFM has no time parameter for switching from one item to the next in the memory set (or on the display). Schneider and Shiffrin used such a parameter to account for the fact that, in most conditions where the $M \times D$ product is equal, the mean response times are larger when M size is larger than D size, which is also the case in our results. SFM easily accounts for this fact: Since display search is guided by diagnostic features, increasing display size does not entail a corresponding increase in the number of feature comparisons performed. Imagine, for instance, that only one feature remains above threshold and that this feature allows all targets to be discriminated from all distractors. In such an extreme situation, attention would immediately be drawn to the target location in positive trials, irrespective of the number of items on display. However, the item in this location would still have to be compared with the items in memory (since they are all potential matches), resulting in a memory set size effect. In negative trials, no display item would possess the critical feature, so that no display item would be located and the memory set would be ignored altogether. Therefore, SFM can also account for the fact that memory load effects were less pronounced in negative than in positive trials in CM conditions, especially in the CM-HETERO condition.

Finally, note that the directionality of the search and comparison processes is critical to SFM's performance. SFM would not be able to account for the pattern of results just described if the memory set items were searched and compared with the display items. By virtue of being potential targets, the members of the memory set necessarily possess diagnostic features, so that just

[5] In an early version of SFM, each feature of a given display item was tested in parallel with all the items in the memory set. The load effects obtained were generally too small compared with those observed in VM conditions.

about any increase in memory set size would inevitably yield an increase in the number of feature tests performed, even in negative trials, if search operated on the memory set instead of the display set. To account for the results, SFM must search and compare the items on the display with those in the memory set, rather than vice versa, in contrast to what Schneider and Shiffrin originally proposed.

Identification process

In models where whole items serve as the basic unit of processing, it must be assumed that the stimuli are identified before they are compared. Sternberg (1966), for instance, postulated that the displayed item was encoded before it could be compared with the items in the memory set. There is no such encoding stage in SFM: The displayed items are recognized while being compared with the memory set items. In VM conditions, there are enough above-threshold features to allow identification of display items during the comparison stage. However, when there are few above-threshold features, as in CM conditions, a target can be detected before it is uniquely identified. This poses an empirical difficulty. As Schneider and Shiffrin (1977) remarked, CM performance does not differ much from VM performance in M1D1 conditions. By contrast, the number of feature comparisons performed by SFM may be as small as 1 in CM conditions and as large as 6 in VM conditions when memory and display set size equal 1. So, SFM's performance would clearly differ across mapping conditions, if the response was emitted immediately upon reaching the diagnosticity threshold.

We solved this difficulty by postulating that the presumed target had to be fully identified before a positive response could be issued. Like the comparison process, the identification process is assumed to be feature-based and sequential but it involves only below-threshold features. In its current implementation, SFM simply adds the number of below-threshold features present in the memory item under consideration to the number of feature tests performed before reaching the threshold. In M1D1 conditions, the identification process exactly compensates for the difference in above-threshold features across CM and VM conditions, since the number of below-threshold features that need to be considered by the identification process is the complement of the number of above-threshold features considered by the comparison process. In short, whether the features are above or below threshold does not matter very much when load is minimal. However, in large load conditions, the number of above-threshold features can lead to large, multiplicative effects on search that are not at all compensated by the smaller, additive effects originating in the identification process. So, the partitioning of the feature list into two different subsets, each serving a different function, turns out to be a fairly simple and powerful device to account for the response times.

Whether the below-threshold features considered by the identification process belong to the display item or to the memory set item has little consequence on SFM's predictions concerning response times. We have chosen to consider the below-threshold features of the memory set item because it allows SFM to make interesting predictions concerning errors. Given the featural representation in Fig. 6, SFM can detect the presence of a target after only one or two feature comparisons in the most efficient conditions. When many memory set items share these diagnostic features, SFM can correctly detect the presence of a target while identifying the wrong memory item as the target. SFM can also identify a target when none was actually present on the display.

The distinction between detection and identification in SFM has other interesting properties, apart from accounting for false alarms. When target detection requires only one or two features, an average of 4.5 additional features remain to be considered before the identification of the candidate memory item is completed, during which display information is not processed. Such an identification process may therefore be at the origin of some of the transient blindness phenomena reported in the literature. For instance, Ward, Duncan, and Shapiro (1996, see also Shiffrin & Schneider, 1977 for a similar result) found that participants could not detect the presence of a second target presented up to five times longer that the base scan rate. They attributed this result to a long dwell time on the first target, which prevented processing of the second target. However, they provided no reason for the extended dwell time. The hypothesis that the participants are blind to the display during the identification process would account nicely for such results.

Termination rule

The purpose of the termination rule in SFM is to account for the performance obtained in negative trials. As discussed previously, exhaustive search cannot accommodate load effects on negative trial SDs in SSTS* models. This is true whether exhaustivity is defined over all items in the memory and display sets, as in traditional models, or over a subset of items containing above-threshold features, as in SFM.

Knowing little about the exact nature of the mechanisms that cause search to stop when no target is present, we decided to adopt a probabilistic termination rule. This rule is implemented as follows: At trial onset, the system specifies the maximum number of feature tests that can be performed, given the number of above-threshold features present in the display and memory set items. SFM performs the maximum number of feature tests with probability p. Processing stops after a random uniform number of feature tests (bounded by the maximum) with probability $1 - p$. In such trials, SFM performs on average half the number of tests done in exhaustive trials. SFM issues a negative response after reaching the criterion number of comparisons set at trial onset, provided that no target has been found beforehand.

The termination rule can cause search to stop before the target is located in positive trials, thereby generating misses. To avoid an excessive number of misses, the *p* parameter is dynamically adjusted throughout training, using a staircase procedure. The odds of exhaustive search are decreased by one-tenth after each succession of 20 correct negative responses—the odds of exhaustive search to random stop being expressed by $r : 1$, and, as usual, $p = r / (r + 1)$—and they are increased by one-tenth after every miss. This procedure ensures that the average percentage of misses will stabilize at about 5% in all conditions.

In summary, SFM continuously attempts to reduce the number of feature tests performed by limiting display search to locations containing diagnostic features and by trying to avoid processing all the diagnostic features present among display and memory items. Two parameters are adjusted throughout training: The number of features that are above the diagnosticity threshold (labeled *B* for boundary), and the probability *p* of exhaustive processing. Although *B* contributes to the definition of exhaustivity, it is important to note that *p* and *B* are adjusted independently of each other, so as to avoid an excessive miss rate in the former case, and an excessive false-alarm rate in the latter case.

