Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Cognitive Psychology 48 (2004) 243–294
Cognitive Psychology
www.elsevier.com/locate/cogpsych

Memory retrieval given two independent cues: Cue selection or parallel access?☆

Timothy C. Rickard* and Daniel Bajic

Department of Psychology, University of California-San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109, USA

Accepted 1 July 2003

Abstract

A basic but unresolved issue in the study of memory retrieval is whether multiple independent cues can be used concurrently (i.e., in parallel) to recall a single, common response. A number of empirical results, as well as potentially applicable theories, suggest that retrieval can proceed in parallel, though Rickard (1997) set forth a model that denies that possibility. In this paper, five quantitative models are developed to test broad candidate principles. In multiple experiments, subjects were trained to retrieve a vocal digit response for each member of a set of letter or color cues. In subsequent test and transfer phases, single cue trials were randomly mixed with dual cue trials on which the two cues always required the same response. For the first few repetitions of each new set of dual cue items, there was no evidence of parallel retrieval over any part of the RT distribution. After more repetitions, dual cue trials were performed faster than single cue trials, but only under conditions that were favorable to development of a "chunked" dual cue representation. These results indicate that associative independence is an important modulating variable that must be heeded in any general model of attention and memory retrieval. Further, the results are most consistent with a model that places the performance bottleneck prior to the retrieval stage of processing.

☆ This research was supported by NIMH Grant No. R29 MH58202 to the first author. We thank Jude Mitchell for programming assistance.

* Corresponding author.
E-mail address: trickard@ucsd.edu (T.C. Rickard).

0010-0285/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2003.07.002

1. Introduction

There is broad agreement in the literature that retrieval from long-term memory involves some type of parallel search. When a cue is presented, all associated pathways are simultaneously activated and compete in some fashion to retrieve a response (for examples, see Nobel & Shiffrin, 2001; Ratcliff, 1978; Ross & Anderson, 1981). It seems a straightforward step to a related principle stating that two independent cues can operate in parallel to retrieve a single, common response. This hypothesis would predict, among other things, that response times (RTs) might be facilitated when two cues are available, provided that they can be perceived in parallel with negligible delay and that their RT distributions overlap.

Consider an example taken from the experiments reported below. In Experiment 1, subjects first learned to make a vocal digit response when presented with each of 12 single letter cues. They then entered a test phase in which the single cue trials were intermixed with dual cue trials (e.g., M L $\rightarrow$ 4). Dual cue trials were always congruent. That is, the answer was always the same as for the component single letter cue trials (e.g., M $\rightarrow$ 4; L $\rightarrow$ 4). Experiments 3–5 had similar designs but the dual cue items were constructed from letter and color cues. According to the parallel facilitation hypothesis outlined above, RTs for dual cue items might be smaller than the RTs for the faster of its component cues when presented individually.

Although no research has directly addressed this hypothesis for the case of memory recall from two independent cues, several lines of work indirectly support it. Consider the coactivation effect observed in go, no-go tasks (e.g., Miller, 1982). Subjects must press a key when a specific stimulus, the target, is presented, but withhold their response when a non-target stimulus is presented. The main effect of interest here is that RTs are faster when two targets are presented than when only one target is presented. Further, Miller and others have shown that the magnitude of the RT facilitation for dual cues often exceeds that expected under a pure separate activation account, in which the two cues "race" independently and in parallel. Miller termed this facilitation effect coactivation, because it implies that the two cues jointly activate a single response node.

Cross-talk effects recently observed in dual task, two-choice RT experiments also suggest that two cues can simultaneously activate a single response. Hommel (1998) showed that RTs for the first task completed are faster if the second task has the same response. This result suggests that the flow of activation to the first response was occurring in parallel from the two task cues. Logan and Schulkind (2000) investigated whether cross-talk effects generalize to semantic memory (two-choice categorization) tasks. Consider Experiment 1 of their study. On each trial subjects were presented with two letters (one above the other), two numbers, a letter and a number, or the spatial reversal of the letter and number condition. For each stimulus, starting with the top one, subjects pressed a button that corresponded to the correct category. Subjects were faster on trials with either two letter or two number cues (i.e., compatible trials). Further, the facilitation effect decreased as the stimulus onset asynchronicity (SOA, the latency between the onset of the first and second task stimulus) increased. The authors interpreted these results as reflecting a cross-talk effect similar to that observed by Hommel (1998); activation was flowing from the second task stimulus to its category representation while subjects were performing the first task. In Experiments 3 and 4 of the same paper, the authors found cross-talk in lexical decision tasks. Logan and Delheimer (2001) found analogous effects for an episodic memory task. Subjects learned a list of words, and recognized a list word faster if both stimuli were from the list.

Superficially at least, the congruent Stroop task provides the closest match to the current tasks. Slight facilitation is occasionally observed for mean RT in the congruent relative to a control condition, but a more robust effect is that the lower portion of the RT distribution is below that of the control condition (Heathcote, Popiel, & Mewhart, 1991; Mewhort, Braun, & Heathcote, 1992; Spieler, Balota, & Faust, 1996, 2000). Eriksen and Eriksen's (1974) flanker task is closely related. In the typical design, subjects make a response to a target that is presented in the middle of the screen. Flanker stimuli, which may have responses that are neutral to, congruent with, or incongruent with the required target response, are presented as well. The general finding is that RTs for target identification are smallest when a congruent flanker is present.

Finally, numerous other studies exploring cued memory access in both animals and humans have shown that performance improves as the number of available cues increases (e.g., Jones, 1976; Rudy, 1974). Jones had subjects study pictures with multiple stimulus dimensions, such as color and location, and then compared recall of items given one or two of these dimensions as cues. He, among others, found higher accuracy when two cues were presented, though he did not report response latencies. However, an increase in accuracy in the context of multiple cues does not necessarily imply parallel response activation. It is quite plausible that when one of the cues could not support retrieval, the other one could. Under those conditions, accuracy should improve even if retrieval through the two cues is attempted sequentially.

2. Potentially applicable theories

In addition to the empirical evidence pointing to the possibility of dual cue facilitation in cued recall, there are several theories that, if applied to our tasks, would appear to predict dual cue facilitation. One straightforward example is modern theories of the Stroop effect (Stroop, 1935; for a review, see MacLeod, 1991), all of which assume parallel processing of color and word dimensions up to the response stage of processing (e.g., Cohen, Dunbar, & McClelland, 1990; Logan, 1980; Phaf, Van Der Heijden, & Hudson, 1990; Schooler, Neumann, Caplan, & Roberts, 1997), resulting in RT facilitation in the congruent condition. If these models are to be viewed as capturing general properties of information processing, extending beyond the Stroop task proper, then they appear to predict dual cue facilitation in our experiments as well (especially in Experiments 3–5, which involve letter and color cues).

Logan's (1988, 1992) instance theory of automatization and Nosofsky and Palmeri's (1997) closely related exemplar based random walk (EBRW) model (see also

Palmeri, 1997) assume that an independent instance is formed on each trial for an item, and that previously encoded instances race to retrieve the response on each trial. These theories might easily accommodate a finding of dual cue facilitation in the current experiments. However, their development to date applies primarily to the case of a single retrieval from a single cue, and as such those models may not make strong a priori predictions for the current tasks. Nevertheless, the current experiments should be of value in guiding any future extension of those models beyond the case of a single cue and a single response.

Wenger (1999) tested a series of memory models that varied along the dimensions of both representation and process. With respect to process, he interpreted his results as supporting a parallel retrieval model. Logan and Gordon (2001) proposed a model that integrates the instance and EBWR models with Bundesen's (1990) visual attention model. They also added a control mechanism that determines, among other things, whether or not two tasks will be performed in parallel. As currently developed, however, that model does not appear to make a priori predictions for the current tasks. It is potentially consistent with a finding of dual cue facilitation, but does not appear to be constrained to make that prediction.

Uniquely, Rickard's (1997, 1999) component power laws (CMPL) model would predict no dual cue facilitation if applied to the current tasks. That model was originally developed to explain the strategy shift to memory-based performance that often occurs with practice on cognitive tasks. At its core, however, is a theory of attention with respect to retrieval from long-term memory. Two assumptions are critical for current purposes. First, learning is assumed to involve formation of a separate representation (node) in long-term memory for each independently acquired conjunction of a specific cue and the general task goal (or set). In this paper we will refer to these nodes as set-cue conjunctions, although the equivalent nodes in the Rickard (1997) simulation were termed "problem level" nodes. For example, if a subject has learned to respond by saying "4" when presented with the letter M, the model assumes that a set-cue conjunction node, representing the conjunction of the general (i.e., non-item-specific) task set, "speak the digit," and the stimulus, M, has been formed. This node, which will be unique for each item in the stimulus set if single cue training takes place on independent trials, is in turn associated with the answer (e.g., "4"). Thus, activation flows from the stimulus and the task set at the first level, to the set-cue conjunction at the second level, and then to the response at the third level.

The second critical assumption is that only one of these set-cue conjunction nodes can be used at any given moment to retrieve a response. As soon as one set-cue node is selected, activation of all other cue-set nodes is suppressed, resulting in a prediction of no facilitation on dual cue trials. If, for example, a dual cue item such as "M L" is presented, only one of the letters, along with its corresponding and unique set-cue node, can be selected for retrieval. Hence, the model proposes a performance bottleneck prior to the retrieval stage of processing, at the set-cue level of representation.

These two assumptions lead to other predictions as well. One prediction, which recently found support in studies by Rickard and Pashler (2003) and Nino and

Rickard (2003), is that two responses cannot be retrieved in parallel from a single cue, provided again that the two cue–response associations are independent. These authors had subjects first memorize a vocal digit response for each of 10 visually presented words. Next, they memorized a left or right key press response for the same 10 visually presented words (in other experiments, equivalent results were found when more than two key press responses were involved). In the final test phase, blocks of the digit and key press tasks were interleaved with dual task blocks, in which both the vocal digit and the key-press responses were performed on each trial. Learning of the single-task associations took place independently for the vocal and key-press tasks, yielding an independent goal, or task set, for each task (i.e., "retrieve the digit" and "retrieve the key-press"). The CMPL theory therefore predicts that two independent set-cue conjunctions were acquired during learning for each stimulus. One of these was the conjunction between the stimulus and the key press goal, and the other was the conjunction between the stimulus and the vocal response goal. It follows from the second assumption above that on dual task trials, retrieval of the two answers must occur sequentially, at least while the two associations continue to be represented independently. The data were consistent with this prediction. Nino and Rickard (2003) showed that the sequential retrieval predictions held even when subjects were given extensive retrieval practice on the single tasks prior to the dual task test phase.

The experiments below explored the reverse issue; two different cues were associated with a single response, under a single task set (i.e., "retrieve the digit"). Whereas the critical factor causing the processing bottleneck in the Rickard and Pashler (2003) experiments was, according to CMPL, the use of two independent task sets, in the current experiments it will be the presence of two independent cues. In both cases, however, the set-cue level of representation is the hypothesized source of the bottleneck.

The above literature review illustrates two related points. First, the question of whether two memory cues can independently and simultaneously activate a single response is central to current work in several related literatures. Second, there is currently no consensus regarding the answer. To our knowledge, no studies have directly addressed the issue in the domain of cued recall.

3. Quantitative predictions of five candidate models

Our primary goal in this line of work is to distinguish between two major classes of models: cue selection models, according to which only one cue can be used for retrieval on a given trial, and parallel models, according to which retrieval can take place through two or more cues concurrently. Toward that end, we present here two quantitative models from the cue selection class and three from the parallel class. These models embody simple and fundamental principles that are subject to straightforward evaluation. We intentionally developed these models at a relatively abstract level. The candidate principles being evaluated will be central to development of more detailed process models, so it is sensible to evaluate them first. To the extent that a particular principle can be falsified at this more general level, any more specific process models that might embody it are also falsified.

The first three models below allow dual cue RT predictions to be derived parameter-free, directly from the single cue data. Merits and limitations of parameter-free models are considered in Section 9. All five of the models are grounded in the assumption that the cue–response association for one cue of each dual cue item is independent of that for the other. The case of *associative independence* is theoretically crucial. To our knowledge, all potentially applicable parallel models in the literature treat it as a sufficient condition for parallel memory access. This assumption is reasonable when applied to the first dual cue trial for an item, since prior to that point all cue–response associations would have been learned and performed on independent trials. Extended practice on dual cue items, however, may lead to violation of it. There is little doubt that sets of independent elements presented together can be "chunked" into a single memory representation under some conditions (element binding in episodic memory is perhaps the most familiar example). Such chunking may well occur for dual cue items after practice in the current experiments. As such, the models outlined below make their strongest predictions for the first block (defined as one randomly ordered trial for each single and dual cue item) of each new dual cue performance phase, although approximate independence may hold beyond that point.

A secondary goal of the study was to explore whether the model that best predicts performance in the case of known associative independence continues to hold after moderate practice, and if not, to make progress in generalizing that model. Subjects were therefore given at least 20 blocks of dual cue testing over the course of each experiment.

3.1. Random cue selection (the RS model)

Both this model and the next assume three sequential and stochastically independent processing stages: cue perception, cue selection, and retrieval and execution of the response from the selected cue. Subjects must select one cue at the expense of the other, and retrieval takes place only through the selected cue. Once a cue is selected, retrieval takes place exactly as it would if that cue were presented alone. A distinguishing property of the RS model is that subjects have no information at the moment of cue selection regarding which cue is likely to yield the faster response. Rather, cue selection on each dual cue trial is assumed to be random, with each cue having an equal selection probability (.5). In the simplest case, in which the cue selection stage has zero latency, this model predicts that, for each dual cue item of each test block,

$$\mu_{\mathrm{D}} = (\mu_1 + \mu_2)/2 + \mu_{\Delta p}, \tag{1}$$

where μ denotes a population mean RT (including all stages, from stimulus presentation to the motor response), μ_{D} is the population mean for a dual cue item, μ_1 and μ_2 are the population means for two single cues constituting the dual cue item (when presented by themselves), and $\mu_{\Delta p}$ represents any increase in perceptual latency when two cues are presented, relative to perceptual latency for a single cue. Note that this equation constitutes a lower bound prediction for the RS model, since it assumes zero cue selection latency. We will set aside $\mu_{\Delta p}$ for the moment and return to it later.

The RS prediction for the dual cue mean can be derived from the single cue data in the following way. Each test block included one trial for each dual cue item (e.g., M L) and one trial for each of its single cue components (e.g., M and L). These three items constitute a matched item triplet. Thus, on each test block, an expected value prediction for each dual cue item can be generated by computing the mean RT for its component single cue trials (e.g., M and L) on the same block. If the RS model is correct at the level of each of these item triplets, then these dual cue predictions, averaged over all item triplets within each practice block, provide the correct prediction for the dual cue mean on that practice block. For this model alone, this result is algebraically equivalent to simply taking the sample mean of all single cue items and using it to predict the dual cue mean.

The model's predictions are not restricted to the mean. It also predicts that the population RT distribution function governing each dual cue RT on each practice block is an equally weighted mixture of the RT distribution functions for its two component single cue items on the same practice block. Details of modeling the dual cue distributions based on the single cue data will be discussed in Section 4.2. Note also that because cue selection is assumed to occur prior to initiation of answer retrieval, error rates, as well as RT distributions on error trials, are also predicted to be equivalent for single and dual cue items.[1,2]

For the case of dual letter cues (Experiments 1 and 2), there is evidence that the $\mu_{\Delta p}$ term is greater than zero. To generate an unbiased RS prediction, its value needs to be approximated. Fortunately, the literature provides information relevant to this issue. A number of studies have shown that two letters can be perceived in parallel (Egeth & Dagenbach, 1991; Pashler & Badgio, 1985; Shiffrin & Gardner, 1972; van der Heijden, 1975). Thus, if a cue selection model fits the data best, that result cannot be attributed to a perceptual bottleneck. Further, the Pashler and Badgio results, confirmed over multiple experiments, indicate only about a 20 ms delay in perception of two digit cues compared to one. Their task is sufficiently similar to ours to expect that $\mu_{\Delta p}$ will be of roughly the same magnitude here. Nevertheless, to provide direct evidence for the magnitude of this delay, we performed an auxiliary experiment (see Appendix A) that was matched as closely as possible to our letter experiments. Consistent with expectation, the results yielded an estimate for $\mu_{\Delta p}$ of about 20 ms. This 20 ms correction was added to the mean prediction for each of the models for Experiments 1 and 2. The standard deviation for the $\mu_{\Delta p}$ estimate in the experiment described in Appendix A was only a few milliseconds. Thus, as a close approximation of the $\mu_{\Delta p}$ effect in the distribution modeling to be discussed later, a constant

[1] Since the stimulus set is relatively small and the items highly familiar, it is reasonable to assume that response errors based on perceptual miscoding were rare.

