Is Repetition Blindness a Visual Phenomenon? A critical review

Clark Fagot

University of California, San Diego

Abstract

This report reviews the literature on whether people have difficulty seeing repeated items in rapid serial visual presentation displays (RSVP): the effect known as “repetition blindness” (RB). It is argued that the data originally cited in support of this hypothesis are weak, and that follow-up work puts the view in doubt. More recent work by advocates of the hypothesis is also discussed. In particular, RB experiments employing signal detection theory measures of sensitivity are carefully examined. It is concluded that signal detection theory is not appropriate for determining whether people fail to see repeated items in these experiments. Finally, the type of evidence that would be necessary to validate the “blindness” view of RB is discussed.

Introduction

The effect commonly named repetition blindness (RB) has been said to reflect a peculiar difficulty in forming conscious representations of repeated items in brief visual displays (hence repetition “blindness”) (Kanwisher, 1987). There are now several reports which claim to provide contrary evidence, putting this view in doubt (Armstrong & Mewhort, 1995; Fagot & Pashler, 1995; Whittlesea, Dorken, & Podrouzek, 1995; Whittlesea & Podrouzek, 1995;). Downing and Kanwisher (1995), however, recently argued that there is “substantial evidence” to support the original claim and that future research should focus on what the effect tells us about visual cognition rather than on whether it is a “real effect” or not. I argue here that while the effect is undoubtedly real in the sense of being replicable, there is reason to doubt that it tells us anything about visual cognition. Instead, it will be argued, substantial evidence suggests that the RB phenomenon is best explained without reference to visual cognition at all.

Repetition Blindness and the Types and Tokens model

The first published report on RB (Kanwisher, 1987) employed three different tasks, all using rapid serial visual presentation (RSVP). In RSVP visual stimuli are presented sequentially at the same location and at rapid rates (between 100 and 250 ms/item in most RB experiments). If the stimuli were the words “cat”, “album”, and “shoe”, then “cat” would be presented for, say, 125 ms, replaced by “album” for another 125 ms, followed by “shoe”, and finally a pattern mask. In the repetition detection task subjects reported whether any of the words appeared more than once in the list. In the sentence report task the words in the list formed a sentence and subjects reported the entire sentence. Finally, in the near-threshold recognition task subjects reported only the final word in the list. In all of the above cases, some of the word lists contained a single word twice (repetition lists), while other lists contained no repeated words (no-repetition lists). The repeated items in the repetition lists and the items in the corresponding locations in no-repetition lists are called the critical items; they will be denoted C1 and C2 in the order of presentation. Results of the first two tasks led Kanwisher to argue that subjects have difficulty processing items that are repeated. The absence of a repetition deficit on the third task led her to theorize a dissociation between type recognition and token individuation (below these terms are expanded upon). We will consider the findings for each of the three tasks in turn.

Repetition detection task

In the repetition detection task subjects reported whether any item in the display was repeated. Kanwisher found that at a rate of 250 ms/item subjects performed at over 90% accuracy (averaged across repetition and no-repetition displays). When the presentation rate was sped up to 117 ms/item, accuracy fell to 45%.

One explanation for the drop in performance at the higher presentation rate is a deficit in processing repeated stimuli at this rate, as Kanwisher supposes. However, there are obvious alternative explanations, some pointed out by Kanwisher. First, it might be that people do not identify all words at the fast rate. Kanwisher argued against this point on the grounds that people can read RSVP sentences with no trouble at these presentation rates, so the difficulty in detecting repetitions cannot simply be in identifying the words. However, in sentence reading context may prime some words and lead to other words being correctly guessed. In addition, it is not the case that people make no errors reporting sentences at this rate (c.f., Forster, 1970; Potter & Lombardi, 1990). Thus, at least part of the problem with detecting repetitions may be that one or both instances is not identified correctly for reasons quite apart from their status as repeated items. Second, to detect repetitions subjects must do more than identify all the words. They must also store the words to compare them to words presented later in the list, and they must compare each word to all those presented before. At faster presentation rates, one or both of these operations might fail. To determine whether subjects have particular trouble identifying repeated items one would need a control task with storage and comparison demands that are equivalent to the repetition condition. It is unclear what such a task would be.

The only firm conclusion that can be drawn from this experiment, therefore, is that at presentation rates of 250 ms/item subjects are generally able to form conscious representations of repeated items; at the faster rates examined in this study, they may or may not be able to do so.

Sentence report task

In the sentence report task, the words in the display formed a sentence, and subjects reported the sentence aloud. In repetition sentences a word appeared twice. No-repetition sentences were generated by replacing the first occurrence of the repeated item in a repetition sentence with another word that still resulted in a grammatical sentence. A dramatic effect was found: unrepeated C2's were reported on 79% of the trials, whereas repeated C2's were reported on only 22% of the trials. The fact that subjects report C2 in the repeated condition less often than C2 in the unrepeated condition implies that somewhere between perception and report subjects have a special difficulty with repeated words.