Results and discussion

The program was submitted to the same conditions as the human participants, using the same stimulus assignments and following the same training schedules. Despite their initial independence, the parameters *p* and *B* ended up being correlated after training, as the top part of Table 3 shows. This correlation can be understood as follows: When the diagnosticity threshold is high (a small *B*), it is unlikely that the few diagnostic features that exist are shared by many distractors. Chances are therefore high that the display locations picked up by the first feature tests will contain a target. If not, chances are slim that further tests would reveal the presence of a target. So, processing does not need to be exhaustive, and *p* can be small. The consequence of self-terminating search in positive trials is that misses will occur. To maintain a miss rate of 5%, 2:1 odds of exhaustive processing appear sufficient in CM conditions. Misses are of course more frequent for higher loads (Zenger and Fahle, 1997). When the diagnosticity threshold is low, it is likely that some of the diagnostic features are also present among distractors. Consequently, the search process should not stop after just a few mismatching tests. In VM conditions, the odds of exhaustive search raise to about 3:1. If absolutely no errors were allowed, the probability would have to climb to 100% in all conditions.

Our modeling efforts were strictly aimed at accounting for response time data obtained in the 12th block of training. Four parameters were estimated to scale the results of the simulation to the RT means. These parameters, which were only roughly estimated, are given in the bottom part of Table 3. E (T) is the expected time of one feature test. It was estimated at 6 ms per feature. When multiplied by 5 or 6 features to yield a comparison time per item, this estimate approximates the results obtained in many experiments involving VM conditions, starting with that of Sternberg (1966). $E\,(\mathrm{I}+)$ and $E\,(\mathrm{I}-)$ correspond to the time taken by intercept processes in positive and negative trials (in milliseconds). Again, the estimates are in agreement with results generally found in the literature. Finally, the duration of the psychophysical stage E_P was set to one tenth of E (T) for each feature on the display. Note that none of the four estimated times is assumed to have any variability: $\mathrm{Var}\,(\mathrm{T}) = \mathrm{Var}\,(\mathrm{I}+) = \mathrm{Var}\,(\mathrm{I}-) = \mathrm{Var}\,(E_P) = 0$. Therefore, any variations in the results come from the number of features considered by the search and identification processes. Note also that the four estimated parameters are identical for all simulated groups (and participants) so that any difference observed between conditions does not depend on these free parameters.

Response time means

Consistent mapping Figure 7 shows the mean simulated response times produced by the model in positive and negative trials in the various $M \times S \times R$ conditions of the experiment. Let us look at the CM conditions first. The top part of Fig. 7 shows that performance was better in the HETERO than in the HOMO condition. The *M* size and *D* size effects are small in the HETERO condition, and they clearly fail to interact. However, the *M* size effects do interact with Response, being more pronounced in positive trials than in negative trials (although this is not very visible in Fig. 7). In short, the pattern of load effects on the means obtained in the HETERO condition corresponds quite well to that obtained with the participants. *M* size and *D* size also affect performance in the HOMO condition. However, the model shows a tendency for these factors to interact. Comparison with the bottom part of Fig. 2 shows this tendency to be more pronounced for the model than for the participants.

The differences between the CM-HOMO and CM-HETERO conditions exhibited by the model are all attributable to the relative discriminability of the targets and distractors used in these conditions. As previously mentioned, for some stimuli, one feature was unique to the digits, so that this feature could rise in the feature list and facilitate the task when the memory set was composed of digits. Other features, unevenly distributed over digits and letters could also have facilitated the task in the HETERO condition. Discriminability affects the number of diagnostic features (*B*), which equals 1.9 in

Table 3 Parameters of the simulation on the 12th block of training averaged over simulated participants

Learned parameters		
Groups	p (%)	B
CM-HOMO	50	3.2
CM-HETERO	38	1.9
CVM-HOMO	62	4.5
CVM-HETERO (one list)	44	3.2
CVM-HETERO (two lists)	42	2.6
VM-HOMO	72	8.7
VM-HETERO (one list)	71	7.7
VM-HETERO (two lists)	72	7.3
Scaling parameters common to all groups		
E (T)	6 ms/feature	
E (I+)	340 ms	
E (I−)	400 ms	
E_P	.6 ms/feature	

the HETERO condition and 3.2 in the HOMO condition when averaged over the simulated participants. M size and D size effects are small and they fail to interact when the number of above-threshold features is small.

Categorical varied mapping The middle part of Fig. 7 shows the performance of the model in the CVM conditions. The solid line represents the results obtained with the HOMO stimuli. Those obtained with the participants in groups CVM-HOMO-O and CVM-HOMO-I are shown in the bottom part of Fig. 4. The magnitude of load effects on the model's performance resembles more that obtained with group CVM-HOMO-I. SFM therefore behaves like the participants who were informed of the set composition of the stimuli and motivated. The model could not possibly fit the results of both groups simultaneously.

Comparison of the middle and top parts of Fig. 7 shows that the M size and D size effects obtained in the CVM-HOMO condition are much more pronounced than in the corresponding CM condition, and so is the $M \times D$ interaction. Because the sets of stimuli switch roles as targets and distractors in CVM conditions, the features that are diagnostic of one set compete with the diagnostic features of the other set to reach the top position in the feature list. Moreover, the threshold cannot be raised very high in the CVM-HOMO condition ($B = 4.5$) since the features that are diagnostic of targets in some trials are found among distractors in other trials.

It is important to realize that the differences between the CM- and CVM-HOMO conditions would not have been obtained had the model updated a separate feature list for each individual stimulus instead of updating a single representation for all stimuli. Since the same sets of stimuli served as targets and distractors in both conditions (switching role within participants in one group and across participants in the other), the average discriminability of each individual target was the same for both groups. SFM would therefore have produced identical asymptotic performance in both conditions, had the learning mechanism reordered and raised a threshold in the feature lists representing each individual target. SFM behaves differently in the CM and CVM conditions because it updates a single feature list. At the theoretical level, this means that search in SFM is driven by a general representation that summarizes information about the sets of stimuli used in the CM- and CVM-HOMO conditions.

The need for summary mental representations (Smith & Medin, 1981) is also evidenced in the CVM-HETERO condition. The dashed line shows the model's performance when SFM updates a single feature list, as described for the other conditions. The M size by D size interaction is slightly smaller than in the CVM-HOMO condition, as can be seen by the amplitude of the respective fans. This difference is again due to differences in the relative discriminability of the stimuli across HOMO and HETERO conditions. However, this facilitation far from accounts for the performance obtained with the real participants, shown in the bottom part of Fig. 3.

The dotted line in Fig. 7 shows the model's performance when SFM updates two separate feature lists, one for letters and one for digits. The two lists are composed of the same features but their order and the position of the diagnosticity threshold can differ across lists. When digits have to be found among letters, the digit feature list is activated and updated in the manner described. Similarly, the letter feature list is activated and updated when the targets are letters. Since different lists are involved in different trials, performance is no longer affected by switching the target and distractor sets. The digit feature lists come to be dominated by features that tend to be shared by digits but not by letters, and vice-versa for the letter feature list. This modification to the model allows the simulation to reproduce the small memory and display load effects shown at the bottom of Fig. 3 for the CVM-HETERO condition. Note that the dual-list modification to the model does not change anything for the CM-HETERO condition where only one feature list is updated since only digits or only letters serve as targets for any given participant. Nor does it change anything for the HOMO conditions since these conditions involved only letters or only digits for any given participant.