[2] Note, however, that the model does not require that the RTs on error and correct trials will have the same distribution. Error RTs might, for example, result from weak and poorly formed cue–response associations.

value of 20 ms was added to each single cue RT that was used to create the predicted distribution.

Two factors should mitigate any remaining concern that the reader may have regarding dual cue perceptual delay. First, in Experiment 1 of the Logan and Schulkind (2000) study, the longest SOA was 900 ms, whereas the first task was probably completed or in the motor stage in less time on most trials. Thus, the 900 ms SOA condition was essentially a single cue condition, making the 0 vs. 900 ms SOA conditions perceptually analogous to the two versus one cue conditions in Experiments 1 and 2 of this paper. The fact that they found dual task facilitation in the 0 relative to the 900 ms conditions for compatible items demonstrates that any perceptual delay that may be caused by the presence of two letter cues is not sufficient to mask facilitation, even in a quickly executed two choice RT task in which the magnitude of potential facilitation is relatively modest. The two versus one letter facilitation effect observed in the go, no-go task also indicates this fact. Those results should also alleviate any concerns regarding our implicit assumption of context independence more generally (i.e., the assumption that subjects do not treat single and dual cue items differently, beyond the effects intended in the manipulation). Single and dual letter items are mixed within blocks in both the cross-talk and go, no-go paradigms, just as in the current experiments, and dual cue facilitation was observed. Second, in Experiments 3–5, cues were colors and letters, and in Experiments 3 and 5 they took the form of colored letters. As discussed in the introduction to Experiment 3, the literature indicates no dual cue perceptual delay at all in those cases. Hence, $\mu_{\Delta p}$ was set to zero in those experiments.

3.2. *Efficient cue selection (the ES model)*

This model is equivalent to the RS model, with the crucial exception that the more efficient cue of each dual cue pair (i.e., the cue yielding faster retrieval *on average* on single cue trials) is always selected on each dual cue trial. This model thus assumes that subjects have perfect information as to which cue is expected to deliver the faster response (i.e., as to which cue is more efficient). For now we will not speculate on the nature of this information, or on the conditions in which this assumption might be psychologically viable, but we will return to this issue later. Again assuming zero cue selection latency, this model's prediction for each dual item on each practice block is:

$$\mu_D = \min(\mu_1, \mu_2) + \mu_{\Delta p}, \tag{2}$$

where $\min(\mu_1, \mu_2)$ refers to the minimum of its component single cue means, μ_1 or μ_2.

The ES prediction forms the lower bound RT prediction for cue selection models as a general class. If dual cue RTs are significantly below this boundary at any point on the RT distribution, and if the two cues can be considered independent, then the entire class of cue selection models can be rejected. The ES model is thus crucial to our long-term goal of discriminating between cue selection and parallel models as general classes.

Note that the RT distribution prediction of the ES model is closely related to the Grice inequality (Grice, Canham, & Gwynne, 1984; Townsend & Nozawa, 1995), which specifies the upper bound RT that is possible for any parallel race model (corresponding to the extreme case of perfect positive dependency between the two retrieval latencies; Colonius, 1990). More specifically, the ES prediction is identical to the upper bound defined by the Grice inequality.

An expected value prediction from the ES model for each dual cue item on each block can be derived from the single cue data using a technique directly analogous to the matched item triplet estimation process that was described for the RS model. In this case, however, only the efficient cue of each pair should be used as the predictor. Of course, the efficient cue of each single cue pair needs to be identified before this selection can be done. We approached this problem by first computing the mean RT for each single cue item, over all test blocks (at least 20 in each experiment). For each dual cue pair, the component single cue with the faster sample mean on single cue trials can then be treated as the ES prediction for the corresponding dual cue item (sans $\mu_{\Delta p}$).

This approach assumes that the sample means for the two cues always have the same magnitude ordering as do their underlying population means. If this assumption is correct for all items, then the cue that is identified as efficient for each cue pair based on the sample data will always be the more efficient cue in the population. However, if this assumption is in error—as it inevitably will be on occasion when only modest sample sizes are available—then this estimate turns out to be too low. Essentially, it capitalizes on chance variations in the sample means, inevitably pushing the estimated prediction below the true ES prediction. A method for adjusting for this bias, applicable at the level of the RT mean and distribution predictions over items and subjects, is described in detail in Appendix B. The analyses in that appendix demonstrate that, to arrive at an unbiased ES prediction in the pooled data, the slower member of a dual cue pair should be eliminated from the ES prediction only when the absolute value of a t test on the sample means of the RT data from the two cues exceeds 1.0. Otherwise, both members of the single cue pair should be included in the computation of the ES prediction for their corresponding dual cue item. In the current experiments, the bias in the mean prediction that would be introduced by not making this correction would be only about 20 ms. Nevertheless, this adjustment was integrated into the derivation of the ES prediction in each experiment to obtain maximum validity. Note that including this adjustment should make it easier to reject the ES model in favor of a parallel model, because it increases the value of the ES prediction.

Once the efficient cue for each dual cue pair was selected, and the other, "inefficient" cue eliminated (when appropriate), the averaging over items and subjects to obtain the ES prediction for the mean on each test block was performed just as it was for the RS model. As was the case for the RS model, the ES predictions extend beyond the mean, to the RT distribution and the error rate.

3.3. Unlimited capacity independent parallel retrieval (the race model)

The race model considered here assumes that both cues can participate independently and in parallel in retrieving the response throughout both their perceptual and retrieval stages. Processing within the retrieval stage is assumed to be capacity unlimited. According to this model, dual cue retrieval can be characterized as a horse race, with the finishing time of each horse, or cue, being the same on average as its finishing time when running alone. The first cue to reach the "finish line" on a dual cue trial determines the RT, the response, and the accuracy on that trial. For discussions of various types of parallel models and their properties, see Colonius and Ellermeier (1997), Colonius and Vorberg (1994), Compton and Logan (1991), Diederich and Colonius (1987), Logan (1988, 1992), Miller (1982), Rohrer, Pashler, and Etchegaray (1998), Schweikert (1983), Townsend and Ashby (1983), Townsend and Colonius (1997), and Townsend and Nozawa (1995). This model's population mean prediction for each observed dual cue RT is

$$\mu_{\mathrm{D}} = \mu_{\min(\mathrm{RT1},\mathrm{RT2})} + \mu_{\Delta p}, \tag{3}$$

where $\min(\mathrm{RT1}, \mathrm{RT2})$ refers to response time for retrieval through the "faster" cue on each dual cue trial.

Assuming that general performance efficiency does not fluctuate much from trial-to-trial within each test block, an expected value race prediction for a given dual cue trial can be obtained by selecting the faster RT from its component single cue trials within the same practice block. For example, consider the dual cue item M L in Experiment 1. By picking the faster RT of the M and L single cue trials within a given block, a random observation from the theoretical RT distribution of the corresponding dual cue item (M L) on that same block is approximated.[3] Note that, unlike the cue selection models, the race model can take advantage of trial-to-trial variability in retrieval latency for each cue on each trial, resulting in RT predictions that are always below that of the ES model, provided only that the RT distributions of the two cues overlap. As was the case for the preceding models, the subset of single cue data that corresponds to the race prediction for the mean also provides predictions for both the dual cue RT distribution and the error rate.

3.4. A limited capacity parallel retrieval model (the LC model)

The hypothesis that dual cue retrieval is parallel but with limited capacity is also viable. Unfortunately, any limited capacity account will require one or more free parameters to describe the nature and extent of the capacity limitation and thus it is not possible to compare it on an equal footing with the models introduced above. Nevertheless, it is informative to consider a simplest case limited capacity account that can be derived as a transformation of the single cue data. Our goal here is not to

[3] In its purest form, this model should include a correction for motor processing latency. Since on dual cue trials a single motor response is executed after the retrieval race is completed, the motor component of the single cue RTs cannot contribute to facilitation resulting from the race. By estimating the race prediction from the single cue data, we implicitly assume that it does. However, simulations in which the motor latency was assumed to be 100–150 ms, with a coefficient of variation of .2, suggest a bias due to this factor of less than 10 ms [for an analogous simulation, see the appendix of Nino and Rickard (2003)]. In light of the data patterns observed in the experiments, this effect size can be ignored for both this model and the next without compromising theoretical inference.

exhaustively test limited capacity models as a general class. Rather, we sought to evaluate the fit quality of one simple implementation and to identify challenges that may remain for future modeling efforts within this framework.

At the core of any model that assumes limited capacity retrieval through separate, non-interactive channels must be a function that transforms the single cue RT distribution into the predicted RT distribution for that cue on dual cue trials. We chose the simplest case function in which each observation from the single cue distribution is multiplied by a coefficient whose value is greater than one. With this coefficient added, the RT distribution has a larger mean and standard deviation than for the same cue when presented alone, but a general shape that is unchanged. Expressed mathematically, the population mean prediction for each observed dual cue RT is:

$$\mu_{\mathrm{D}} = \mu_{[\min(\mathrm{RT1} * c, \mathrm{RT2} * c)]} + \mu_{\Delta p}, \tag{4}$$

where c is the limited capacity coefficient. This model has the same assumptions that were set forth for the race model. This equation differs from the race equation (Eq. (3)) only by the parameter c.[4] As for the other models, the LC model assumes cue independence, and thus its strongest prediction is for the beginning of each test phase. The value of its free parameter, c, was thus optimized for the RT data from the beginning of test.

There are of course other candidate parameters in the general class of limited capacity models, accounting for other possible factors such as trial-to-trial variability in total available capacity, random fluctuations in capacity allocation, strategic capacity allocation, multi-parameter and non-linear RT transformations, and mixtures of self-terminating and exhaustive retrieval trials. There are also other types of parallel models that assume a multiple retrieval random walk process instead of a single retrieval on each trial (e.g., Nosofsky & Palmeri, 1997). We did not attempt to fit these more complex models, though we do discuss several of them later.

3.5. Coactivation

This model adopts the parallel retrieval assumption of the race model, but instead of assuming that each cue activates a separate response token, it assumes that activation from the two cues converges on a common response. It is not possible to derive instructive RT predictions from the single cue data for this model, because

[4] Multiplying single cue RTs by the coefficient c inflates not only the retrieval latency, as desired, but also the perceptual and motor latencies, which would not be expected to increase since the limited capacity adjustment applies only to the retrieval stage of processing. However, any distortion caused by these factors would be small relative to the mean dual cue RTs of about 1200 ms on the first test block. If the perceptual and motor components could be removed, and only the retrieval component of the RT inflated, the primary consequence would be a modest reduction in the distribution variance. Given the consistent pattern of results obtained in all five experiments, and the highly similar shape of the race and LC distribution fits, such distortions could have no effect on the theoretical conclusions drawn. Indeed, given the outcome, removal of those variance components from the LC model would only worsen its fits. Note that the motor bias (Footnote 3) and the bias described in this footnote work against each other, partially canceling their effects for the LC model.

the function that would combine dual cue activation at the output node is not known. However, if dual cue RTs fall significantly below the prediction of the race model, coactivation is implied and analytical techniques are available to determine whether or not any type of separate activation parallel model can account for the data (e.g., Miller, 1982).

4. Experiment 1

Subjects were first trained to a learning criterion on each of 12 letter–digit associations. They then began a test phase in which all single cue items were interleaved with six dual cue items on each of 20 blocks. Letter order for each dual cue item was left–right reversed from block-to-block. Thus, if M L was a dual cue item on block 1, then L M was a dual cue item on block 2.

4.1. Method

4.1.1. Subjects

Twenty University of California at San Diego undergraduate students participated for course credit.

4.1.2. Materials, design, and procedure

The test stimuli consisted of 12 letters, presented either singly or in pairs. Each letter stimulus was 3 mm wide and 5 mm tall, and the distance between letters presented in a pair was 3 mm. Responses consisted of spoken digits, 3–8. Each digit response was mapped to exactly two letter stimuli, and letters presented in a pair were always associated with the same response. The complete stimulus set thus consisted of all 12 single-letter stimuli, along with 6 dual-letter stimuli.

Subjects were tested individually using IBM-compatible personal computers. All experiments were programmed using Micro Experimental Laboratory software (version 2.01). This software, along with the voice key apparatus (model 200A), was purchased from Psychology Software Tools. Each subject was seated about 50 cm from a 35.5 cm color monitor, and approximately 5 cm from a microphone. Subjects were instructed to place their elbows near the edge of the table, with their hands over their arms. This procedure kept the subject at a constant distance from the microphone and monitor. The program was then initiated and the experimenter read aloud the instructions presented on the screen while the subject read along silently.

The first three blocks of items were always *study* blocks. In these blocks, only the single-letter stimuli were presented. On each trial, the subject was simultaneously presented with a single-letter stimulus (e.g., M) the answer (e.g., 8), and instructions to memorize the answer. After 5 s, these instructions were replaced by instructions to make the correct response when ready. After the subject responded, the computer proceeded by presenting the next item. Each letter appeared in exactly two trials per block, randomly ordered. After completing two study blocks, the subjects were asked whether they felt sufficiently comfortable with the task to proceed to the trials in which they would be required to generate the answers from memory. If they responded "no," they were allowed one additional study block.

Before beginning the dual cue test phase, subjects were given a series of single cue performance blocks in which they had to demonstrate a minimum degree of proficiency in recalling the answers. Each block again consisted of 24 randomly ordered trials, with each single letter stimulus presented twice. Every trial proceeded as follows: (1) the message "Get Ready!" appeared in the center of the screen for 400 ms, (2) the screen went blank for 200 ms, (3) an asterisk fixation point appeared in the center of the screen for 400 ms, (4) the screen went blank as before, and (5) the letter stimulus was presented. The letters always appeared directly to the left or right of the location where the fixation asterisk had appeared; each letter appeared in both positions in every block. Upon presentation of each stimulus, subjects were required to speak into the microphone the answer they had earlier memorized, and to do so as quickly as possible while being as accurate as possible. After the subject responded and the voice key tripped, the experimenter used the computer keyboard both to enter the subject's response and to record whether the voice key tripped properly. The computer then provided accuracy feedback and, if the subject was in error, presented the correct response. Each block concluded with a listing of the percent of accurate responses and the mean correct RT. The subjects continued to receive these blocks until accuracy of at least 90% was attained, at which point the learning phase ended.

Before the start of the testing phase, instructions were presented on the screen and simultaneously read aloud by the experimenter. These instructions informed the subject that some trials would consist of single letters, and some trials would consist of letter pairs. It was stated that each pair would consist of two letters that had been associated with the same digit during the learning phase. Thus, subjects understood that both letters always provided convergent information about the correct response. It was also made clear to subjects that they should give only one response, not two, when presented with a letter pair.

Trial events during the testing phase matched those from the training phase. Each block consisted of 18 trials, with one presentation of each of the 12 single-letter items, and one presentation of each of the 6 dual cue items. Single letter cues always appeared to the left or right of the location where the fixation asterisk had appeared. When pairs were presented, the two letters straddled this central location. However, for both single and dual cue items, there was block-to-block reversal in whether a particular letter appeared to the left or right of the fixation point. Thus, if M and L were letters that had been associated with the same number, the dual cue item would be "M L" in some blocks, and "L M" in an equal number of other blocks. There were a total of 20 test blocks. Subjects were permitted to take a brief break at the midpoint of the experiment.

4.2. Results

Fig. 1 shows the test phase mean dual cue RTs (with errors and incorrect voice key trips excluded) as a function of test block, along with the predictions of the

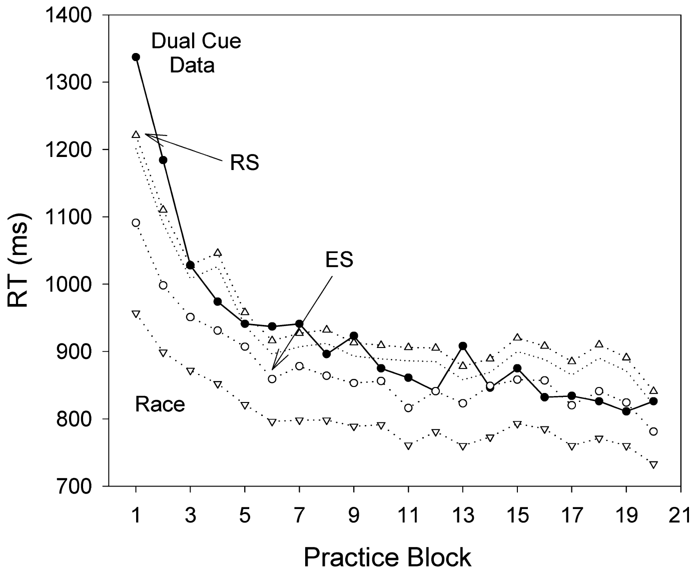

Fig. 1. Mean dual cue RTs for Experiment 1 as a function of test block, plotted against the predictions of the RS, ES, and race models. The dotted line just underneath the RS prediction is the mean of the single cue data, unadjusted for $\mu_{\Delta p}$.