The two feature lists used in the CVM-HETERO condition can be viewed as representing categorical knowledge, the above-threshold features being those that best allow the members of one category to be distinguished from those of another. Inasmuch as digits and letters form two distinct categories, it is normal that each has a distinct categorical representation (i.e., ordered feature list). It is very likely that our educated participants had long acquired such representations for letters and digits. This would explain why the performance of the CVM- and CM-HETERO groups was very efficient from the beginning of the experiment.

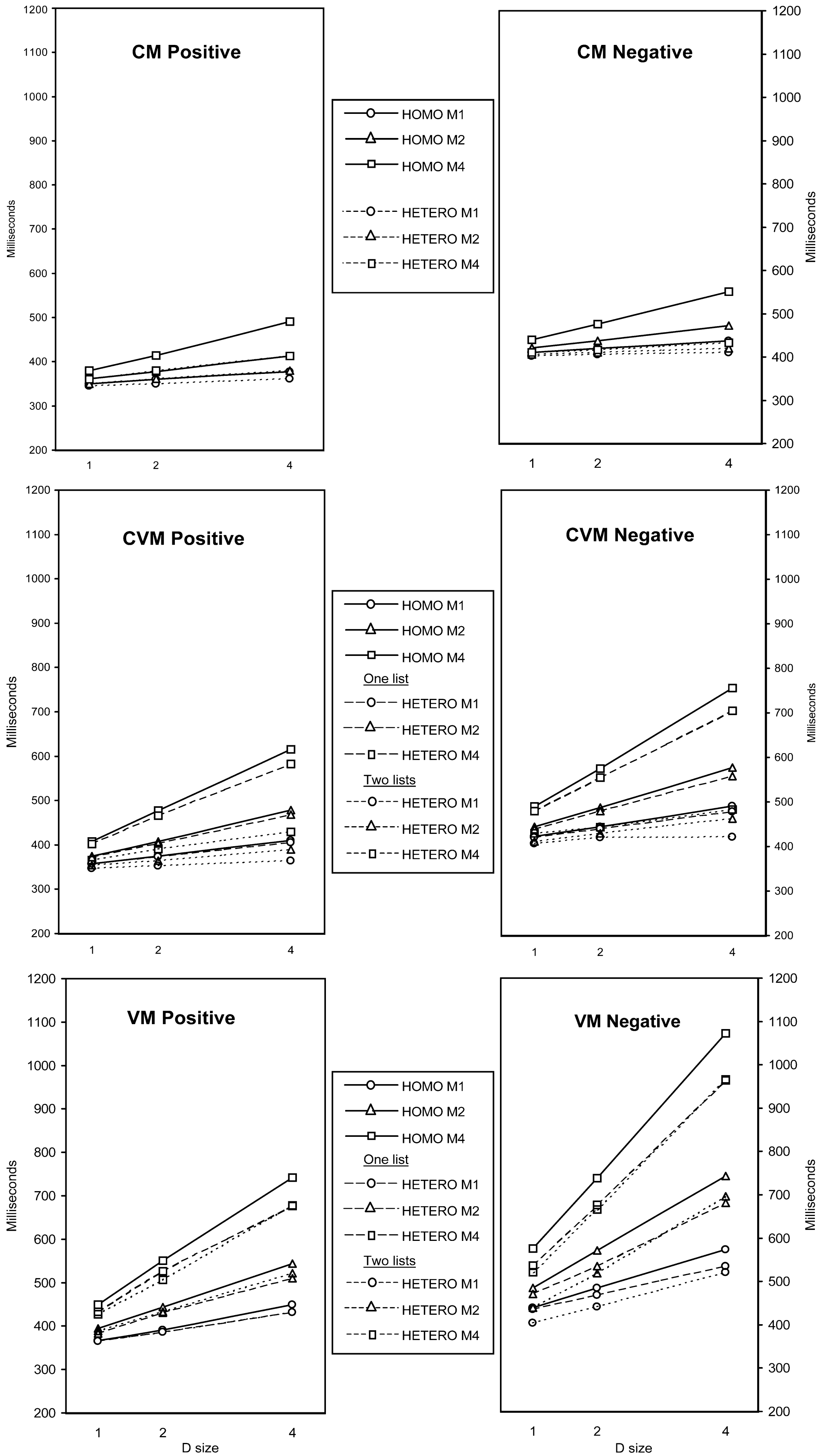


Fig. 7 Mean simulated response times obtained in the 12th block of training under the six original Mapping × Stimulus conditions of the experiment. Results are presented separately for positive (*left panels*) and negative (*right panels*) trials, for CM (*top*), CVM (*middle*), and VM (*bottom*) mapping, as a function of stimulus composition (HOMO vs. HETERO), memory (*M*), and display (*D*) size. For CVM-HETERO and VM-HETERO, predictions of the model are obtained using either two lists of features or one list of features

Varied mapping The bottom part of Fig. 7 shows SFM's performance on the 12th training block in VM conditions. The results obtained in the VM-HOMO condition exhibit the largest multiplicative load effects of all HOMO conditions. Part of the reason for this poor performance is that, in the absence of a set distinction separating targets and distractors, each stimulus has to be discriminated from seven others. Each such discrimination generally requires a larger number of features than in CM or CVM conditions where fewer stimuli serve as distractors. But, even if a subset of discriminating features can be found for a given target, the same set of features will also belong to a distractor when the same stimulus serves as distractor in another trial. As a result, the threshold cannot be raised much with practice ($B = 8.7$).

The dashed lines in Fig. 7 show SFM's performance in the VM-HETERO condition when only one feature list was used. As shown, the load effects are slightly smaller than those produced by SFM in the VM-HOMO condition, as could be expected from the analogous difference obtained in CVM conditions. To determine what would happen with two feature lists (while changing the learning process used in other conditions as little as possible), we modified SFM so that it invoked and updated the list of the category to which the majority of stimuli in the memory set belonged. Choice between the two lists was random when there was an equal number of digits and letters in the memory set. Dotted lines show the results. As can be seen, the use of one vs. two separate feature lists does not make much difference after extensive training under VM conditions.