RS, ES, and race models (the grand mean of single cue items, with no $\mu_{\Delta p}$ adjustment, is also shown as a dotted line for reference). Block means were computed first over items for each subject, and then across subjects. The fits of the LC model will be shown only on the RT distribution plots to be discussed later. For reference, however, that model's fit to the dual cue means can always be exact on the first test block, due to its free parameter. In all cases it predicts that dual cue RTs decrease across test blocks at a rate similar to that predicted by the race model.

On the first three blocks, the mean dual cue RTs were above not only the ES prediction but also the RS prediction. They never fell systematically below the ES prediction throughout the 20 test blocks. In fact, the dual cue means converged on the ES prediction at about block 12 and remained roughly parallel with it for the remainder of the test. To evaluate whether the faster rate of speed up for the dual cue items is statistically significant, a within subjects analysis of variance (ANOVA) with factors of Fit (prediction versus dual cue data), Block, and their interaction was conducted for each model. For this and all other statistical tests, α was set to .05. In all cases, the effect of Block was highly significant. The ANOVAs confirmed the interactions between Fit and Block evident in the figure; the F statistics, each with 19 and 361 degrees of freedom, were 2.76, 4.40, and 7.94 for the RS, ES, and race models, respectively (all p's$<.001$). Because of these significant interaction effects, there is little to be gained by reporting the main effects of Fit. Next we investigate the model fits more precisely through RT distribution analyses conducted separately for data from the beginning and end of the test.

4.2.1. Cumulative distribution fits

Cumulative distribution tests were performed to determine whether the model fit results for the mean held across the entire RT distribution. No attempt is made here to fit parametric distribution models to the data. Rather, we simply compared the empirical distribution shapes for the observed and predicted dual cue RTs for each model.

For the race and LC models, each of the six dual cue RTs on a test block is matched to a single predicted RT that was derived from the single cue data on the same test block, in the manner described above, yielding six predicted dual cue RTs for each model. For the RS model, there are two predicted values for each dual cue item; one from each of its two component cues on the same test block. To equate the number of data points used in the distribution fits for each model, these 12 predictions on each block for the RS model were reduced to six predictions by randomly selecting one of the two single cue items as the prediction for each dual cue item. This random selection process exactly simulates the random cue selection process assumed in that model. For the ES model, there was a single prediction for each dual trial for most item triplets, but for some triplets both single cues constituted predictions. This situation resulted from the method used to correct the ES prediction for statistical bias (see Appendix B). Just as for the RS model, one of the two single item RTs for each of those item triplets was randomly selected as the ES prediction for the distributions fits. Analyses of the distributions that resulted from the simulations in Appendix B verified that the ES distribution prediction derived in this manner matched the true, known ES distribution, demonstrating the validity of the approach.

By hypothesis for each model, each of the six predicted RT values are random deviates taken from the same RT distribution as their six matched dual cue datum. Thus, if a given model is correct, the predicted and observed dual cue cumulative distributions for data averaged over items and subjects must be identical in the populations, and statistically identical in the samples. This is the case no matter how the data are treated, either without pooling, or by pooling over items or blocks to form quantiles of averaged data, provided that there is a matched single cue prediction for each dual cue observation. The data that are presented below were pooled into average quantiles based on a rank ordering of the item RTs for each subject within each test block. Separate cumulative plots of the non-pooled, raw data yielded essentially identical results. The results below are thus not an artifact of the particular data pooling procedure that was used.

Cumulative distribution plots were constructed in the following way. The RTs for each subject and each test block were rank-ordered over items from fastest to slowest, separately for dual cue items and for the sub-sets of the single cue data that constituted the predictions for each model. In the current experiment there were six dual cue items and six corresponding single-cue predictions per model for each block. This ranking procedure thus generated an estimate for each of six quantiles of the dual cue distribution, along with an estimate for each of six quantiles for each model, separately for each subject and test block. These rank-ordered data were then averaged, by rank, over the appropriate subset of the test blocks for a given fit (see below), and then over subjects, yielding one grand mean for each of the six quantiles for the dual cue items, and one grand mean for each of the six quantiles for each of the model predictions. Thus, all data that were in rank position one were averaged together over blocks and then over subjects to produce a grand mean for quantile one, etc. Note that the item ranking was done purely on the basis of RT, so that a given item might be in different rank positions in the observed and predicted distributions due to random fluctuations. All responses, regardless of accuracy, were included in this analysis. Typically this procedure raises the possibility that error RTs may distort the distributions. However, in this case, all fitted models make RT predictions for both correct and incorrect trials, thus ruling out this type of bias.

For the test phase of this and all subsequent experiments, separate distribution fits were performed for data from the first few test blocks and for data from about the last half of the test. The fits to the first few test blocks were designed to test the models under conditions in which cues are most likely to be independent. The blocks to be included in these fits were determined separately in each experiment by first inspecting the quantile values on the first block for the dual cue items, relative to the values predicted by the models, and then including all successive test blocks on which the ordinal relations between the dual cue RTs and model predictions on each quantile remained the same. For example, if on block one all RT quantiles for the dual cue data were above the prediction of the RS model, then data from subsequent, consecutive blocks were averaged with the first block data only if dual cue RTs continued to be roughly equal to or above the RS prediction on all quantiles on those blocks.[5] In this way, noise in the distribution fits to the data from the beginning of the test phase could be reduced without the risk of masking the ordinal level outcome. Selection of blocks to include in the distribution fits toward the end of the test was done by visually inspecting the graphs of the mean RTs and estimating the first block at which roughly steady-state performance had been reached. That and all subsequent test blocks were included in the distribution fit. In no case was the distribution fit materially affected when the estimate of the onset of steady-state performance was moved to a later block.

Cumulative distribution fits of the various models are shown in Fig. 2. Data averaged over the first three test blocks are shown in Panel a and data averaged over the last 9 test blocks are shown in Panel b. For each model fit discussed below, the mean over quantiles of the absolute differences between dual cue data and the model prediction (the objective function in fit optimization for the LC model) is provided. For the LC model, the parameter value, c, for the optimal fit is also provided. Optimization for that model was achieved by iterating through values of c, from 0 to 1.0, in increments of .01, to find the smallest mean absolute deviation. The LC model fit was optimized for the cumulative distribution fit to the beginning of test, and c was set to the same value for the fit to the end of test. After optimization, matched

[5] Blocks exhibiting a numerically minor reversal for one or two quantiles that did not replicate on the following block were also included under the assumption that these reversals were due to random fluctuations.

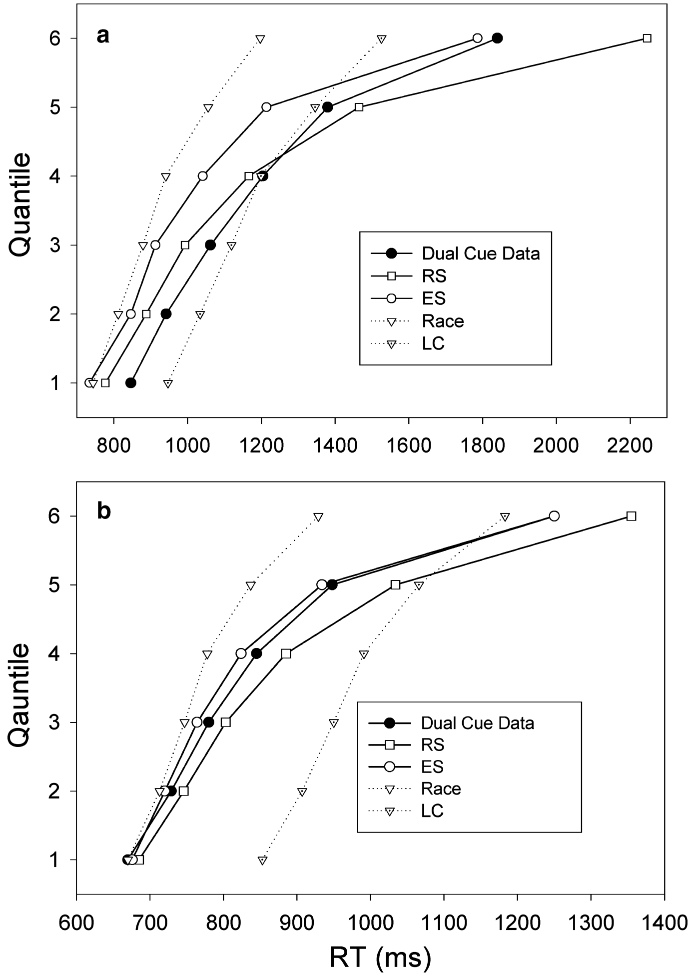

Fig. 2. Model fits to the cumulative RT distribution for dual cue items in Experiment 2. Data in Panel a are averaged over the first three blocks of the test and in Panel b over the last nine test blocks.

t tests were performed on the subject level means for each quantile for each model (see Miller, 1982, for simulations speaking to the validity of this approach). Testing of distribution fits in this way has an advantage over global fit tests, such as χ^2, in that it provides explicit information about quality of fit at multiple locations of the distribution. It is especially useful when there are multiple opportunities for replication, as was the case here.

For the first three blocks (Panel a), the RS model under-predicted RTs on the lower quantiles (1–3, 120; i.e., the predicted and observed dual cue RTs were significantly different for the first three quantiles and the mean absolute prediction deviation was 120 ms). The ES (1–5, 123) and race (1–6, 274) models also under-predicted RTs. The

LC model fit the data best (1–3, 6; 100; $c = 1.28$). Note, however, that the LC model must fit the central tendency well due to its free parameter. For the last nine test blocks (Panel b), the RS model over-predicted (1–6; 47), and the race model under-predicted (2–6; 92) the dual cue RTs, but the fit of the ES model was quite good (nil; 11). The LC model fit poorly (1–5, 143).

4.2.2. Error analysis

Error rates were analyzed for the same subsets of data that were analyzed in the distribution fits. Inspection of the error data in all experiments revealed substantial floor effects as well as pronounced right skew. Some subjects made no errors, most made occasional errors, and typically one or two subjects made frequent errors, particularly at the beginning of a new dual cue test phase. As such, ANOVAs on the mean error data are not appropriate. Instead, we used the non-parametric Wilcoxon test for matched samples (Hayes, 1988). The dual cue error rate collapsed over the first three test blocks was .04, compared to predictions of .061 by the RS model ($p = .010$), and .038 and .034 by the ES and race models, respectively (p's>.2). On the last nine blocks of the test, the dual cue error rate was .008, compared to predictions of .023 for the RS model ($p = .059$), and .014 and .019 for the ES and race models, respectively (p's>.2).

A similar basic pattern of results was obtained in Experiments 2–5 below. Error rates in those experiments were low with substantial floor effects and did not meaningfully discriminate among the models. As such, the error results will not be discussed further.

4.3. Discussion

The results are more consistent with the cue selection class of models. The finding that the mean dual cue RTs were initially above even the RS prediction is striking and unanticipated by the literature. But if subjects must select one cue at the expense of the other, it makes sense. It is unclear how subjects would immediately know, on the first dual cue trial, which cue is more efficient for each pair. The RS model did not fit the data very well at the distribution level, however, suggesting that the slow initial performance did not reflect purely random cue selection. The convergence of RTs on the ES prediction for both the mean and the entire distribution toward the end of practice suggests that subjects become much more efficient at cue selection over the course of the test, possibly through mechanisms to be considered in Section 9. The simplest interpretation of the test results is that subjects underwent a transition from some type of "inefficient" cue selection to efficient selection.

One of the more powerful sources of evidence against the LC model is that it, along with the race model, exhibited systematic under-prediction of the variance and skew in the distribution fits. These effects can be understood by noting that slower RTs are filtered out by the competition between the two cues in parallel models. This phenomenon is not specific to the parallel models that were fit. Rather, it applies to the optimal fits of many types of parallel models in which the first cue to retrieve a response determines performance.

5. Experiment 2

This design is identical to that of Experiment 1, with the following three exceptions. First, during the initial study phase, both dual and single cue items were presented. This modification allowed us to determine whether the results of Experiment 1 might reflect some type of dual cue novelty effect at the outset of the test phase. Second, the left–right ordering of the dual cues was not reversed from block-to-block. It is possible that, if the dual cues are always in the same spatial order, dual cue facilitation will be observed after sufficient practice. Third, we included a transfer phase after the main test phase, in which all of the old items, as well as the left–right spatial reversal of all items, were presented.

5.1. Method

5.1.1. Subjects

Sixteen University of California at San Diego undergraduate students participated for course credit.

5.1.2. Materials, design, and procedure

Each subject received three mandatory study blocks, and, if they so chose, a maximum of two additional study blocks. Subjects next went into the test phase, in which all single and dual cue items were mixed within each block. After the 20 test blocks were completed, subjects were given three transfer blocks. On each transfer block they were presented with (1) all old single and dual cue items, (2) reversed single cue items (if left of fixation asterisk during test, then right of asterisk at transfer, and vice versa), and (3) reversed dual cue items.

5.2. Results and discussion

The mean correct dual cue RTs are shown in Fig. 3, along with the model predictions. The dual cue RTs closely match the ES prediction until about the sixth test block, after which they fell below it, approaching the race prediction by the end of practice. The ANOVAs (identical to those performed in Experiment 1) revealed a significant Fit by Block interaction effect for the race model, $F(19, 285) = 2.71$, $p < .001$, but not for either the RS, $F(19, 285) = 1.17$, $p > .2$, or ES model, $F(19, 285) = .66$, $p > .2$. A supplemental linear regression on the difference scores between the ES and dual cue RTs on each block, however, revealed a reliable slope, $t(15) = -2.8$, $p = .012$, indicating that there was an interaction even for the cue selection models.

Although the inclusion of the dual cue items in the initial study phase of this experiment resulted in significantly faster mean dual cue RTs at the beginning of test, those RTs were still most consistent with the ES model. Note also that associative independence may not have held at test in this case due to the prior dual cue practice. The consistent left–right ordering of the dual cue stimuli is presumably a factor behind the substantially faster dual cue RTs by the end of test in this experiment, in

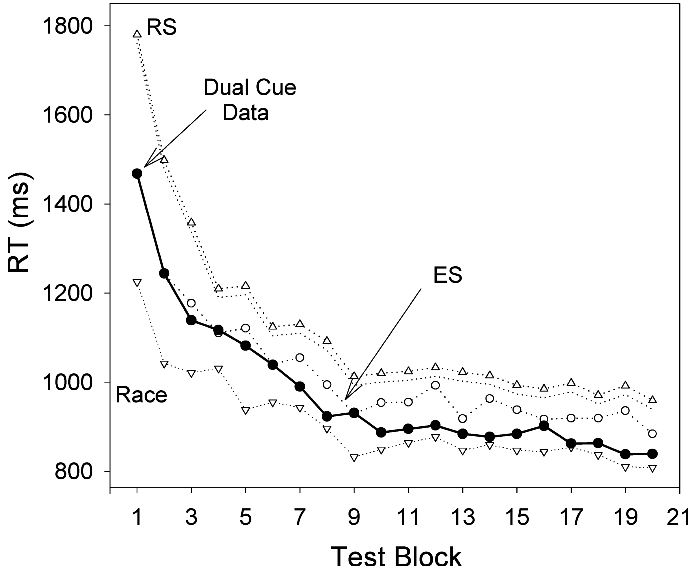

Fig. 3. Mean dual cue RTs for Experiment 2 as a function of test block, plotted against the predictions of the RS, ES, and race models. The dotted line just underneath the RS prediction is the mean of the single cue data, unadjusted for $\mu_{\Delta p}$.

comparison to Experiment 1, although the introduction of dual cue items during the initial study phase might also have played a role in that outcome. In any case, the ES model clearly does not hold by the end of practice in this case.