In summary, a comparison of Fig. 7 with the bottom parts of Figs. 1, 2, 3, and 4 shows that the model provides a good qualitative fit to the mean RTs obtained in various Mapping × Stimulus conditions, except for the CVM-HOMO-O condition. The quality of fit also includes the fact that, whenever $M \times D$ is equal within a given Mapping × Stimulus × Response condition, the model tends to produce longer mean RTs when M size is larger than D size, as is the case for real participants. Table 4 gives an indication of the quality of the quantitative fit across conditions, as computed by a regression of the MNs obtained in the different $M \times D \times R$ conditions against those produced by the model. From now on, when talking about the model's predictions, we will refer to the results obtained with SFM, using one feature list for the HOMO conditions and two feature lists for the HETERO conditions, thereby limiting the presentation to SFM's best profile. For positive trials, SFM (with the appropriate number of lists) predicts between 84 and 98% of the variance in the participants' means, depending on the groups, and the slopes of the regressions are all close to 1.0, again with the sole exception of the CVM-HOMO-O group. The same holds for the means obtained in negative trials, except for the CM-HETERO group where predictions are not as good, but the data have an amplitude (difference between M1D1 and M4D4) smaller than one standard deviation. If, as we have argued, the same mechanism underlies performance in all conditions, the mean RTs should all fall on the same line. A regression performed on positive trial MNs of all groups (except group CVM-HOMO-O) accounted for 95% of the variance with a slope of .98. A similar regression performed on negative trial MNs accounted for 98% of the variance with a slope of .97.

Table 4 Fit of the simulation in the 12th block of training averaged over simulated participants. r^2 correlation between RT and simulated RT

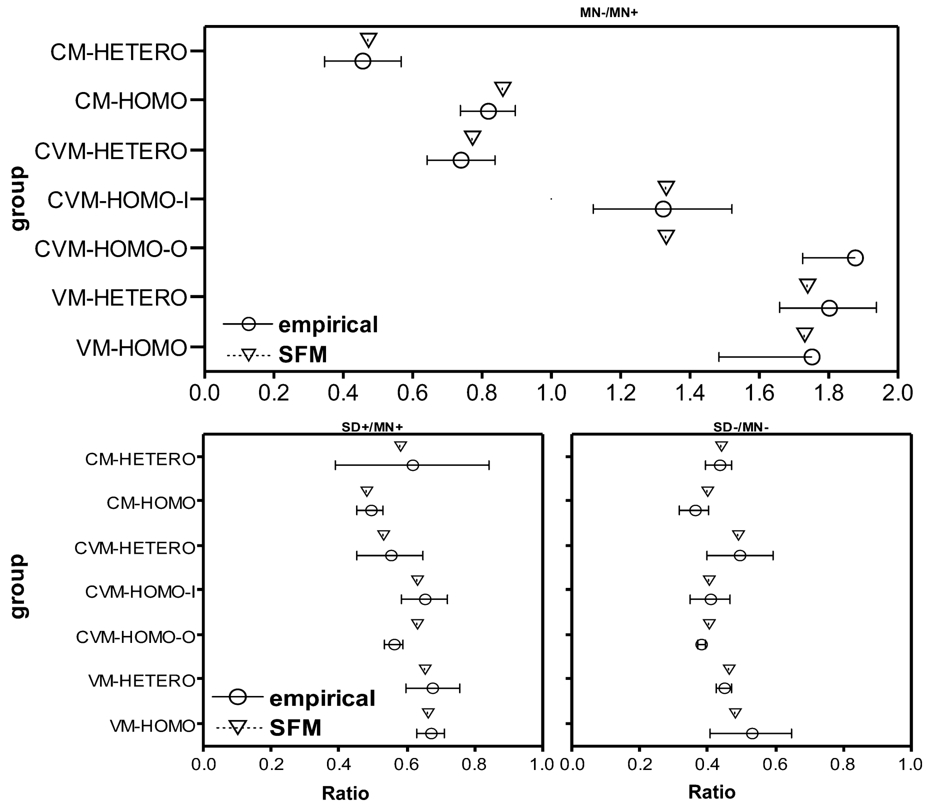
Groups	r^2		Slopes	
	Positive	Negative	Positive	Negative
CM-HOMO	.84	.92	1.08	1.19
CM-HETERO	.95	.61	.97	.69
CVM-HOMO-I	.98	.98	.98	.93
CVM-HOMO-O	.97	.98	.75	.48
CVM-HETERO (one list)	.80	.78	1.97	2.58
CVM-HETERO (two lists)	.84	.88	.92	.89
VM-HOMO	.98	.99	1.05	.94
VM-HETERO (one list)	.97	.99	1.00	.90
VM-HETERO (two lists)	.97	.99	.98	.97

Turning now to the ratio of load effects on the means of negative vs. positive trials. In the top part of Fig. 8, we reproduce the mean slopes reported in the top part of Table 2, along with the standard error intervals inferred from variability among participants within each group. The arrowheads show SFM's best predictions (i.e., using one list for the HOMO conditions and two lists for the HETERO conditions). With the exception of the CVM-HOMO-O group, SFM's predictions are well within one standard error and quite close to the participants' means. Ratios as low as .50:1 and as high as 1.80:1 are produced by the model. These ratios result mainly from differences in the probability of exhaustive processing, ranging from .38 to .72 over conditions, which varies with the number of diagnostic features. The small display load effects stemming from the psychophysical phase also contribute a little to the ratios produced by SFM. By being constant over Response, the psychophysical stage reduces the differences in load effects between the positive and negative trials, thereby reducing the $\Delta MN-:\Delta MN+$ ratios.

Standard deviations

The principles underlying SSTS* models that are embodied in SFM guarantees that the load effects on positive trial SDs will be linearly related to the load effect on MNs. So, SFM guarantees that the pattern of load effects on SD+ will have the same shape as that on MN+, within every Mapping × Stimulus type condition. The rationale is the same for negative trials. The termination rule adopted guarantees that, for a given $M \times D$ condition, SFM will perform half as many comparisons, on average, when processing stops early as

Fig. 8 Plots of the three ratios ΔMN−:ΔMN+ (*top*), ΔSD+:ΔMN+ (*bottom left*), and ΔSD−:ΔMN− (*bottom right*) obtained in the 12th block of practice under the six original Mapping × Stimulus conditions of the experiment and the CVM-HOMO-I replication. Error bars are standard error intervals. *SFM* sufficient features model



when it is exhaustive. Since the probability of exhaustive processing within a given Mapping × Stimulus condition does not vary with load, the contribution of exhaustive trials to the MNs and SDs is also proportional to load, causing the former to increase and the latter to decrease with respect to early termination trials. So, the pattern of load effects on SD− should have the same shape as that on MN−, within every Mapping × Stimulus condition. The question is whether SFM can reproduce the ratios of load effects on SDs vs. MNs that were obtained. This is not guaranteed even for the .60 value derived from SSTS* models for positive trials. Although the above-threshold features are selected randomly by the comparison process in SFM, as also assumed by SSTS* models, the learning processes introduce some biases in the number of comparisons performed, thereby affecting variability.