It is apparent that none of the models developed in the introduction can fully account for dual cue performance after practice in this experiment. One elaborative account is that independent parallel retrieval may only emerge with increasing automatization of single cue items, and perhaps also with increasing automatization of the task context and goals. According to such a model, dual cue processing would initially reflect cue selection, but would eventually transform into independent parallel processing. A related account is that the capacity that is required to retrieve through a given cue decreases as retrieval gets less effortful with dual cue practice. Neither of these models can directly explain why dual cue facilitation after practice was observed here and not in Experiment 1, though it could be that introduction of dual cue items during the initial study phase is needed for such a transition to occur. A third possibility is that in this experiment the cue independence assumption was violated after sufficient practice in the form of new, chunked dual cue representations, somehow leading to the differentially greater speed up for dual cue items. Although this account also does not directly predict the greater dual cue facilitation in this experiment than in the first experiment, it would be natural to assume that a chunking mechanism has enhanced impact on performance when the spatial cue ordering is constant for each dual cue item.

The transfer phase allowed us to discriminate among these possibilities. According to both the automatized parallel retrieval and the decreasing capacity demand accounts, spatial reversal of dual cue items at transfer should have no effect on RT, provided that spatial reversal of the single cue items has no effect on RT (i.e., provided that single cue learning during the test did not code spatial location of the cues relative to the preceding fixation asterisk, or to any other spatial reference point). According to the chunking account, on the other hand, reversing the cue ordering at transfer should disrupt performance, provided only that the chunked dual cue representations that formed during test were specific to spatial cue ordering.

The transfer test had six conditions: left old, right old, left reversed, right reversed, dual old, dual reversed. Here, the terms "left" and "right" refer to the spatial location of the cue (left or right of the fixation point) during the test phase. Old items were identical to items seen during test. "Left reversed" refers to single cues presented on the left side of fixation during the test but on the right of fixation during the transfer, and "right reversed" refers to the opposite. Reversed dual cue items were the same as old dual cue items, with the exception that the cues were left-to-right reversed.

The results, collapsed across the three transfer blocks, are shown in Fig. 4. These data were analyzed by a within subjects ANOVA with a six level factor of Condition. Four single degree of freedom contrasts were performed. First, it is immediately apparent that reversal of cues on dual cue trials had a marked impact on performance. Reversed dual cue items were 118 ms slower than old dual cue items, $F(1, 17) = 16.31$, $p < .001$, and 40 slower than even the RS prediction for the transfer test, $F(1, 75) = 2.94$, $p = .090$. However, a contrast comparing old to reversed single cue trials (collapsed over the left, right distinction) was not significant, $F(1, 75) < 1.0$. Evidently, location of a cue relative to the preceding fixation asterisk was not coded as part of the single cue representation. These results rule out both the

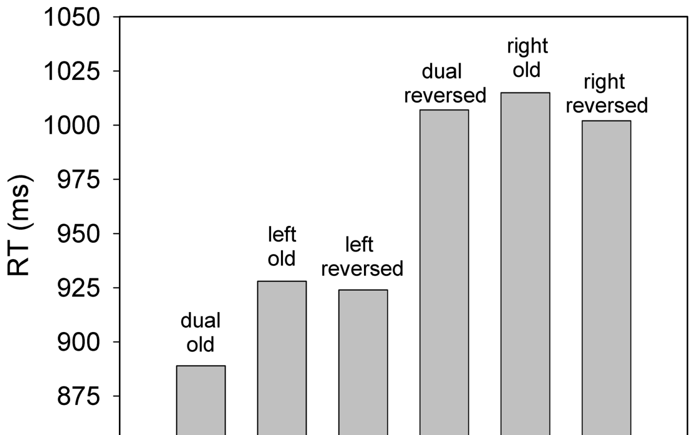

Fig. 4. Mean RTs for each of the transfer conditions in Experiment 2, collapsed over the three transfer blocks.

automatized parallel retrieval and the decreasing capacity demand hypotheses. Instead, they suggest a chunking account in which access to the dual cue representation—for purposes of retrieval at least—is possible only when the cues are presented in the previously learned spatial order.

A second notable transfer effect is that left-old and left-reversed single cue items were performed faster than right-old and right-reversed single cue items, $F(1,75) = 15.42$, $p < .001$. Because cue location was counter-balanced across subjects, this result cannot be attributed to differences in intrinsic retrieval difficulty of left- vs. right-side cues. Also, as noted above, spatial location relative to fixation was not coded into the single cue representations during the test. It thus appears that the RT advantage for single cue items that were always presented on the left during test is solely a result of a preference for retrieval through those cues on dual cue trials. If some subjects adopted this simplifying strategy, the left-side cue would get twice the retrieval practice of its companion (once as a single cue and once in its dual cue form within each test block), leading in turn to faster RTs for that cue on single cue trials. To explore this possibility, we compared the means for single cue items presented on the left during the test to those presented on the right. This analysis was restricted to test blocks 10–20, on which dual cue performance was stable relative to the model predictions. The left-side single cue items had a mean RT of 950 ms, 64 ms faster than the mean RT for right-side items. This effect size is in the same range as the 82 ms difference between left-old and left-reversed vs. right-old and right-reversed single cue items on the transfer test. This effect was reversed (but not significantly so) toward the beginning of test, suggesting that the left-side bias developed as a consequence of learning over test blocks. It appears that subjects did preferentially retrieve through the left-side cue on dual cue trials.

Note that within the cue selection framework, it would be natural for subjects to discover a simplifying selection strategy when possible during initial practice blocks. The retrieval architecture requires them to select only one cue for retrieval on dual cue trials. If they cannot always choose a cue efficiently at the item level, then a strategy of cue selection based on a category distinction (i.e., left vs. right side) is as good as any. On the other hand, according to the race model, the cue selection problem does not exist. Strategic retrieval through the left-side cue could only hurt performance, because it would not capitalize on the facilitation effect that results from a race. In the LC model, preferential retrieval through one cue could lead to optimal performance under conditions in which the capacity limitation more than cancels out the statistical facilitation due to parallel retrieval. However, placement of capacity allocation under strategic control complicates parallel models, and vastly increases their flexibility. We will return to this issue in Section 9.

Despite its substantial impact on performance, the left-side retrieval bias alone cannot explain the dual cue facilitation during test in this experiment. By the end of the test, the ES prediction was below the left-side cue RTs, and the dual cue RTs were significantly below the ES prediction. A dual cue chunking account is needed to fully explain these results.

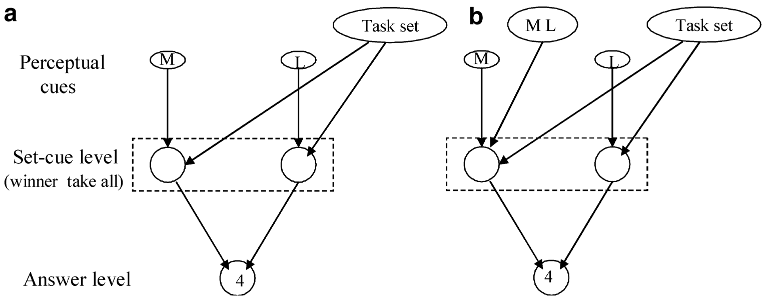

Fig. 5. A candidate chunking model within the cue selection framework.

5.2.1. Candidate chunking accounts

First consider the cue selection class of models. One plausible chunking account, an elaboration on the CMPL architecture (Rickard, 1997), is depicted in Fig. 5. Prior to dual cue practice (Panel a), the associative pathways from the two cues to the response are separate and independent, and there is no dual cue representation. After dual cue practice, a chunked dual cue representation might develop at the perceptual level (Panel b).[6] Note that although CMPL limits activation flow from the set-cue level forward, it places no constraints on the number of representations that can be active at earlier stages of processing, nor on the number of input nodes that can send activation in parallel to a given set-cue node.

We also assume that subjects are always retrieving the response through the more efficient cue of each pair by the time a stable dual cue chunk begins developing.[7] According to this chunking model, the dual cue chunk becomes directly associated with the pre-existing set-cue conjunction node of the efficient cue (M in the figure) as an automatic consequence of dual cue item repetition (Panel b). This account fits naturally within the CMPL framework. A new set-cue conjunction node is not needed to support the chunking effect, because a stable associative pathway to the required response already exists via the set-cue conjunction node for the efficient cue. On each dual cue trial, that set-cue conjunction node is activated by the efficient cue, and an association develops between it and the concurrently activated chunk. Once that association has formed, activation from both the efficient cue and the dual cue chunk may pass in parallel to the corresponding set-cue node.

There is insufficient information at present to justify any specific claims regarding how a newly associated dual cue chunk might facilitate retrieval through the efficient

[6] We will not address the issue of how the dual and single cue representations might be connected, if at all, at the perceptual level. We simply assume that the dual cue chunk can be approximately modeled, with respect to its effect on performance, as a third cue at the perceptual level. Since the data compel a chunking interpretation, some third neural representation, connectivity, or activation pattern must exist that is not engaged when the cues are presented alone (i.e., something new at a neural level must be generating the chunked facilitation effect). Considered at this most abstract level, it is this aspect of the representation that is hypothesized to become associated with the set-cue node for the efficient cue.

[7] In Section 9 we consider factors that justify this assumption.

set-cue node, but there are numerous plausible and non-exclusive possibilities, including a facilitated rate of activation of the set-cue node, an increased maximum activation level of the set-cue node, and a strengthened association from the set-cue node to the response that is fully tapped only when the dual cue is present. In all cases, the rate of activation at the response level could be facilitated on dual trials, even though response activation is still occurring through only one set-cue node.

The facilitation effect due to chunking can reasonably be assumed to be proportional to the efficient cue RTs over all dual cue items for a given subject. Thus, the simplest approach to implementing this model mathematically is analogous to that used for the limited capacity model; the RT for the efficient cue of each cue pair is simply multiplied by a coefficient, j, whose value in this case is less than 1.0. The mean RT prediction of this model, which we will term the efficient selection plus chunking (ES + C) model, is:

$$\mu_{\mathrm{D}} = j * \min(\mu_1, \mu_2) + \mu_{\Delta p}. \tag{5}$$

As was the case for the parameter c in the limited capacity model, the fit of this model was optimized in the RT distribution analyses by finding the value of j that minimized the sum of the absolute differences between the actual and predicted dual cue quantiles.

Chunked dual cue representations can be incorporated into parallel models as well, though the implementation is not quite as straightforward. One natural account would be that an independent dual cue chunk forms with practice and becomes a third racer when a dual cue item is presented. However, from the parallel retrieval perspective, capacity appears to already be taxed by two cues, resulting in performance at or above the ES level when the cues can be considered independent. Since the independent chunk would form later than the single cue associations, it would presumably be a weak racer, rarely winning but taking significant capacity from the single cue racers. A third racer under these conditions would most likely hurt rather than facilitate performance.

Alternatively, an approach analogous to that used for the ES + C account could be adopted. The dual cue chunk could somehow become part of the retrieval pathway for one or both of the single cues, facilitating retrieval through those pathways. Only two racers would be present on dual cue trials, eliminating the deficiency of the account outlined above. Such a chunking mechanism could be added to the LC model just as for the ES + C model, by multiplying the single cue retrieval RT by a coefficient of less than 1.0. However, within a limited capacity framework, one would not expect a chunked association to develop only for the more efficient cue of each pair, because so long as the RT distributions overlap, either cue can deliver the response first on any given trial throughout the test blocks. Rather, the dual cue chunk would become associated with each cue–response pathway in proportion to the frequency with which that cue wins the race on dual cue trials. This proportion is not knowable directly but can be approximated by tallying the proportion of blocks on which a cue yields a faster RT than its companion cue on single cue trials. The idea here is that the faster cue as indexed by single cue performance is more likely to win the race on dual cue trials.

We developed a simplest case quantitative model based on the second account above by computing, for one cue of each cue pair, the proportion of test blocks on which the RT for that cue (a single data point in each block) was faster than that of its companion cue. The result was a weighting variable, w, for estimated the chunking effect for that cue on dual cue trials. For a given cue pair, w will refer to the weighting factor for one cue (cue 1), and $(1 - w)$ to the weighting factor for its companion (cue 2). The overall chunking facilitation effect for a given cue pair was given by the free parameter, k.[8] The dual cue RT facilitation coefficients for each cue pair were then computed as:

$$k_1 = 1 - w(1 - k), \tag{6}$$

$$k_2 = 1 - (1 - w)(1 - k). \tag{7}$$

If, for example, the overall chunking parameter, k, is set at .9 (implying a 10% overall facilitation effect), w is .4, and $1 - w$ is .6, then the facilitation coefficient for cue 1 is $1 - .4(1 - .9) = .96$, and that for cue 2 is $1 - .6(1 - .9) = .94$.

This limited capacity plus chunking (LC + C) model predicts that:

$$\mu_{\mathrm{D}} = \mu_{[\min(k1*c*RT1,\, k2*c*RT2)]}. \tag{8}$$

For the ES + C and LC + C models to provide complete accounts of dual cue practice effects, functions would be needed that describe how j (the chunking parameter for the ES + C model) and k, respectively, change with practice. However, given the ease with which decreases in retrieval RTs with practice are usually fit with relatively simple practice functions, there is little reason to believe that these models could be discriminated along those lines. Further, in our view it is of more value to first determine whether either model can fit the RT distributions under roughly "steady-state" conditions after sufficient practice (i.e., under conditions in which there is minimal block-to-block speed-up, so that practice effects can be ignored).[9] If either model cannot fit those data, there is little reason to develop it further. Hence, in this paper the candidate chunking models will be evaluated only for the approximate steady-state case, which corresponds to roughly the second half of the test in all experiments. Finally, because these models, like the LC model, could fit the mean arbitrarily, they will be evaluated only for the RT cumulative distribution fits.

[8] The method for estimating the predictions of both the ES + C and the LC + C models reduces not just the retrieval component of the RT distribution, as desired, but also the perceptual and motor components of those distributions. However, for the ES + C model, the parameter j typically took values of around .95. Given this small magnitude transformation, these factors would have little if any measurable impact on distribution shape. For the LC + C model, the values of the parameter, k, were often much smaller. However, the deflation of the RT distribution caused by this parameter partially reverses the inflation for each cue brought about by the transformation through the parameter, c. In any case, given the similarity in the shapes of the race, LC, and LC + C models, it is clear that this potential bias is not a factor in the quality of the fits.

[9] By steady-state, we do not imply asymptotic, but rather an interval of blocks over which the sample RTs changed minimally, if at all, and thus over which the cognitive processes can be assumed to be stable.

5.2.2. Cumulative distribution fits

As for Experiment 1, the distribution fits generally mirrored the results for the means (see Fig. 6). Over the first six test blocks (Panel a), the ES model provided the best fit (nil; 43). The RS model significantly over-predicted dual RTs (1–6, 284) and the race model under-predicted them (1–6, 161). The LC model faired better, but was rejected on the first and sixth quantiles (1, 6; 85; $c = 1.11$).

On the last 10 test blocks (Panel b), the RS (1–6, 114) and ES (1–5; 53) models over-predicted dual cue RTs, whereas the race model under-predicted them

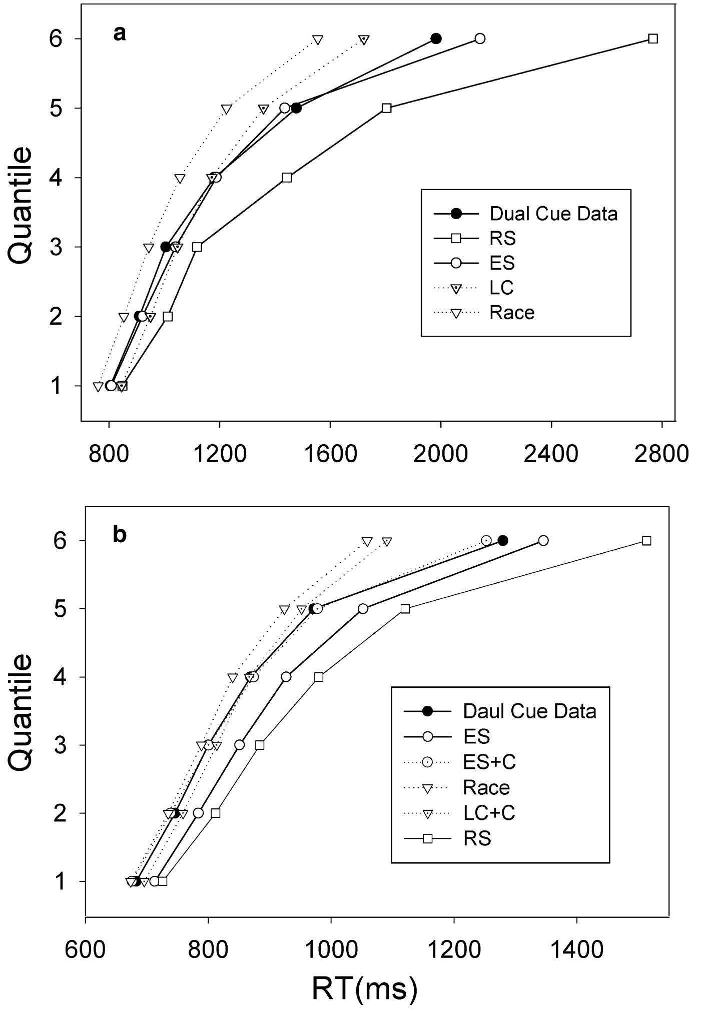

Fig. 6. Model fits to the cumulative RT distributions for dual cue items in Experiment 2. Data in Panel a are averaged over the first six blocks of the test and in Panel b they are averaged over the last 10 test blocks.