The bottom part of Fig. 8 reproduces the mean slopes of regressions of SDs on MNs (also found in Table 2), along with the standard error intervals inferred from variability across participants within each group. Arrowheads in Fig. 8 represent SFM's predictions. It is interesting to note that for the CM-HOMO condition, this interval fails to include the .60 value predicted by item-based SSTS* models for positive trials. SFM can nonetheless account for the ratios obtained. Except for the CVM-HOMO-O condition, SFM's predictions also fall within one standard error of the ratios obtained in negative trials. We would not have achieved this quality of fit had we used a single *p* parameter for all groups. It is because *p* is related to the number of diagnostic features in each Mapping × Stimulus type condition that SFM can provide an integrated account of the performance of each group in negative as well as positive trials. Note that, although SFM's termination rule guarantees that the ratios of load effects on SD− vs. MN− will never exceed .60 by much (which requires that search always be self-terminating in negative trials), it does not guarantee that the load effects on SD− will always exceed those on SD+. Load effects on SD− decrease with an increase in the probability of exhaustive processing, which is in turn related to the number of diagnostic features. So, the less discriminable the targets are, the more likely it is that load effects on SD+ will come to exceed those on SD−, as indeed happened in Schneider and Shiffrin's VM-HOMO condition (1977). Because Schneider and Shiffrin's experiment involved a larger number of stimuli, the number of discriminative features may have been smaller, thereby increasing the likelihood of exhaustive search in negative trials.

Figure 9 shows the evolution of load effects on MNs and SDs throughout training for the CM-HOMO group. More specifically, Fig. 9 presents the mean slopes of the regressions of the SDs on the MNs obtained in both positive and negative trials in each of the first 12 blocks of the experiment, along with the standard error intervals inferred from variability across participants within the group. We have chosen to focus on the CM-HOMO group because the cross-sectional analyses reported earlier showed the performance of this group to differ most in the first and twelfth block of training. Remember that early performance in the CM-HOMO

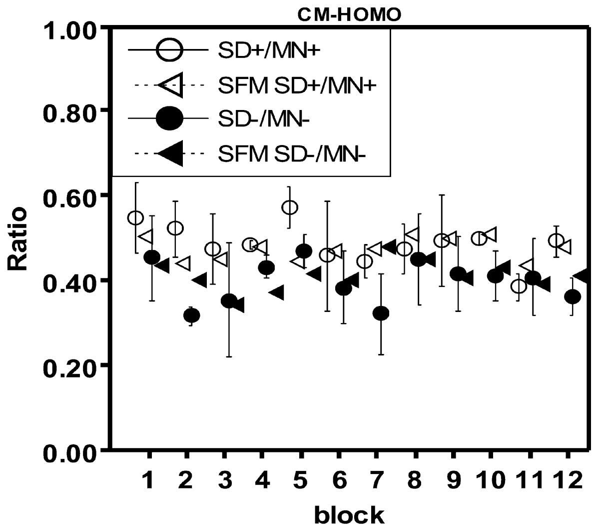


Fig. 9 Plots of the ratios ΔSD+:ΔMN+ and ΔSD−:ΔMN− obtained during the first 12 blocks of practice under the CM-HOMO condition. Results are presented separately for positive (*open circles*) and negative (*filled circles*) trials. Error bars are standard error intervals. Predictions of the models are shown for ΔSD+:ΔMN+ (*open triangles*) and ΔSD−:ΔMN− (*filled triangles*)

condition was characterized by large multiplicative effects of Memory and Display size, characteristic of controlled processing. By contrast, on the 12th block, these effects were additive, the memory and display search slopes being relatively small, though still greater than zero. Although we could argue that performance was still not fully automatic, the performance of this group is the one that conforms best to the predictions of theories that postulate a qualitative change in processing during training under consistent mapping conditions. If performance was indeed controlled by different processes early and late in training, then we might expect these processes not only to exhibit different load effects on means, but to also have different variability under different load conditions, so that the ratio of load effects on MNs and SDs might change between the first and last training sessions. On the contrary, Fig. 9 shows that the ratios of load effects on MNs and SDs remain fairly stable throughout training. This result suggests that the same processes underlies performance throughout training in the CM-HOMO condition, a conclusion that at least agrees with, if it does not reinforce, our earlier proposal that the same processes underlie performance in every condition of the experiment.

SFM's predictions concerning the CM-HOMO group are represented by arrowheads in Fig. 9. These predictions were derived using 72 trial blocks, each target being presented exactly once per block in each $M \times D$ condition to the simulation. The reason for using smaller blocks than in the experiment is that the simulated participants learn faster than the real ones, not being affected by momentary lapses in attention, motivation, or memory. As seen, SFM accounts fairly well for the ratio of load effects on search time MNs and SDs in the CM-HOMO condition throughout training, SFM's predictions being within one standard deviation of the observed ratios inn most of the 12 blocks of training.

General discussion

Information reduction

Three main conclusions emerge from the present study. The first concerns the feasibility of information reduction as an explanatory alternative to dual-process views of automatization in visual-memory search tasks. In the model developed here, search is guided by the features that are most susceptible to revealing the presence of a target on the display. Information reduction is achieved by ignoring features that are not deemed diagnostic enough. In the absence of a pre-existing categorical distinction among the stimuli, the number of feature tests performed by the model is determined by the relative discriminability of targets and distractors. When every stimulus has to be discriminated from all others, as in VM-HOMO conditions, the number of feature tests required to locate a target is large and heavily dependent on memory and display size. The set composition of the stimuli that characterizes CVM-HOMO conditions allows a reduction in the number of feature tests performed since each target has to be discriminated from a smaller number of distractors. The number of feature tests decreases even further when a smaller number of stimuli consistently serve as targets, as in CM-HOMO conditions, resulting in even smaller load effects.

Fisher, Duffy, Young, and Pollatsek (1988) speculated that performance in varied mapping conditions could eventually equal that observed in consistent mapping conditions, if amount of practice and target-distractor discriminability could be kept constant. As previously mentioned, target-distractor discriminability was the same in our CM and CVM conditions. Total amount of practice (and amount of practice per target, see Footnote 4) was also equated. Nonetheless, the experimental results obtained in the CVM-HOMO condition exhibited large multiplicative effects of memory and display size, even when the participants were aware of the set composition of the stimuli, whereas the results obtained in the CM-HOMO condition exhibited small additive memory and display load effects. Although SFM is based on target-distractor discriminability, it succeeded in accounting for this diverging pattern of load effects because it computes discriminability over all potential targets and because it assumes that information reduction occurs at a representational level that encompasses all stimuli. A larger number of feature tests is therefore needed to discriminate the targets in the CVM-HOMO condition because the same stimuli also served as distractors.