(4–6; 54). The LC + C model fit reasonably well but again under-predicted the skew (6, 42, $k = .81$). The fit of the ES + C model, however, was nearly exact (nil; 9.5; $j = .95$). The fit of the LC model (not shown), was substantially above and worse than that of the LC + C model, just as for the end of practice in Experiment 1.

The overall results are again more consistent with the cue selection class of models. The fact that the one parameter ES + C model fit the steady-state distribution data so well even when the means were well below the ES prediction suggests that a single associative pathway, perhaps mediated by a single set-cue conjunction node, may in fact continue to govern performance indefinitely. Of course, only moderate levels of practice were given in these experiments. The question of whether the ES + C model would describe the dual cue RTs after extended practice remains to be addressed. To our knowledge, however, there is no evidence in the literature that processes underlying retrieval performance change fundamentally after the rate of speed up with practice has already become so small (but see Schumacher et al., 2001, for evidence that RT interference decreases markedly after practice on a dual choice RT task).

6. Experiment 3

In Experiments 3–5, dual cue items consisted of one letter and one color cue instead of two letter cues. Use of these two cue dimensions has a number of useful consequences. First, there is strong evidence that letter and color stimuli can be perceived in parallel with essentially no processing delay (Mordoff & Yantis, 1993; More & Osman, 1993), eliminating this factor as a nuisance variable. As such, no perceptual correction was applied to the model predictions derived from the single cue data (i.e., $\mu_{\Delta p}$ was assumed to be zero). Second, it is possible that, whereas two cues from the same category cannot be used in parallel for retrieval, two cues from different categories, such as letters and colors, can be. Third, in the current experiment and in Experiment 5, dual cue items took the form of colored letters. This design allowed us to explore whether spatially integrated cue dimensions might yield parallel retrieval from the outset of practice. Finally, use of color and letter cues brings these experiments somewhat closer to the congruent condition in the Stroop task. It may be of interest to compare results from these two related paradigms.

6.1. Method

6.1.1. Subjects

Twenty University of California at San Diego undergraduate students participated: 13 for course credit, and 7 for a financial reward.

6.1.2. Materials, design, and procedure

The single cue stimuli consisted of seven single letters (presented in light grey on a black background) and seven colored X's. The dual cue stimuli were colored letters. For each stimulus triplet (i.e., a colored X, a letter, and the corresponding colored letter), total luminosity, number of activated pixels, and the rectangular space covered, were equated.

The single-cue learning phase of this experiment was divided into two subsections: one for the letter stimuli, and the other for the color stimuli. The order in which these two subsections were presented was counterbalanced. Within each subsection, the structure closely matched the learning phase from Experiment 1: i.e., study blocks (three mandatory; one optional) followed by single-cue performance-blocks (with each appropriate stimulus presented once per block). The first subsection concluded when the subject completed two consecutive performance blocks with 100% accuracy, and a mean RT of 1000 ms or less on the last block. For the second subsection, there was the added requirement that the mean RT on the last block could be no higher than that from the final block of the preceding subsection. This procedure helped match the RTs for color and letter cues. These learning criteria were slightly more stringent than those of the earlier experiments, and were included in an effort to increase accuracy during the test phase.

Next, each subject received one mixed single cue block. This block was structured like a block from the testing phase, except that no dual cue stimuli were presented. Each letter and color stimulus was presented once in this block, mixed randomly. This block was included to refresh subjects' memory for the items learned during the first learning phase and to familiarize them with presentation of both cue categories in a mixed fashion.

At the start of the test phase, subjects were informed that some of the following trials would involve letter cues, some would involve color cues, and some would involve colored letter cues. Subjects were explicitly told that, whenever a colored letter was presented, the color and letter would always correspond to the same digit (learned previously), and that this digit would be the appropriate answer. To ensure that each subject understood this, all subjects were required to summarize the instructions in their own words.

In the testing phase, the basic format from the earlier experiments was used; however, the message "Get Ready!" was replaced with an asterisk fixation point which flashed twice in the center of the screen. Each stimulus appeared in the exact location of the preceding fixation point. Each testing block consisted of 21 trials, with each color, letter, and colored letter stimulus presented once, in random order. A total of 20 test blocks were presented.

6.2. Results and discussion

Fig. 7 shows the mean dual cue RTs, along with predictions of the RS, ES, and race models. On the first test block, mean dual cue RTs were again above the RS prediction, but fell below it on the second test block, and converged with the prediction of the race model by the last test block. The interactions between Fit and Block evident in the figure were confirmed by ANOVAs: the $F(19, 361)$ values were 2.38, 4.2, and 5.37, for the RS, ES, and race models, respectively (all p's<.001).

The cumulative distribution fits are shown in Fig. 8. On the first test block (Panel a) the RS model prediction was not significantly different from the dual cue RTs on

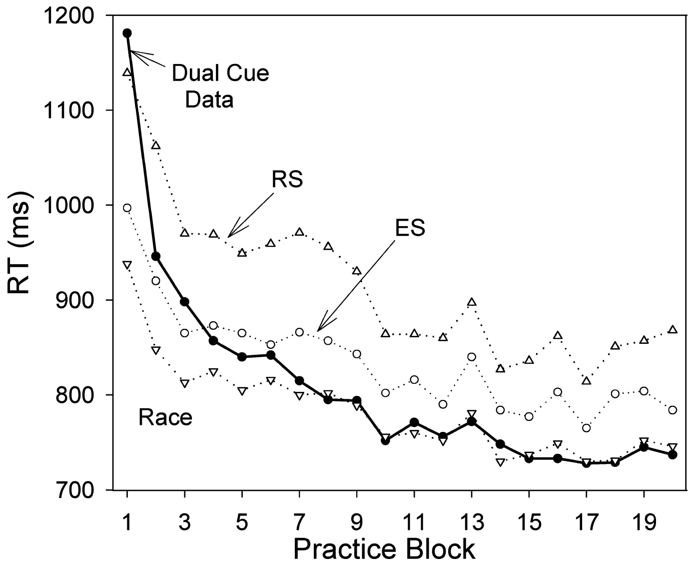

Fig. 7. Mean dual cue RTs for Experiment 3 as a function of test block, plotted against the predictions of the RS, ES, and race models.

any quantile (nil; 67). The ES (2–6, 165) and race (1–7, 289) models dramatically under-predicted the RTs. The fit of the LC model was also relatively poor (1–3, 7; 132; $c = 1.29$). On the last 13 test blocks (Panel b) the RS (1–7; 113) and ES (1–6 ; 40) models both over-predicted the dual cue RTs. The race (7; 26) and LC + C (7; 26; $k = .56$) faired better. The fit of the ES + C model was again excellent (nil; 7, $j = .945$).

It appears that even in the case of mixed color and letter cues, subjects must select and retrieve from only one cue when the cues are independent. With practice, the dual RTs decreased quickly, nearly matching the race prediction (with respect to the means) by the end of practice. Nevertheless, the dual cue RT distribution was still best fit by the ES + C model. This finding lends additional support to our hypothesis that a single set-cue conjunction node continues to mediate performance after dual cue practice, even when dual cue chunking has occurred and clear RT facilitation is present. It is important to note, however, that in this case the best fitting cue selection model has one free parameter, whereas the best fitting parallel model, the race model, has none.

From the perspective of the parallel retrieval framework, the accelerated rate of dual cue speed-up, and the convergence of the dual cue mean on the race prediction, raise the possibility that either the automatized parallel retrieval or the decreasing capacity demand hypotheses, which were ruled out for the case of two letter cues, might nevertheless hold for the case of mixed color and letter cues. Color and letter cues may be stored and processed in different neural networks, perhaps reducing interference on dual cue trials. Alternatively, since the letter and color cues are integrated in this experiment, conditions may also be optimal for dual cue chunking.

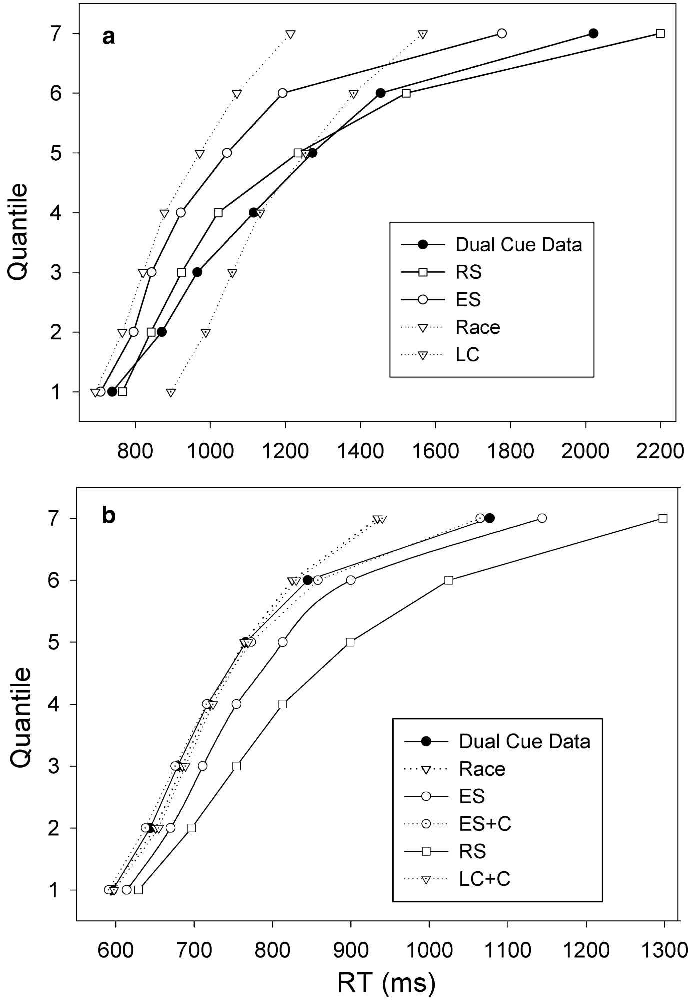

Fig. 8. Model fits to the cumulative RT distribution for dual cue items in Experiment 3. Data in Panel a are for the first block of the test and in Panel b averaged over the last 13 test blocks.

It may be much easier for the memory system to encode a colored letter as a single object than to encode two letters (or a color and a letter) presented side by side as a single object. In Experiment 4, we tested these accounts by spatially reversing the ordering of the dual cue dimensions from block-to-block.

7. Experiment 4

This experiment was designed to be as similar as possible to that of Experiment 3, while incorporating the side-by-side, block-to-block reversed, dual cue format of

Experiment 1. Since chunking in this case should be more difficult than in Experiment 3, the cue selection account predicts that dual cue RTs will be above the ES prediction for significantly more practice blocks than in that experiment. On the other hand, the automatized parallel retrieval account does not depend on dual cue chunking to predict an onset of dual cue facilitation with practice, and thus predicts that dual cue performance will be analogous to that observed in Experiment 3.

7.1. Method

7.1.1. Subjects

Twenty-one University of California at San Diego undergraduate students participated for course credit.

7.1.2. Materials, design, and procedure

The experimental design closely matched that of Experiment 3, with the following exceptions. First, there were six, rather than seven, items in each stimulus category. Second, each single color cue was presented in the form of a 3 mm × 5 mm rectangular patch instead of an X. Dual cue stimuli consisted of a letter and a color patch presented side-by-side. As in Experiment 3, members of each stimulus triplet were equated for total luminosity, number of activated pixels, and rectangular area covered. As in Experiment 1, single color and letter cues were left–right reversed from block-to-block with respect to fixation, and the two cues of each dual cue item were also left–right reversed from block-to-block. There were 20 test blocks.

7.2. Results and discussion

Fig. 9 shows the dual cue RTs, along with predictions of the RS, ES, and race models. Dual cue RTs on the first four test blocks were again above the RS prediction. With practice they gradually fell below it, roughly converging on the ES prediction by the end of practice. ANOVAs again revealed significant interactions between Fit and Block for all models: The $F(23, 460)$ values were 2.38, 2.4, and 3.56 for the RS, ES, and race models, respectively (all p's<.001).

Cumulative distribution fits are depicted in Fig. 10. On the first three test blocks (Panel a), the RS model provided the best fit (nil; 62), whereas the ES (1–6, 177) and the race models (1–6, 257) under-predicted RTs. The fit of the LC model was better, but still poor on the distribution tails (1, 2, 6; 166; $c = 1.24$). On the last 10 test blocks (Panel a) the RS model over-predicted (3–6; 101) and the race model under-predicted (1–6; 94.5) dual cue RTs. However, the fit of the ES model (nil; 12) was once again quite good. The LC + C model was somewhat competitive in this case (6; 45, $k = .74$), but exhibited the under-estimation of skew that has been characteristic of all of the parallel model fits.

These results mirror those of Experiment 1. It appears that spatial reversal of cues from block-to-block severely disrupts chunking for not only dual letter cues, but also for letter and color cues. These results speak against both the automatized parallel retrieval and the decreasing capacity demand account of practice in Experiment 3.

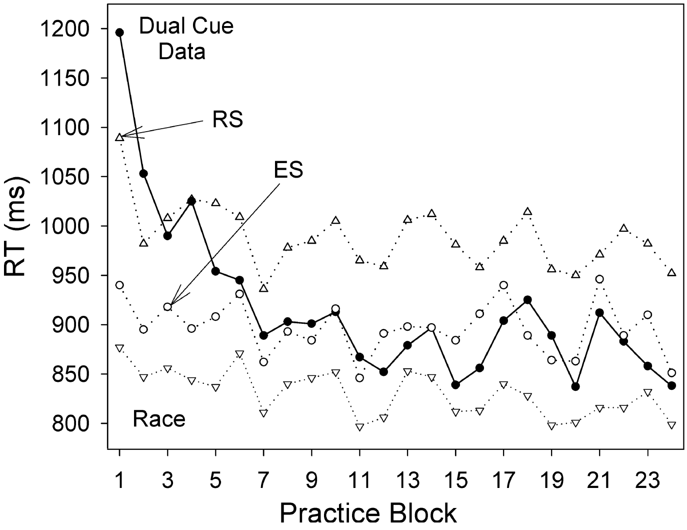

Fig. 9. Mean dual cue RTs for Experiment 4 as a function of test block, plotted against the predictions of the RS, ES, and race models

When chunking is not possible, or is at least substantially delayed, performance over test blocks is well captured as a straightforward transition from random, or at least inefficient, cue selection, to efficient cue selection.

The good fit of the ES model to the steady-state data when cue order is reversed from block-to-block (Experiments 1 and 4) raises a potentially serious question that must be addressed in any future modeling efforts within the parallel framework. If dual cue performance reflects limited capacity parallel cue processing, then the zero free parameter ES prediction is arbitrary, having no more significance than any other random function. Why, then, would the dual cue data conform so well to the ES prediction for the case of steady-state performance? The same question is raised regarding the one parameter ES + C model fits to steady-state data in Experiments 2 and 3.