The number of diagnostic features is not the only factor that contributes to SFM's overall performance but, if the stop criterion (which was correlated with the number of diagnostic features) is excluded, it is the only factor that varies across mapping conditions. So, the model's performance in the HOMO conditions shows that it is possible to account for widely differing patterns of load effects on mean RTs on the sole basis of the number of feature tests performed. This suggests, in turn, that the amount of information reduction achieved in the various mapping conditions is the main determinant of the efficiency level achieved after training. If this is true, then the existence of fundamentally different processes to account for CM vs. VM performance no longer needs to be postulated.

It could be argued that the model's success in accounting for CM performance is critically dependent on the small number of stimuli used as targets and as distractors. Increasing the number of stimuli would indeed reduce the likelihood that diagnostic features exist. Remember, however, that the featural representation used by the model was very impoverished, being composed only of straight-line segments. So, the possible weakness of the model's current featural representation in no way invalidates the feasibility of the underlying information reduction principle. What this principle predicts is that there should be no improvement in search when all features are evenly shared by targets and distractors (although intercept processes could evidently improve).

The information reduction principle embodied in the model is similar to that proposed by Haider and Frensch (1996, 1999) to account for performance observed in an alphabet verification task. The task investigated by Haider and Frensch probably lends itself more easily to conscious strategic choices than the visual-memory search task. Indeed, the effects reported by Haider and Frensch were found to be more pronounced for participants who reported being aware of the irrelevance of some parts of the stimuli. Some results reported by Compton and Logan (1991) show that the same may be true of the alphabet-arithmetic task (and possibly also the dot-counting task) used by Logan (1988, see also Lassaline & Logan, 1993). In SFM, information reduction occurs at a level that probably lies outside the scope of conscious strategic control, namely the featural level. So the information reduction principle put forth by Haider and Frensch, may not result only from changes in consciously penetrable cognitive strategies. Our results and model suggest that it may also underlie perceptual learning.

If perceptual learning occurs in visual-memory search tasks, some of the knowledge acquired during training could transfer to other perceptual tasks. Treisman, Vieira, and Hayes (1992) found little evidence of such a transfer. They concluded that training under consistent mapping conditions did not cause the formation or use of new pre-attentive feature detectors, sensitive either to local aspects of the target, such as conjunctions of line segments, or global aspects such as figure shape. They argued further that whatever changes during training is specific to the stimuli and task used. SFM is compatible with both conclusions. It can account for the improvement in performance produced by training without creating any new features. The amount of transfer observed would depend on the usefulness of the diagnostic features isolated during the search task. It is not clear that the required level of control over the featural composition of the stimuli was present in Treisman et al.'s experiments. The issue of transfer deserves more investigation.

There is a second sense in which information reduction is present in our model. The global list of diagnostic features that the model constantly updates can be viewed as grouping and summarizing information about all the targets (and distractors) encountered. Information reduction by means of grouping is present in the literature on automaticity under the name of chunking (Miller, 1956), and it has been proposed (Newell & Rosenbloom, 1981) as the main mechanism underlying training effects in many tasks. In the case of visual-memory search tasks, we could imagine that search is initially guided by the representation of individual targets and that these representations are gradually merged into more global ones. By contrast, SFM relied on summary representations right from the start of the experiment. Since the stimuli used, and the categories to which they belong, were all over learned, the most important effects of chunking could well have occurred prior to the start of the experiment. A study involving artificial stimuli and/or unfamiliar categories might be more likely to reveal the effects of chunking in progress, so to speak.

Categorical knowledge

The second major conclusion stemming from this study concerns the influence of pre-existing categorical knowledge on search. CVM-HETERO performance failed to exhibit the large multiplicative pattern of memory and display size effects on means, which characterized asymptotic performance in the CVM-HOMO condition. The load effects obtained in the CVM-HETERO condition were even smaller than those obtained in the CM-HOMO condition. Mapping the category and set distinctions consistently onto decisions or responses, as in the CM-HETERO condition, added little further improvement compared with CVM-HETERO performance.

SFM was totally unable to account for the performance obtained in the CVM-HETERO condition when a single feature list was used for all targets. More specifically, the model greatly underestimated the difference in load effects obtained in the CVM-HETERO vs. CVM-HOMO conditions, despite the fact that there were more featural differences between categories than within category in the character representation used. So, the failure of the single-list model suggests that there was more to the categorical effects observed than a mere difference in discriminability between letters and digits.

The problem was easily remedied by postulating separate feature lists for digits and letters in HETERO conditions, which suggests an interpretation of the feature lists in terms of category representations. Such an interpretation is not ad hoc or farfetched since we already had to assume that information reduction occurred at the level of a single summary representation in the HOMO conditions, i.e., when stimuli were composed of only digits or only letters.

Jonides and Gleitman (1972) reported that the same target could be detected more efficiently among letters than among digits when it was conceived as a zero, the results being opposite when the ambiguous target was conceived as the letter O. This phenomenon, known as the 'oh-zero' phenomenon, was interpreted as indicating that categorical effects on search are not reducible to a mere matter of digit-letter discriminability, since the same ambiguous target served in within- and between-category search. However, both the phenomenon and its interpretation have been contested. The phenomenon was difficult to replicate (e.g., Duncan, 1983; Cardosi, 1986) and many authors have argued that category effects are entirely due to differences in discriminability. The model proposed here constitutes a halfway position. The model is clearly endowed with separate categorical representations for digits and letters. So, it could promote different features when searching for the letter O vs. the digit 0. However, the presence or absence of the 'oh-zero' phenomenon would still depend on the discriminability of the target, as conceived by the participants, and the distractors used.

The two-lists representations in our model may be viewed as implying too much on-line adaptation of the search processes, since the list that is invoked in HETERO conditions can change from trial to trial as a function of the items in the memory set. In fact, our results suggest that participants cannot fully adapt to the context provided by the memory set in every successive trial. Had they been able to do so, their performance would have been identical in CM-HOMO and CVM-HOMO conditions, since the composition of the memory sets was identical in both conditions over participants. Apart from random variations, the same was also true of the display sets. The fact that performance was not identical in these conditions suggests that participants could not optimize search as a function of the specific items in the memory set. Likewise, the model does not order the feature tests performed in a given trial so as to start with the diagnostic features that are most common among the memory set items. Rather, search is driven by general representations that are modified mainly through long-term exposure to the stimuli. So, it is not the case that the model's performance is entirely determined by the momentary context provided by the memory set. By virtue of being featural, the categorical representations that drive search can be conceived as perceptual in nature. So, it is not the case either that the model involves a large amount of top-down, cognitive-to-perceptual influence (see Pylyshyn, 1999).

Serial processing

The third conclusion to emerge from this work is that performance in all experimental conditions investigated here probably involved some seriality in processing. In the model, serial processing occurs when the features of the display items are compared with those in memory and when the memory items are identified. The number of features successively considered by the comparison process is the main determinant of the pattern of load effects on the reaction time means and on their standard deviations. By assuming that the individual comparison times have negligible variability (an assumption that is much more likely to be true when the comparison process operates on elementary features than on complex characters), SFM was able to account for the ratios of load effects on positive trial MNs and SDs, which were relatively constant across conditions in the experiment (hovering around the predicted .60 value).

In order to account for the smaller but also relatively constant ratios of load effects on negative trial MNs and SDs, which hovered around .40, we had to postulate that search sometimes stops before all the required feature comparisons are performed. The same mixture of self-terminating and exhaustive processing also allowed the ratios of load effects on negative vs. positive trial MNs, which were very variable across conditions in the experiment, to be accounted for. Similar variability is also found across experiments involving conjunctive search (see Wolfe, 1998). SFM suggests that this variability is linked to the amount of processing required to discriminate targets from distractors, the probability of exhaustive processing being larger when there are few diagnostic features than when there are many. Although the mixture of self-terminating and exhaustive processing greatly affects the number of feature tests performed by SFM, it affects the MNs and the SDs in a similar way in every condition, which explains why the ratios of load effects on negative trial MNs and SDs remained fairly stable across conditions.

The relative constancy of the ratios of load effects on positive and negative trial MNs and SDs, observed in all conditions of the experiment, poses a difficulty for dual-process theories. By postulating fundamentally different processing, such theories can easily account for the different patterns of load effects on MNs obtained in CM and VM conditions. However, it would be quite a coincidence if such qualitatively different processes produced exactly proportional effects on SDs. So, the results obtained strongly suggest that the same search process underlies both automatic and controlled performance, in contrast to Schneider and Shiffrin's (1977) claim. This conclusion is reinforced by the ratios of load effects on MNs and SDs obtained in the condition that showed the largest improvement in performance with training, namely the CM-HOMO condition. Early performance in this condition exhibited large multiplicative load effects on MNs whereas small additive effects characterized late performance, in agreement with predictions made by dual-process theories based on consistent stimulus-decision mapping. If fundamentally different processes were indeed responsible for early and late performance, it would be surprising that the variability in search times would have the exact same proportionality at the beginning and at the end of the training. The fact that this was not the case suggests that the same processes were involved throughout training.

The validity of the preceding argument rests on the premise that constant ratios of load effects on MNs and SDs imply identity of processing. The argument would have little strength if the ratios of load effects on MNs and SDs were always found to be constant, irrespective of task, condition or training. There is evidence that this is not the case. Rickard (1997) investigated the alphabet arithmetic task used by Logan (1988). This task allows two different strategies, one of which is algorithm-based and the other, memory-based. Automatization of performance in the alphabet arithmetic task therefore appears to involve a shift from one type of processing to another. Consequently, the ratio of MNs to SDs should not be constant throughout training. Rickard did not analyze such ratios, focusing instead on the rate of decrease of the MNs and SDs with practice. Fortunately, the ratio of MNs and SDs remains constant when both measures decrease at the same rate, irrespective of the exact shape of the learning curves (Heathcote, Brown, & Mewhort, 2000), and it does not when the MNs and SDs decrease at different rates. Rickard found that the MNs failed to decrease at the same rate as the SDs (see also Delaney, Reder, Staszewski, & Ritter, 1998; Palmeri, 1999), which guarantees that the ratio of MNs to SDs was not constant. So, Rickard's results add weight to the premise underlying our argument, as well as Logan's (1988) predictions, that constant ratios imply singularity of processing.

Concluding remarks

Although the work presented here synthesizes close to a quarter of a million response times, collected in six different experimental conditions, we readily see that it has important limitations. First of all, the experiment involved only one task, namely the single-frame task of the venerable visual-memory search paradigm. So, the single-process view of automatization proposed here is not intended to apply to all experimental paradigms in the literature on automaticity, as the preceding discussion shows. Secondly, we have been concerned with only one measure of load effects in the task, namely the effects of memory and display size on response times. This study was partly motivated by the fact that, despite extensive research involving the visual-memory search paradigm, little was known about the extent of load effects in a fairly critical condition, namely CVM-HETERO. The fact that these effects were barely more pronounced than in the CM-HETERO condition led us to attribute much more importance to categorical distinctions among the stimuli than to mapping. However, it could be argued that the differences between these two conditions are not visible because the task had become too easy with practice. However, differences would emerge if participants had to perform a concurrent task. Although possible, we do not believe that this is very likely.

The model proposed also has limitations. Its grain is very coarse. The features on the display could be represented by activation or location maps similar to those found in contemporary models of visual attention (e.g., Bundesen, 1990; Logan, 1996; Treisman & Sato, 1990; Wolfe, 1994). Only local features are used by the model to guide search, but global features having to do with the overall shape of the characters may also be involved (see Kyllingsbaeck et al., 2001). The reordering of features may stand for some deeper process having to do with the strength of the feature-to-category associations or with the number of feature-to-category associations. In addition, the feature lists are certainly only a shorthand notation for true categorical representations. The probabilistic nature of the termination rule also hides some ignorance about the factors that cause search to terminate in any given trial. The various processes in the model are also much too modular. On the input side, the order of feature search at display onset could be partially determined by the rate at which the psychophysical stage makes the various features available, which is not the case in the current implementation of the model. On the output side, the termination rule was programmed independently from the diagnosticity threshold, but the simulations show these two mechanisms to be closely related. Other assumptions, such as the postulate that switching from item to item takes no time, probably have little more than a simplifying value. Finally, we certainly make no claims concerning the neurological plausibility of the model. The model provides a strictly functional and schematic characterization of the processing that occurs in visual-memory search.

Despite all these limitations, the model proposed could account for the pattern of load effects on the mean response times obtained in all realizable mapping conditions of the single-frame, visual-memory search task. It also accounted for the relationship between means and standard deviations in just about all experimental conditions, not only at asymptote but throughout training in the condition most affected by practice. In short, the model provides the most integrative view of performance in the task to date. The model is not only integrative at the empirical level. It also integrates many existing theoretical notions in the field of search and automaticity. It integrates the parallelism of disjunctive feature search with the seriality of conjunctive item search (Treisman & Gelade, 1980). It allows the diagnosticity of feature search (Bundesen, 1990; Fisher, 1986) to serve an information-reduction purpose (Haider & Frensch, 1996, 1999). It shows how discriminability effects (Duncan, 1983; Krueger, 1984) can originate at the level of categorical representations (Jonides and Gleitman, 1972). The role of categorical representations in the model is compatible with Logan's (1990) suggestion that it is the mapping of the stimuli to their interpretation that matters. Finally, the model shows that the termination rule used in positive and negative trials may not be as different as the dichotomy between self-terminating and exhaustive processing has led us to believe, in agreement with Wolfe's suggestion (1998; Chun & Wolfe's, 1996). So, although the model constitutes a radical departure for traditional dual-process views, we think that it provides an account of the automatization of visual-memory search that is not only feasible, but also very plausible.