8. Experiment 5

Despite our earlier arguments to the contrary, it is still conceivable that the slow dual cue performance on the first few test blocks in Experiment 3 reflects some unspecified novelty effect that was unique to the dual cue items. Perhaps parallel retrieval from independent cues is possible only when each of the following conditions is met: (1) each cue is from a different category, (2) the cues are two dimensions of a single object, (3) single cue performance has been sufficiently automatized, and (4) any possible dual task novelty effects at the beginning of test have been eliminated. This experiment tests this possibility. The basic design is similar to that of Experiment 3, with the following exceptions. First, in this case eight letter and

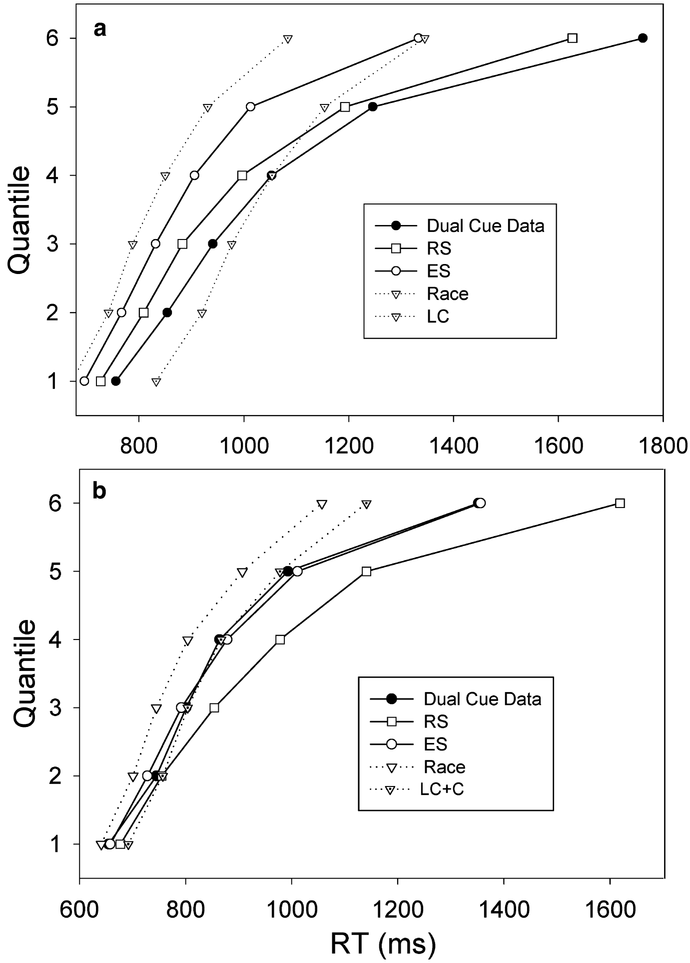

Fig. 10. Model fits to the cumulative RT distribution for dual cue items in Experiment 4. Data in Panel a are averaged over the first three blocks of the test and in Panel b over the last 10 test blocks.

eight color cues were associated with only four responses (one response for each stimulus pair in each cue category) during training. Note that this design results in two non-overlapping sets of dual cue pairings for each response. For example, consider a case in which two letter cues, A and B, and two color patch cues, Red and Green were all associated with the response "3" during single-cue learning. One possible set of dual cue pairings is (A, Red) and (B, Green). A second set of dual cue pairings is (A, Green) and (B, Red). During test, all single cues, but only one set of dual cue pairings for each response was presented, yielding a total of 24 items per test block. Drawing on the example above, subjects would see A, B, Red, Green, (A, Red), and (B, Green), all having the same response, "3." Second, the test phase in this experiment was followed by a transfer phase. The transfer phase was identical to the test phase, with the exception that here the dual cue items from test were replaced with the second set of dual cue items. In the example, (A, Red) and (B, Green) were removed from the stimulus set above and replaced with (A, Green) and (B, Red). There were still 24 items per block in the transfer phase. As in Experiment 3, dual cues were dimensions of a single object (i.e., a colored letter). However, the single color cues were presented as rectangular color patches, as in Experiment 4.

We expected the results of the test to be similar to those observed in Experiment 3. The transfer phase provided the critical test of the models. On the first transfer block, subjects were already familiar with performing dual cue trials, and had already received considerable practice with all the individual cues. Subjects were also highly familiar with the task requirements and with the experimental context and flow. The cues of each dual cue item, however, had not been seen together before, so any dual cue chunking that may have occurred during the test phase was completely eliminated. This manipulation cuts to the heart of the models. Any cue selection model must predict that dual cue RTs will be at or above the ES boundary on the first block of the transfer phase across the entire distribution, regardless of the level of dual cue performance during test. The automatized and decreasing capacity demand parallel models, on the other hand, predict that cue recombination in phase two will have no effect on dual cue performance.

8.1. Method

8.1.1. Subjects

Sixteen University of California at San Diego undergraduate students participated for course credit.

8.1.2. Materials, design, and procedure

The design and procedure were similar to those of Experiment 4, with a few exceptions. As noted above, two color and two letter stimuli were associated with each of four digit responses (4–7) during the learning phase. The dual cue items were colored letters. There was a 10 block test phase followed by a 20 block transfer phase. Crucially, the cue members of new dual cue stimuli that were presented in the transfer phase had never been seen together before, and thus are independent. The onset of the transfer phase occurred without notification to the subject. Indeed, only one subject reported even noticing the recombination of the dual cue stimuli in a post-experimental interview. Subjects were given a brief rest before the start of the first test phase, and also after the 8th, 16th, and 24th blocks of testing.

8.2. Results and discussion

The RT results are plotted in Fig. 11. The vertical dotted line demarks the transition from the test phase to the transfer phase. The pattern for the test phase is familiar. Dual cue RTs were initially above the RS prediction, but decreased quickly on a trajectory toward RTs that are faster than the efficient selection prediction.

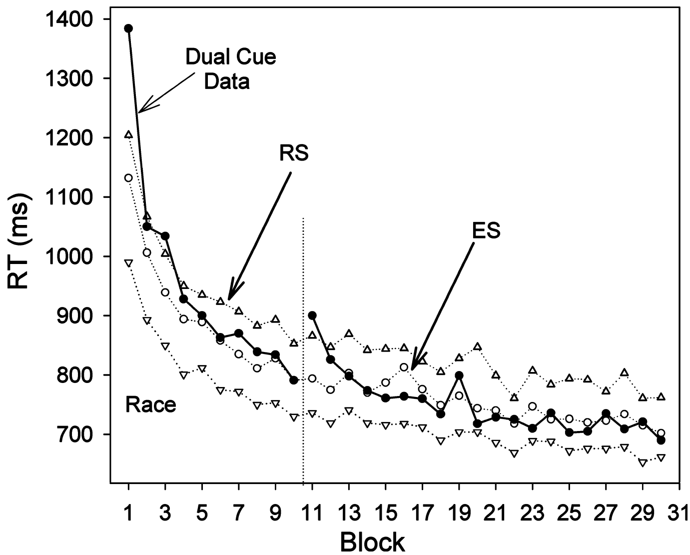

Fig. 11. Mean dual cue RTs for Experiment 5 as a function of test block, plotted against the predictions of the RS, ES, and race models.

The interactions between Fit and Block were all significant: The $F(15, 135)$ values were 5.42, 5.03, and 6.44 for the RS, ES, and race models, respectively (all p's<.001).

Crucially, dual cue RTs were once again above the ES and RS prediction on the first block of the transfer phase. Paired t tests were performed comparing the dual cue and ES means on blocks 10 (the last test block) and 11 (the first transfer block). On block 10, the 3 ms difference was not significant, $t(15) = -.025$. However, on block 11, the difference was substantial (109 ms) and significant, $t(15) = 2.4$, $p = .029$. As usual, dual cue RTs decreased substantially over the course of the transfer blocks. The interactions between Fit and Block were significant for all models: The $F(19, 285)$ values were 2.44, 2.2, and 2.39, respectively, for the RS, ES, and race models (all p's<.01).

Cumulative distribution fits are shown in Fig. 12. On the first three test blocks (Panel a), the RS model provided the best fit (4, 5; 90), whereas both the ES (2–7; 155), and the race models (1–8, 282) consistently under-predicted RTs. The LC model fit was better than the race fit (1–3, 8; 142; $c = 1.25$), but again lagged that of the zero free parameter RS model. On block 11 (Panel a), the first block of the transfer phase, the RS model again provided the best fit (4, 5; 40). The ES model fit the lower quantiles well, but under-predicted the upper three quantiles (5, 6, 7; 100), and the race model under-predicted all quantiles (1–8; 163). The LC model also fit poorly (1–4; 109).[10] On the last 10 blocks of the transfer phase (Panel c), the RS model over-predicted almost all quantiles (1–7; 66), the ES model faired well but still

[10] The fit was not improved materially by allowing c to vary as a free parameter.

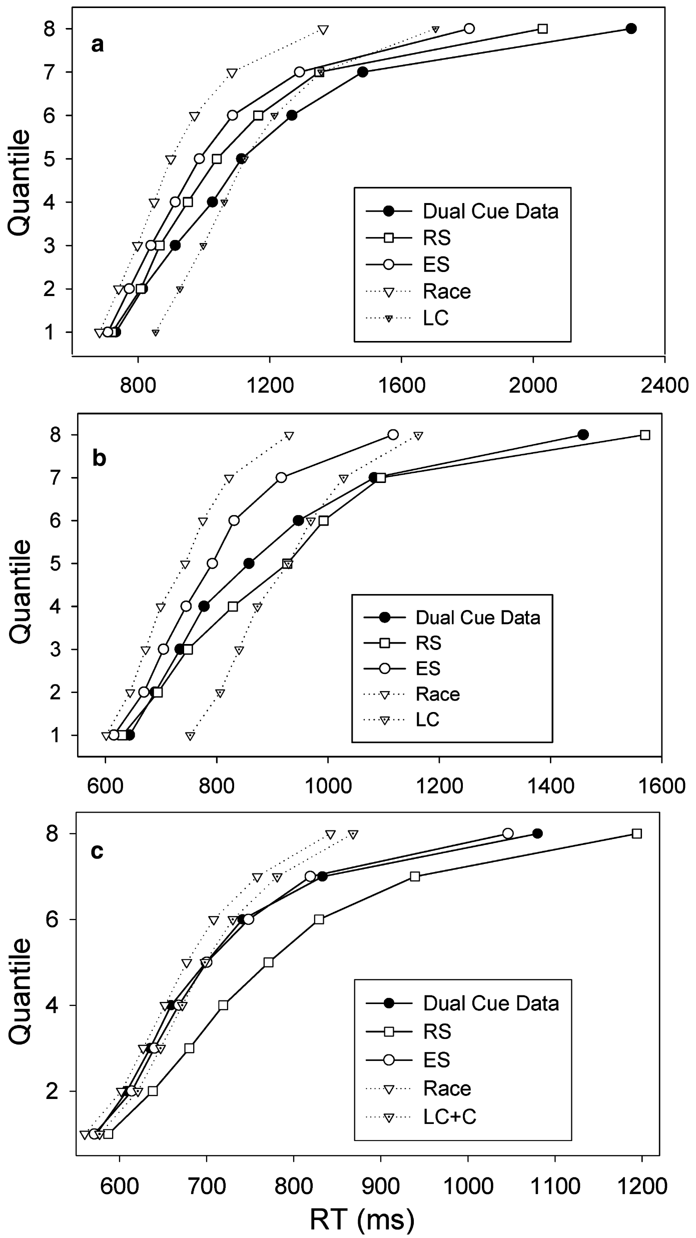

Fig. 12. Model fits to the cumulative RT distribution for dual cue items in Experiment 5. Data in Panel a are averaged over the first three test blocks, in Panel b for the first transfer block, and in Panel c are averaged over the last 10 blocks of the transfer phase.

over-predicted two of the low quantiles (2, 4; 10), and the race model under-predicted almost all quantiles (1, 3–8; 50). The LC + C fit was rejected on five quantiles (2–4, 7, 8; 39; $k = .65$). In this case the ES + C model could not improve on the ES fit in terms of mean absolute deviation, but it did eliminate the significant quantiles (nil; 9; $j = .99$; not shown in Fig. 12).

These results provide strong converging evidence for the cue selection class of models. Even when (a) general task level learning is eliminated at transfer, (b) cues are a color and a letter, and (c) the cues are merely different properties of a single object, two independent cues render no RT advantage over one cue.

9. General discussion

Our primary long-term goal in this line of investigation is to distinguish between two broad and theoretically pivotal classes of real time retrieval models under conditions of cue independence; the data so far support the cue selection class over the parallel class. The lower bound RT prediction for any member of the cue selection class is the ES prediction, which corresponds to the expected retrieval latency for the faster, or more efficient, cue. In contrast, parallel models as a class can predict dual cue facilitation below that level. Among the seven distribution fits at the beginning of a new test phase, involving 35 quantiles, there was no case in which the dual cue latency fell below the ES prediction. Indeed, dual cue RTs were substantially above it—in the range of the RS prediction—the majority of the time. This result was quite unexpected based on most of the literature reviewed in Section 1.

Though there was clearly no violation of the ES lower bound RT prediction at the group-level on the first few test blocks, there may still be individual differences in performance, such that RTs for some subjects violated the ES boundary condition. We investigated this possibility for the RT data from the beginning of test in Experiments 1, 4, and 5. In those cases there were sufficient numbers of blocks that passed the criterion for inclusion in the distribution analyses to perform subject level analyses. For each analysis, a difference score was obtained comparing the dual cue RT to the ES prediction (dual-ES) for each dual cue item on each test block for each subject. For example, the first three test blocks were included in this analysis for the beginning of test in Experiment 1. For each subject, there were six difference scores for each test block, and three blocks, yielding a total of 18 difference scores. Separately for each subject, these difference scores were analyzed using a one-sample t test. Across all three experiments, there were no subjects whose dual cue RTs fell significantly below the ES prediction. The group-level results appear to hold consistently at the subject-level.

It is important to acknowledge the possibility that subjects choose not to retrieve in parallel during the first few test blocks, even though in principle they could have. But for now that account is at best incomplete, because there is no obvious reason why subjects would have chosen such a strategy. One might argue that the cognitive system "prefers" to retrieve through one cue at a time, perhaps because capacity limitations on parallel retrieval are severe, but such an account has limited appeal unless there is some underlying principle from which one can predict a priori whether or not facilitation will be observed. The associative independence principle, in concert with the cue selection class of models, satisfies that condition.

The evidence favoring the cue selection class of models is just as strong for the second half of the test in each experiment, when roughly steady-state performance had been achieved. During the second half of the test in Experiments 1 and 4, performance on dual cue items exhibited a striking convergence with the zero-parameter ES prediction. In Experiments 2, 3, and 5, in which cues were not spatially alternated during test, dual cue RTs fell significantly below the ES boundary over the course of test and transfer. However, the ES + C model, which incorporates the simplest possible chunking account within the cue selection framework, uniquely fits those data quite well, with no evidence of systematic error in any case. It is a bit of a misnomer to refer to the ES + C model as a cue selection model, since it assumes that both a single cue and a dual cue chunk can participate in parallel in retrieval. However, the more central point is that this model can account for performance while preserving the assumption of a retrieval bottleneck at the set-cue level. The ES + C model is a member of the "set-cue" selection class of models.

In contrast, all of the parallel models were rejected on multiple distribution quantiles in nearly all cases. As we noted already, the ES and ES + C predictions are arbitrary from the perspective of parallel retrieval models, and they have zero and one free parameter, respectively. It therefore seems quite unlikely that dual cue performance would match the predictions of those models so well if retrieval was occurring in parallel from both cues, be it capacity limited or not.

9.1. Toward an integrated cue selection account

Because random selection is the most viable account of initial dual cue performance among the models considered, it is important to consider how the apparent transition to efficient cue selection might have taken place with dual cue practice. The simplest account rests on the assumption that subjects begin to systematically select one of the cues for each dual cue item within the first few test blocks. Cue selection might become systematic, for example, if subjects randomly chose a cue on the first exposure to a dual cue item, strengthening that cue's associative pathway, and in turn increasing the likelihood that that cue would be selected on the next test block, etc. Alternatively, as in Experiment 2, some subjects may have chosen cues based on cue category (e.g., left- or right-side). In either case, the result, for most cue pairs, would be the onset of systematic retrieval through one of the cues within the first few practice blocks. The preferred cue of each pair would then receive twice the retrieval practice of its companion cue (once on its single cue trial and once on its dual cue trial on each block). Since retrieval practice is a primary determinant of retrieval speed, it seems inevitable that dual performance would transition to roughly the efficient selection level with sufficient practice. This may simply be an automatic consequence of performance.

The excellent fits of the ES + C model to the steady-state data in Experiments 2 and 3 further suggest a two-stage transition process with practice. The first stage is from roughly random to efficient cue selection, and the second stage is from efficient selection to chunked efficient selection, provided that chunking is possible in the task. This general framework has the potential to account for virtually all of the RT patterns in the five experiments.

9.2. Limited capacity retrieval models

The LC model only modestly improved the distribution fits over the race model. Its primary limitation was under-prediction of the skew on the upper tails of the RT distributions. Even for the second half of the test in Experiment 3, on which mean dual cue RTs were quite close to the race model prediction, the under-prediction of skew by the race and LC+C models was clear. In contrast, ES and ES + C models fit the upper tail of the dual cue RT distributions for data from the second half of the test quite well in all experiments. As such, the failure of the parallel models to fit those data cannot easily be dismissed as reflecting outlier effects in the dual cue data drawn from different populations than assumed by those models (i.e., possibilities such as subjects being confused or hyperconservative on some dual cue trials, etc). Such events may well occur sometimes on the first few test blocks, but would not be expected toward the end of test. One certainly would not expect such factors to result in variance and skew that repeatedly matched the predictions of the RS, ES or ES + C models.