Acknowledgements This research is dedicated to the memory of two of our professors: Y. Dagenais and M. Strobel. The research presented here involved intense collaborative work, which started when the first author was a graduate student under the supervision of the second author. Both authors want to express their gratitude to J. Carignan and A. Archambault for their assistance with the participants, and to M. Lassonde, I. Peretz, and P. Schyns for lending access to their computing facilities. Finally, this research and/or the resulting report benefited enormously from comments by M. Arguin, Y. Lacouture, G. Logan, G. Lacroix, C. Lefebvre, D. Fisher, J. Neely, and J. Tzelgov.

References

Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review*, *89*, 599–607.

Briggs, G. E., & Johnsen, A. M. (1973). On the nature of central processes in choice reactions. *Memory & Cognition*, *1*, 91–100.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523–547.

Cardosi, K. M. (1986). Some determining factors of the alphanumeric category effect. *Perception and Psychophysics*, *40*, 317–330.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternate accounts of skill acquisition. *Psychological Review*, *92*, 414–423.

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, *30*, 39–78.

Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory & Cognition*, *19*, 151–158.

Cousineau, D., & Shiffrin, R. M. (2003). Termination of a difficult visual search. *Vision Research* (manuscript submitted for publication).

Czerwinski, M. Lightfoot, N., & Shiffrin, R. M. (1992). Automatization and training in visual search. *American Journal of Psychology*, *105*, 271–315.

Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science*, *9*, 1–7.

Dumais, S. T. (1979). *Perceptual Learning in Automatic Detection: Processes and Mechanisms*. Doctoral thesis, Indiana University.

Duncan, J. (1983). Category effects in visual search: A failure to replicate the "oh-zero" phenomenon. *Perception and Psychophysics*, *34*, 221–232.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.

Fisher, D. L. (1984). Central capacity limits in consistent mapping, visual search tasks: Four channels or more? *Cognitive Psychology*, *16*, 449–484.

Fisher, D. L. (1986). *Hierarchical Models of Visual Search: Serial and Parallel Processing*. Cambridge, MA: Society for Mathematical Psychology.

Fisher, D. L., Duffy, S. A., Young, C., & Pollatsek, A. (1988). Understanding the central processing limit in consistent-mapping visual search tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 253–266.

Fisk, A. D., & Schneider, W. (1984). Consistent attending versus consistent responding in visual search: Task versus component consistency in automatic processing development. *Bulletin of the Psychonomic Society*, *22*, 330–332.

Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, *30*, 304–337.

Haider, H., & Frensch, P. A. (1999). Eye-movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 172–190.

Heathcote, A., Brown, S. & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*, 185–207.

Hintzman, D. L. (1986) "Schema abstraction" in a multiple-trace memory model. Psychological Review, 93, 411–428.

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.

Hopkins, G. W., & Kristofferson, A. B. (1980). Ultrastable stimulus-reponse latencies: Acquisition and stimulus control. *Perception and Psychophysics*, *27*, 241–250.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*, 575–577

Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: Effects with single- and combined-feature targets. *Journal of Experimental Psychology: General*, *118*, 258–279.

Jonides, J., & Gleitman, H. (1972). A conceptual category effect in visual search: o as letter or as digit. *Perception and Psychophysics*, *12*, 457–460.

Kramer, A. F., Strayer, D. L., Buckley, J. (1990). Development and transfer of automatic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 505–522.

Krueger, L. E. (1984). The category effect in visual search depends on physical rather than conceptual differences. *Perception and Psychophysics*, *35*, 558–564.

Kyllingsbaeck, S., Schneider, W. X., Bundesen, C. (2001). Automatic attraction of attention to former targets in visual displays of letters. *Perception and Psychophysics*, *63*, 85–98.

Lassaline, M. E., & Logan, G. D. (1993). Memory-based automaticity in the discrimination of visual numerosity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*, 561–581.

Logan, G. D. (1978). Attention in character-classification tasks: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General*, *107*, 32–63.

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

Logan, G. D. (1990). Repetition priming and automaticity: common underlying mechanisms? *Cognitive Psychology*, *22*, 1–35.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 883–914.

Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, *103*, 603–649.

Logan, G. D., & Stadler, M. A. (1991). Mechanisms of performance improvement in consistent mapping memory search: Automaticity or strategy shift? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 478–496.

McCarley, J. S., Wang, R. F., Kramer, A. F., Irwing, D. E., & Peterson, M. S. (2003). How much memory does oculomotor search have? *Psychological Science*, *14*, 422–426.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. I. An account of basic findings. *Psychological Review*, *88*, 375–407.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Miller, J. (1988). A warning about median reaction time. Journal of Experimental Psychology: Human Perception and Performance, 14, 539–543.

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

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.

Palmeri, T. J. (1999). Theories of automaticity and the power law of practice. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 543–551.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*, 341–423.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, *126*, 288–311.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception. II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*, 60–94.

Schneider, W., & Fisk, A. D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. *Perception and Psychophysics*, *31*, 160–168.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing. I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.

Schneider, W., & Shiffrin, R. M. (1985). Categorization (restructuring) and automatization: Two separable factors. *Psychological Review*, *92*, 424–428.

Shiffrin, R. M., & Dumais, S. T. (1981). The development of automatism. In J. R. Anderson (Ed.), *Cognitive Skills and Their Acquisition* (pp. 111–140). Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.

Shiffrin, R. M., Dumais, S. T., & Schneider, W. (1980). Characteristics of automatism. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 223–238). Hillsdale, NJ: Erlbaum.

Smith, E. E., & Medin, D. L. (1981) Categories and concepts. Cambridge, MA: Harvard University Press.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652–654.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, UK: Cambridge University Press.

Townsend, J. T., & Colonius, H. (1997). Parallel processing response times and experimental determination of the stopping rule. *Journal of Mathematical Psychology*, *41*, 392–397.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*, 15–48.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 459–478.

Treisman, A., Vieira, A., & Hayes, A. (1992). Automaticity and preattentive processing. *American Journal of Psychology*, *105*, 341–362.

Von Mühlenen, A., Müller, H. J., & Müller, D. (2003). Sit-and-wait strategies in dynamic visual search. *Psychological Science*, *14*, 309–314.

Ward, R., & McClelland, J. L. (1989). Conjunctive search for one and two identical targets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 664–672.

Ward, R., Duncan, J., & Shapiro, K. (1996). The slow time-course of visual attention. *Cognitive Psychology*, *30*, 79–109.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*, 33–39.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419–433.

Zenger, B., & Fahle, M. (1997). Missed targets are more frequent than false alarms: A model for error rates in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1783–1791.