Nevertheless, there are several candidate elaborations of the LC model that might provide better fits to the data from the beginning of the test. One could assume that, instead of retrieval being self-terminating, it is exhaustive on some trials (i.e., both retrievals are completed before a response is executed). This mixture of self-terminating and exhaustive trials might generate the increased skew needed to fit the data from the beginning of the test phases, because RTs on exhaustive trials would be equal to or slower than those of the inefficient cue of each cue pair. However, subjects were instructed to optimize performance in the current experiments, so it seems counter-intuitive that they would adopt this strategy. Another approach to increasing the skew in the LC fits would be to use a non-linear transformation function to convert single cue RTs into predicted dual cue RTs. Again, however, in neither of these scenarios would one expect dual cue RTs to match the ES and ES + C predictions so well during roughly the second half of the test in all five experiments.

A third approach would be to construct a parallel model in which the processes and predictions of the cue selection models constitute special cases. Introduction of a strategic capacity allocation process could accomplish this goal. When all needed capacity is always allocated to the more efficient cue of each pair, the RT predictions could approximate those of the ES model. In fact, if there is just enough capacity for one retrieval, or if retrieval through the inefficient cue can be actively inhibited, then such a model is isomorphic with the ES account. If, on the other hand, all available capacity is allocated to one cue on each trial, and that allocation is randomly determined, the LC model would be isomorphic with the RS model. Given the fit quality of the highly constrained cue selection models, however, there is for now little reason to consider this more flexible class of parallel model. For a parallel model to be preferred on scientific grounds, it should not only provide an equally good account of the current data with a comparable number of free parameters, but must also make a correct prediction that cannot be accommodated within the cue selection framework.

9.3. Comments on parameter-free modeling

The parameter free (or parameter minimizing for the ES + C, LC, and LC + C models) approach used in this paper has some important advantages, along with some disadvantages, in comparison to the more traditional parameterized modeling approach. The relative value of these approaches may well depend crucially on context. In the current case, the task domain lends itself to models with few or no parameters that capture theoretically important issues at a quite general level. Hence, they facilitate evaluation not just of specific process models but also entire classes of models. The ES prediction, for example, establishes a theoretically crucial lower bound performance level for cue selection models as a class.

In the alternative approach of full parametric modeling, both the single and dual cue data must be modeled, without use of the single cue data as a constraint. As such, one must be concerned not only with the quality of the fit to the dual cue data, but also with the quality of fit to the single cue data. If one's goal is more ambitious than ours here, and includes modeling of the properties of both single and dual retrieval, then a fully parameterized model may be best suited. However, our current interest is restricted to the question of how concurrent access to memory from two cues might be limited. Hence, use of single cue data as a known quantity (having only random error in its sample means), and building of theoretical predictions based on that known quantity, is a potentially much more efficient and powerful approach. Any systemic problems with such a model's fit must be attributed to the inability of that model to account for the manner in which the brain manages *dual cue* retrieval.

An example of the inferential power provided by our approach is the RT distribution fitting. It is well documented that quite large amounts of data are needed before an empirical distribution will faithfully represent its underlying population function and parameter values (e.g., Van Zandt, 2001). In any fully parametric model this fact must be heeded, making it difficult in many cases to conclusively test a theoretical distribution form. Our approach, in contrast, is completely non-parametric with respect to distribution form. We did not seek to test any parametric distribution model. Rather, we sought to compare two empirical distributions for each model, one being the dual cue data and the other being the model prediction derived from the single cue data. In this case, the potential presence of bias in the sample cumulative distribution shapes, relative to their true population shapes, is not relevant. Under the null hypothesis for each model, whatever shape bias is present will be statistically identical for the observed and predicted dual cue RT distributions. It is only the relative shape of the distributions being compared (e.g., whether they crossover) that is relevant here.

Despite these favorable attributes, our approach does have potential pitfalls. Most importantly, it requires that any estimated correction factors must be close enough to their true values that they will not lead the researcher to an incorrect conclusion. In the current case, $\mu_{\Delta p}$ was of primary concern. We dealt with this issue by conducting an experiment designed to estimate its value. Though there is no guarantee that that estimate is exact, the likelihood that it is significantly in error, given the similar estimates drawn from other studies, seems quite low. Note also that, in any fully parametric model, the dual cue perceptual delay would still need to be dealt with. It could be treated as a free parameter, but boundary conditions on parameters are preferred when possible. Thus, our empirically based estimate of $\mu_{\Delta p}$ should be just as useful in fully parameterized models as in our models.

Several other corrections, however, appear to be idiosyncratic to our approach. These include the ES correction derived through simulation (Appendix B), along with three other bias factors in the LC and chunking models that were discussed in footnotes. Fortunately, the strong statistical basis for the ES correction, along with the small magnitude of the potential bias relative to the main effect sizes for the race, LC, and chunking models, made these correction issues generally non-problematic. However, this may not be the case when parameter-free models are applied in other experimental paradigms. In any situation, the vastly decreased model flexibility gained through our parameter-free approach (along with its tendency, in our view, to focus the researcher on broad theoretical principles), should be weighed against potential biases built into its assumptions and correction estimates. In the end, we believe that the strongest conclusions will be achieved through converging evidence from both modeling traditions.

9.4. Reconciling our results with coactivation and cross-talk results

Although the cue selection models provide a good account of the data from these experiments, it appears to be at odds with the coactivaton and cross talk results reported in other papers (e.g., Hommel, 1998; Logan & Schulkind, 2000; Miller, 1982). In an initial effort to integrate these findings, we suggest that a set-cue bottleneck, in a generalized form, may underlie performance in all of these cases.

First consider the coactivation results observed in go, no-go tasks. In this case, performance is presumably driven by a rule held active in working memory: "If the target (e.g., an A) is detected, make a key press response." We propose that this single activated rule constitutes the task set. If a target is present, then the rule condition is met, and in essence a set-cue conjunction is formed at the "condition side" of this rule in working memory. When two targets are present (e.g., two A's), there is still only one active set-cue conjunction, but in this case the two cues both match the condition side of the rule (a type of chunking effect), presumably facilitating its execution, and in turn reducing response latency. Thus, facilitation is observed without violation of the principle that only one set-cue conjunction at-a-time can mediate performance. The relative facilitation in the identical and response compatible noise conditions in Eriksen and Eriksen's (1974) flanker task can potentially be interpreted within the same framework. There, the condition side of the rule specifies two potential targets, and the rule fires faster if both are present.

Consider next the cross-talk experiments of Hommel (1998). In those experiments, subjects were presented with a letter and a color cue, combined as a color letter. Subjects were instructed that one of the letter and one of the color cues matched the abstract rule "left," and that the other member of each stimulus category matched the abstract rule "right." They were required to first make a left or right key press response for the color cue, and then a German equivalent of a "left" or "right" vocal response for the letter cue. It is quite plausible that subjects dealt with this task by categorizing the stimuli according to their output requirements. For example, for a given subject, the "H" and "green" cues might be grouped as requiring a "left" response, whereas the letter "S" and "red" cues might be grouped as requiring a "right" response. The plausibility of this stimulus grouping strategy derives from the fact that it reduces the working memory load from four simple rules (e.g., H $\rightarrow$ right, S $\rightarrow$ left, Green $\rightarrow$ right, Red $\rightarrow$ left) to two only slightly more complex rules (e.g., if H or Green $\rightarrow$ right, if S or Red $\rightarrow$ left). By direct analogy to the argument for coactivation phenomena, the active rule(s) constitute the task set, kept dynamically active in working memory. On incompatible trials, the two stimuli activate different rules. The two rules, requiring different responses, receive one dose of activation each. As a result, neither rule may fire as fast as it otherwise could, and competition effects may slow rule selection for the first response. In contrast, on compatible trials, both cues match only one of the two candidate rules on each trial. That rule gets two doses of activation, potentially facilitating RT. Further, there is no other activated rule competing to be selected, and thus no interference effect. Following modern central bottleneck theory (for a review see Pashler, 1997), the system must then reset and initiate second task performance. Relative facilitation occurs for compatible trials on both tasks 1 and 2 even though only one set-cue conjunction (i.e., one conjunctive rule condition) is mediating performance. The important point here is that the cross-talk effects need not reflect independent, parallel flow of activation from each stimulus directly to its response.

Cross-talk has also been observed by Logan and colleagues. In the Logan and Schulkind (2000) experiments that were described in the introduction, subjects presumably had a task set in working memory directing them to match the first task cue to the appropriate category (in Experiment 1, a letter or number), and then to map that category to the required key press response. This task set served to prime the two category nodes. When the cues were presented, activation from each cue presumably flowed in parallel into its appropriate category representation. The most activated category node won a competition and determined the key press response for the first task. If we interpret a set-cue node broadly in terms of its abstract function, the winning category node constitutes the active set-cue conjunction for the first task, with the alternative set-cue conjunction suppressed. If both cues are members of the same category, their activation combines at the winning set-cue node, facilitating first task RTs relative to incompatible trials. Again, facilitation is observed even though only one set-cue conjunction is mediating task 1 (and task 2) performance. Note that associative independence as we have defined it is violated here and above due to the converging activation from two cues on a single node at the set-cue level (a chunking effect).

In Experiments 3 and 4 of the Logan and Schulkind (2000) study, subjects performed a dual lexical decision task. First task responding was faster if both stimuli were words. Here again, there is the possibility that the word stimuli activate a general pre-existing category for words. If both stimuli are words, two doses of activation make it the word category node, potentially resulting in faster response latency.

In the Logan and Delheimer (2001, Experiment 1) study of cross-talk effects in episodic memory, subjects first studied a list of words, and were later presented with trials consisting of either two words from the list, one word from the list and one distracter word, or two distracter words. Subjects pressed one key if the first word was from the list, and another if it was not. Again, first task responding was faster when both words were from the list than when only the first task stimulus was from the list. If we assume that this study created a category for the list, to which all exemplars were associated, and if we treat that category as a set-cue conjunction (by analogy to the previous examples), then a cross-talk effect of a type consistent with the set-cue model could be observed. Here and above, our claim is not that parallel processing of some sort is not occurring, but rather that parallel processing under the case of associative independence, as we have defined it, may not be occurring.

The discussion above makes the general point that the set-cue model, interpreted broadly, potentially allows for dual cue facilitation in some of the working memory and categorization tasks studied to date. It also allows for dual cue facilitation after cue pairs have been presented together a sufficient number of times, as in the current tasks. Essentially, it says that some type of dual cue chunk which allows for convergence of activation at a single set-cue level node must be present for dual cue facilitation to occur. Cued recall is perhaps the most natural domain in which to test this model, since it is easy to create new, arbitrary, and hence independent associations in the laboratory immediately prior to the dual cue test phase. It may be possible, however, to create the same independence conditions in other tasks, in particular working memory tasks. Research along these lines may be quite useful in the effort to pin down the cognitive architecture underlying dual cue, and more generally dual task, performance.

9.5. Implications for theories of memory retrieval and automatization

Modern theories of the Stroop effect typically assume that the color and word dimensions both contribute in parallel to response activation (e.g., Cohen et al., 1990; Logan, 1980; Phaf et al., 1990; Schooler et al., 1997). These theories should apply to the current tasks, because they have no built-in mechanisms that are motivated solely by unique properties of the Stroop task. The fact that we did not find dual cue facilitation in Experiments 3–5 appears to constitute a challenge for them (for other challenges, see Heathcote et al., 1991; Mewhort et al., 1992; Spieler et al., 1996, 2000).

Wenger (1999) concluded in favor of a parallel model of dual cue recall (model 5). He tested parallel and serial models at a general level, so any of a broad range of specific parallel model could be consistent with his results. As such, the implications of the current work for that work are essentially the same as for the class of parallel models more generally. As noted earlier, Logan's (1988) instance theory and Nosofsky and Palmeri's (1997; see also Palmeri, 1997) ERRW theory do not in their current form make predictions for the case of retrieval from two independent cues. However, the current results appear to set a boundary on the extent to which their parallel processing assumption might be extended beyond that case. It may be possible to integrate their models with the set-cue model introduced here. Such integration would require few if any changes to those models for the case of a single retrieval from a single cue.

The results are generally consistent with Rickard's (1997) CMPL model. As noted earlier, the set-cue conjunction model incorporates two basic principles of that model. If applied directly to the current tasks, the simulation model described in Rickard (1997) would predict zero cue-selection latency, and would make predictions that are equivalent to those of the ES model. It could thus account for the results from about the last half of practice in Experiments 1 and 4 without modification. The addition of the chunking mechanism outlined in the discussion of Experiment 2 (i.e., the ES + C model) should allow that model to account for the facilitation effects toward the end of the test phase in Experiments 1, 3, and 5 as well. The model would then be able to fit the steady-state RT data in all experiments. Modification along these lines remains to be implemented, but at present there is no obvious reason to doubt that it would be possible.

In its current form, however, the CMPL model cannot account for the evidence that cue selection is roughly random on the first few test blocks. It seems likely that it could be modified to accommodate that finding, however. If the strengthening rates of the associative links from the stimulus to the set-cue level were made independent of the strengthening rates for links from the set-cue level to the answer level in the simulation model (Rickard, 1997), then the model should be equivalent to the RS model at the beginning of practice (for averaged data, the same cue would always be selected for each cue pair). A transition to efficient selection with practice might then occur by way of the practice mechanism outlined earlier. The sufficiency of this proposed modification, however, remains to be demonstrated.

Although CMPL was not developed within the sampling-recovery retrieval framework common to many memory models, it is worth noting how it might be integrated with it. Without intending to take a strong theoretical stance on the matter, we suggest that parallel activation at the set-cue level may be understood as a sampling process, whereas flow of activation from the winning set-cue node to the response level node may be understood as a recovery process. CMPL is not competitive with global memory models as a general account of memory, nor is it intended to be. Rather, it provides a promising account of memory access constraints in several task domains involving cued recall, aspects of memory retrieval that will eventually need to be integrated with memory models that have historically focused on other topics.

9.6. Accounting for interference effects within the cue selection framework: A response buffer hypothesis

We suspect that, if a Stroop-type design were applied to our tasks (i.e., if subjects were told to attend to only one dimension, and if the cue pairs were neutral, congruent, or incongruent), interference would be observed in the incongruent condition, just as in the Stroop task. Such interference would seem to contradict cue selection models. After all, in the case of cue independence, subjects cannot know at the time of cue selection whether the cues are congruent or incongruent, so that stage of processing should be equivalent in congruent and incongruent conditions. Latencies of the subsequent retrieval stage do not reflect a race, and thus cannot be affected by congruency status either. Further, if only one cue can be selected at-a-time, why would subjects not always select the target cue, and make their response without reference to the to-be-ignored dimension? Alternatively, if subjects cannot avoid an occasional inadvertent selection of the to-be-ignored cue, the result could only be an error. No interference, as measured by RT, would be expected by the cue selection models as developed so far.[11]

Here we propose one candidate scenario by which an elaborated cue selection model might be able to account for Stroop-like interference effects for cases of associative independence (associative independence seems unlikely to hold in the classic Stroop task, and thus we do not seek to model that task here). The basic assumption is that of a post-retrieval error-monitoring buffer period. Subjects can interrupt or delay overt responding after the retrieval stage is completed, if newly accruing evidence suggests that the first retrieval was executed in error through the non-target set-cue node. The existence of an error checking mechanism is consistent with many people's subjective experience with the Stroop task, as well as with other commonly learned tasks (e.g., single-digit multiplication). There are at least three sources of information that could modulate the latency of a response buffer period. First, at a general task level, subjects could adjust the buffer latency depending on the perceived likelihood that the first retrieved response might be an error, in conjunction with the importance of accuracy in task performance. Thus, the buffer period provides a foundation for task level speed–accuracy trade-off effects.

Other factors might modulate the buffer latency at the trial level. Here, it is assumed that on some trials subjects inadvertently select the non-target set-cue node first. On some of those trials, this might result in an overt error. On other trials, the buffer mechanism might suppress overt responding and instead wait for retrieval through the other cue. Two sub-mechanisms might be involved in this suppression of overt responding. First, the system may be able to detect a mismatch between the target cue given the task-set (e.g., "retrieve through the color cue") and the cue through which the first retrieval is actually being executed. We will term this a set-cue mismatch. If such a mismatch is present, then a signal may be sent to increase buffer latency so that a second retrieval, from the target cue, can be executed in a sequential fashion (thus obeying the set-cue bottleneck). Second, the system may continuously monitor response activation within the set of

[11] In the case of one versus two cues, however, dual cue interference resulting in slowed RTs is possible at the cue selection stage of processing (i.e., if cue selection latency is greater than zero).

viable candidate responses. If a second retrieval is initiated, perhaps automatically, after the first retrieval is completed (but before the response for the first retrieval is executed), and if the second retrieved response is different from the first (i.e., on incongruent trials) then this detector might become activated. The buffer period could then be extended to allow for evaluation of the second candidate response. Hence, RT interference effects could be generated in the incongruent condition by both a set-cue mismatch detector and a response level detector, resulting in slowed RTs. In the congruent condition, the set-cue mismatch detector would trigger if the non-target cue is selected first, extending the buffer period. The response level detector would not trigger, however, since the same response would be activated by both cues. If the two cues have identical or similar RT distributions, then the result will be slowed RTs on this subset of trials, compared to a hypothetical true neutral condition. However, if retrieval through the non-target cue is significantly faster than retrieval from the target cue, as in the Stroop task for example, then whether or not interference would be observed would depend on whether the RT saving due to inadvertent first retrieval through the faster cue is offset by the increased buffer latency resulting from the set-cue mismatch. As such, the buffer model is potentially consistent with findings of facilitation, no facilitation, or even interference, in a congruent compared to neutral condition. This model, while speculative, seems quite plausible as an account of at least one source of interference effects in such tasks, and it has the unique advantage of accommodating the evidence favoring cue selection in the current study.

The error monitoring mechanism proposed above might be related to the interrupt mechanism in the stop signal paradigm (for a review, see Logan, 1994). In that paradigm, subjects are presented with a cue to perform a simple task, and on some trials they are presented with a stop signal prior to making their response. The error monitoring signals proposed above are directly analogous, with the only obvious exception being that they are triggered by endogenous factors.

The discussion above shows that the crucial factor in differentiating between cue selection and parallel models is not with respect to interference, but rather with respect to facilitation. In their purest forms, neither class of models predicts RT interference. In Logan's (1988) instance theory, for example, no RT interference would be expected. Stroop-type interference can only be generated by elaborating on the fundamental instance theory architecture, as in the case of Nosofsky and Palmeri (1997; Palmeri, 1997). Similarly, the cue selection interference account suggested above constitutes an elaboration on the basic cue selection class of models, which of themselves predict no RT interference for Stroop-like tasks. On the other hand, predictions about facilitation go to the heart of each framework. The cue selection framework predicts that the presence of two or more cues cannot facilitate performance beyond the ES boundary at any point on the distribution, provided only that those cues are independent. In contrast, parallel models are consistent with dual cue performance below the ES boundary and they generally predict dual cue RT distribution shapes having far less skew than was observed in the data. On these grounds, the evidence now favors the cue selection class of models.

Appendix A

The following experiment was conducted to estimate $\mu_{\Delta p}$. Either one or two letters were presented on each trial, using a trial-level presentation procedure identical to that of Experiment 1. The task was to determine whether the target letter "R" was present. If an R was present, subjects pressed the "M" key on a standard keyboard. Otherwise, they pressed the "V" key. For one-half of the trials (crossed with the manipulation of number of letters presented) the target was present once, and for the other half it was absent. The non-target letter stimuli were: M, L, C, K, F, W, S, Q, Z, G, O, A, P, B, D, and T. The first 12 of these were stimuli in Experiments 1 and 2, the last four were filler stimuli to increase feature overlap between the target letter and the set of non-target letters. Stimuli having these four letters were always paired together on dual cue target-absent trials and were excluded from the analyses. Note that each feature of the letter R, as defined within the Gibson (1969) system, is present in at least two of the non-target letters, a property that should render a feature-based search strategy ineffective.

Each of 19 subjects performed a total of 640 trials over 5 practice blocks. They were allowed to take a brief rest after each block. Each block contained 128 trials, including 32 target-absent dual cue trials, 32 target-present dual cue trials, 32 target-absent single cue trials, and 32 target-present single cue trials. For each dual cue item, each cue was on the left in 50% of the trials, just as in Experiment 1. Presentation order within each block was random.

In the target-absent conditions, subjects must exhaustively search (i.e., in the two letter case, they must test both letters for the presence of R) to achieve high accuracy. The mean RT slowing in the two-letter target-absent case, relative to the single letter target-absent case, thus provides an estimate of the sum of $\mu_{\Delta p}$ and the time required to perform the detection task for one additional letter. It is probable that the additional detection task in the two-letter case constitutes a non-trivial proportion of any observed RT difference. Thus, this difference can be seen as providing an estimate of the upper bound value for $\mu_{\Delta p}$.

Overall accuracy was 96%. Because each of the three major features of the target letter were present in at least two of the 16 non-target letters, best case accuracy of a feature-based search strategy should not exceed 87.5%. Error trials were eliminated from all analyses below.

Of primary interest were the RT results for the one and two letter target-absent conditions. RT means and standard deviations (*SD*s) were first computed for each of the four cells of the design (crossing target status [present or absent] and number of letters) separately for each block and for each subject. Thus, for each subject, a total of 20 means and *SD*s were computed (four per block). These means and *SD*s were then averaged over other variables, as appropriate, in the following analyses.

The grand mean for each block, averaged over all other variables, decreased from 550 ms on the first block to 491 ms on the fifth block. A one-way within-subjects analysis of variance, performed on the subject-level grand means for each block, confirmed this effect, $F(4,18) = 6.46$, $p > .001$. Subsequent analyses, however, showed that the differences in the RTs and *SD*s for the critical one- vs. two-letter target-absent conditions did not vary systematically over blocks. Thus, data were averaged over blocks in the subsequent analyses.

The grand mean RTs (i.e., averaged over blocks and subjects) for the four cells defined by target status and number of letters were 468 and 514 ms for the single and dual letter target-present conditions, respectively, and 504 and 525 ms for the one and two letter target-absent conditions, respectively. The critical result was the 20.5 difference in mean RTs for the single vs. dual letter cells in the target-absent condition. The median difference was 18.7 ms. The difference in mean *SD*s for the single vs. dual letter cells in the target-absent condition was 2.1 ms, and the median difference was .98. We addressed the statistical significance of these effects by first computing, for each subject, the difference score between the one and two letter target-absent conditions for the means, medians, and *SD*s. These subject-level difference scores were then tested against the null hypothesis of zero, using both one-sample t tests and sign tests on the median. For the RT difference, the 20.5 ms slowing was highly significant, as measured by both the t test, $t(18) = 4.78$, $p < .001$, and the sign test, $M = 7.5$, $p < .001$. For the *SD* difference, however, the effect was not significant, $t(18) = .79$, $p > .2$, and $M = -.5$, $p > .2$. The standard error for the RT difference scores was 4.29 ms, and that for the *SD*s was 7.35 ms.

The mean $\mu_{\Delta p}$ estimate of 20.5 ms is remarkably similar to the roughly 20 ms estimate indicated by the Pashler and Badgio (1985) study. In that study, the task was to search for the largest digit in a multi-element display. By design, that task assures that subjects are not using a visual feature search to find the target. In the current experiment, we created an analogous condition by assuring that subjects could not obtain a high level of accuracy using any known feature strategy. It seems very unlikely that subjects used such a strategy. If they did, it is unclear why the difference scores were not smaller than in the Pashler and Badgio study. An account based on task differences is untenable. If anything, the perceptual requirements of their tasks were greater than those of our tasks (e.g., their tasks required substantial visual search, whereas in ours subjects were always pre-cued about letter location, which never varied, and number perception is surely no more automatized than letter perception). The quite palpable subjective experience of both the authors and a few students who were informally interviewed was that judgments were based on letters rather than on features.

Based on these results, we set the mean of $\mu_{\Delta p}$ to a value of 20 ms in Experiments 1 and 2, which employed dual letter cues. There was no statistical difference in the *SD*'s for the single and dual letter target-absent conditions. As such, we treated $\mu_{\Delta p}$ as a constant rather than as a stochastic component in the distribution fits. Given the fact that the coefficient of variation, σ/μ, is usually less than .2 for simple cognitive tasks, the standard deviation of $\mu_{\Delta p}$ seems unlikely to be more than about 5 ms, and such a small effect would be negligible relative to the RTs and *SD*s observed in the retrieval data.

Appendix B

The method for obtaining the efficient selection (ES) prediction outlined in the introduction presumes that the sample means for each pair of cues always have the same magnitude ordering as do their underlying population means. If this assumption is not violated, then the obtained ES prediction will be unbiased. However, since the sample mean fluctuates around the population mean, there are bound to be some errors in identifying the cue with the faster population mean using that approach. Consider a cue pair in which each member has the same population mean. The best (least variable) estimate of the ES mean for that cue pair is the average of the RTs for the two cues (in this special case, the ES and RS models make identical predictions). However, in the sample data, inevitably one mean will be smaller than the other, sometimes substantially so, even if the population means are the same. If the cue with the smaller sample mean is then selected for the ES prediction, then the ES prediction for that particular item will be too fast. Since the method always selects the cue with the minimum mean, this bias will always have the effect of under-predicting the ES mean, never of over-predicting it.

The way to avoid this biasing is to only eliminate the slower cue for cue pairs in which the difference between the sample means is sufficiently large that the sample cue mean RTs are unlikely to be reversed relative to the population values. However, if this criterion difference is made too large, then the reverse bias will be obtained, leading to an efficient selection RT estimate that is too large. This problem can be framed more precisely in terms of t tests performed on the sample means of each cue pair, collapsing over test blocks. If the value of the obtained t score is close to zero, then there is a good chance that the faster sample mean is faster simply because of random sample variation. In these cases, data from both cues should be included in the efficient selection estimate to avoid biasing that estimate to be too fast. On the other hand, when the absolute value of the t score is large, then there is likely to be a real population difference between the two cues, and removal of the cue with the larger mean is necessary to avoid the ES estimate being biased to be too slow. The key to finding an unbiased ES estimate can thus be stated as a problem of finding the correct t threshold for deciding whether or not to eliminate the slower cue. An absolute threshold of 1.0 is a reasonable candidate, since it is what is expected on average when differences between sample means are due to purely random factors.

To find the correct t threshold value, we executed the following simulation. First, we specified a population distribution for each of 192 simulated single cue items, 12 items for each of 16 simulated subjects. For simplicity we assumed normal distributions for all simulated cues. We set the means and variances of those distributions so that, as a group, they roughly matched the means and variances of the observed single cue RTs (calculated over practice blocks) in Experiment 1.We did this simply to assure that the RTs in the simulated data were close, on average, to the RTs in the actual data. However, there is no reason to suspect that the outcome is dependent on this procedure.

Next these 192 distributions were randomly assigned into 96 pairs. Twenty observations were then randomly drawn from each cue distribution, and two sample t tests were conducted for each cue pair to determine the absolute value of t for that cue pair. We then selected a positive t threshold value (starting at zero and working up in increments of .01), applied it to the absolute value of the t score for each cue pair, and eliminated the slower cue of a pair only if that t value was greater than the t threshold value. On each iteration through t, we computed the ES grand mean prediction for the entire simulated experiment and compared it to the known population ES grand mean (i.e., the mean of the population means of the 96 simulated cues that had a smaller mean than their companion cue). We iterated through this process until the t threshold was found that provided the best match to the known ES value of the population. This entire process was repeated 50 times, yielding a distribution of 50 t threshold values. The mean of that distribution was 1.001, with a standard error of .01. We applied this absolute threshold value of 1.0 in all five experiments.

In all cases the procedure used on the data was identical to that described above, with the exception that a paired t test was performed on the difference scores for each cue pair computed on each test block, instead of a two-sample t test. This test was more appropriate for the experimental data since the factor of test block accounted for significant variance in the RTs. The slower cue of each cue pair with an absolute t value greater than 1.0 was then removed prior to calculation of the ES prediction.

References

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

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332–361.

Colonius, H. (1990). Possibly dependent probability summation of reaction time. *Journal or Mathematical Psychology, 34*, 253–275.

Colonius, H., & Ellermeier, W. (1997). Distribution inequalities for parallel models of reaction time with application to auditory profile analysis. *Journal of Mathematical Psychology, 41*, 19–27.

Colonius, H., & Vorberg, D. (1994). Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology, 38*, 35–58.

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

Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component. *Psychological Research, 1987*, 23–29.

Egeth, H., & Dagenbach, D. (1991). Parallel versus serial processing in visual search: Further evidence of subadditive effects of visual quality. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 551–560.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target response in a non-search task. *Perception & Psychophysics, 16*, 143–149.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Grice, G. R., Canham, L., & Gwynne, J. W. (1984). Absence of a redundant-signals effect in a reaction time task with divided attention. *Perception & Psychophysics, 36*, 565–570.

Hayes, W. L. (1988). *Statistics* (4th ed.). Texas: Holt, Rinehart, & Winston.

Heathcote, A., Popiel, S. J., & Mewhart, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109*, 340–347.

Hommel, B. (1998). Automatic stimulus–response translation in dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1368–1384.

Jones, G. (1976). A fragmentation hypothesis of memory: Cued recall of pictures and of sequential position. *Journal of Experimental Psychology: General, 105*, 277–293.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12*, 523–553.

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

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. (1994). On the ability to inhibit thought and action. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory Processes in Attention, Memory, and Language*. San Diego: Academic Press.

Logan, G. D., & Delheimer, J. A. (2001). Parallel memory retrieval in dual task situations: II. Episodic Memory. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1072–1090.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual task situations. *Psychological Review, 2*, 393–434.

Logan, G. D., & Schulkind, M. D. (2000). Parallel memory retrieval in dual task situations: I. Semantic memory. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1072–1090.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163–203.

Mewhort, D. J. K., Braun, L. G., & Heathcote, A. (1992). Response time distributions and the Stroop task: A test of the Cohen et al. (1990) model. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 782–882.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*, 247–279.

Mordoff, J. T., & Yantis, S. (1993). Divided attention between color and shape: Evidence of coactivation. *Perception & Psychophysics, 4*, 357–366.

More, C. M., & Osman, A. M. (1993). Looking for two targets at the same time: One search or two? *Perception & Psychophysics, 4*, 381–390.

Nino, R., & Rickard, T. C. (2003). Practice effects on two retrievals from a single cue. *Journal of Experimental Psychology: Learning, Memory, and cognition, 29*, 373–388.

Nobel, P. A., & Shiffrin, M. R. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 384–413.

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

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 324–354.

Pashler, H., & Badgio, P. C. (1985). Visual attention and stimulus identification. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 551–560.

Pashler, H. (1997). *The psychology of attention*. Cambridge, MA: MIT Press.

Phaf, R. H., Van Der Heijden, A. H. C., & Hudson, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22*, 273–341.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108, Cognition.

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.

Rickard, T. C. (1999). A CMPL alternative account of practice effects in numerosity judgment tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 532–542.

Rickard, T. C., & Pashler, H. (2003). A bottleneck in retrieval from a single cue. Revised manuscript submitted for publication.

Rohrer, D., Pashler, H., & Etchegaray, J. (1998). When two memories can and cannot be retrieved concurrently? *Memory & Cognition, 26*, 731–739.

Ross, H. B., & Anderson, J. R. (1981). A test of parallel versus serial processing applied to memory retrieval. *Journal of Mathematical Psychology, 24*, 182–233.

Rudy, J. (1974). Stimulus selection in animal conditioning and paired-associate learning: Variations in the associative process. *Journal of Verbal Learning and Verbal Behavior, 134*, 282–296.

Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E., & Meyer, E. D. E. (2001). Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science, 12*, 101–108.

Schweikert, R. (1983). Latent network theory: Scheduling of processes in sentence verification and the Stroop effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 353–383.
Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology, 93*, 78–82.
Schooler, C., Neumann, E., Caplan, L. J., & Roberts, B. R. (1997). A time course analysis of Stroop interference and facilitation: Comparing normal individuals and individuals with schizophrenia. *Journal of Experimental Psychology: General, 126*, 19–36.
Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 461–479.
Spieler, D. H., Balota, D. A., & Faust, M. E. (2000). Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 506–526.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.
Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
Townsend, J. T., & Colonius, H. (1997). Parallel processing response times and the experimental determination of the stopping rule. *Journal of Mathematical Psychology, 41*, 392–397.
Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*, 321–359.
van der Heijden, A. H. C. (1975). Some evidence for a limited capacity parallel self-terminating process in simple visual search tasks. *Acta Psychologia, 39*, 21–45.
Van Zandt, T. (2001). How to fit a response time distribution? *Psychonomic Bulletin & Review, 7*, 424–465.
Wenger, M. J. (1999). On the whats and hows of retrieval in the acquisition of a simple skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1137–1160.