At first glance, this result would seem to constitute a compelling demonstration that subjects have difficulty seeing repeated items. If people saw the item, the argument goes, then why should they drop it during report, when doing so makes the sentence ungrammatical? However, this reasoning does not hold up under scrutiny. To start with, it assumes that people process the sentences completely "on-line", i.e. that they understand the sentence as it is presented. Although people do a reasonable job understanding RSVP sentences at these rates (Potter, 1984), there is also evidence that off-line processes play some role in report (e.g., Mitchell, 1979; Potter & Lombardi, 1990). One possibility is that people understand the first several words of a sentence as they are presented, but simply store the later words for later retrieval, not connecting them to sentence context at all.

Furthermore, even if subjects understand the sentence on-line, they may still drop repetitions during retrieval. Why would this happen if the subject fully understood the sentence on-line and knew it was grammatical? The answer is simple: understanding a sentence on-line does not imply also storing the sentence verbatim for later retrieval. It is true that people accurately make grammaticality judgments (Is this a grammatically correct sentence?) and plausibility judgments (Does this sentence make sense?) of RSVP sentences (see Potter, 1984, for a review). However, this does not imply that they have a stable representation of the exact words that were presented to them, and, in fact, there is evidence to the contrary (e.g., Mitchell, 1979; Potter & Lombardi, 1990).

It is also worth noting that when reporting RSVP sentences people have a tendency to "grammatically regularize" the sentence. So, for instance, Kanwisher (1986) found that subjects reported referent pronouns (for example, "it") in place of repeated C2's 17% of the time. Whittlesea et al. (1995) found that when people were presented with sentences such as "The yellow car passed our on the left", a common type of error was to replace the word "our" with "ours". These regularization errors occurred even though subjects were told that the sentences would sometimes have missing words and that they should report exactly what they see. Thus, although it is perhaps natural to assume that subjects simply report what they see, a fair number of post-presentation "corrections" are made.

Finally, there are two potential artifacts that may encourage subjects to drop repeated items. First, the repetition sentences used by Kanwisher (1987) are not necessarily as well-formed as the no-repetition sentences (Fagot & Pashler, 1995; Whittlesea et al., 1995). Compare: "We asked for water although water was unavailable" versus "We asked for wine although water was available" (Kanwisher, 1987). Clearly, the later rolls off the tongue more easily. In particular, subjects could potentially replace the second "water" in the first sentence with "it" just to make the sentence sound better, not because they do not see the word. Second, it is common practice in the RB literature to include "blank" sentences which are derived from repetition sentences by removing C2, making it more likely that subjects would drop an item at the expense of grammaticality (Fagot & Pashler, 1995; Whittlesea et al., 1995).

In short, the sentence report task does not demonstrate that loss of repeated items from report is due to on-line failures, whether or not RSVP sentences are understood on-line. Furthermore, the sentences used in these experiments may have induced a bias against repeated words. Finally, it should be mentioned that the majority of research following up on Kanwisher (1987) has used report of unstructured lists of stimuli (i.e., random word or letter lists). The findings in these experiments correspond to those with the sentence report task: subjects have difficulty reporting repeated items. There is no guarantee in these case either, however, that the difficulty is with the perception of repeated stimuli.

Near-threshold recognition task

In the near-threshold recognition task, subjects viewed an RSVP list of words and reported the final word (which was presented for a shorter interval and masked). In this task Kanwisher (1987) found that repetition did not reduce subjects ability to report the final item, and, in fact, led to better performance. This result led Kanwisher to the distinction between type recognition and token individuation. Type recognition corresponds to identification of an item in the display (that the word "fish" is present), whereas token individuation corresponds to seeing an item as a new and separate instance of that type (the word "fish" is in the third serial position). According to this view, for subjects to accurately report an RSVP sentence with a repeated word or detect a repeated item in an RSVP display, both occurrences of the repeated item must be "individuated". In the near-threshold recognition task, however, there is no requirement to individuate the first instance of the repeated item (C1), so individuation of the second is not hampered[1]. However, the type

[1] Kanwisher (1987) also suggested that subjects might be able to perform the near-threshold recognition task with type information alone, i.e., by not "individuating" even the last item. It is not clear how it could be, however, that subjects are able to read-out the last item in the display without seeing it as a separate instance. Perhaps the theory is that type activation quickly decays so that subjects can invoke a process to extract the most recently activated type (sans the pattern mask). The hypothesis mentioned in the text, that subjects only "individuate" the last item is not without suspicion itself. It would require that

of C1 is recognized, producing the priming effect found in the near-threshold recognition task.

More recent experiments on reporting the final word in an RSVP display have provided conflicting results. Kanwisher and Potter (1990, Experiment 6) found a deficit for the repeated item, while other experiments have found no effect either way (see Park & Kanwisher, 1994, pp. 502, for mention of unpublished data). Park and Kanwisher argue that all of these data can be explained within the types and tokens framework by assuming that two mechanisms operate: priming caused by type activation of C1, and a negative effect caused by the occasional "individuation" of C1. Thus, depending on particulars of the experiment, one or the other of these effects may dominate.

In summary, the near-threshold recognition task provides no support for the hypothesis that people have difficulty forming conscious representations of repeated items. It is consistent with the types and tokens model in the sense that a story can be spun to explain all the data, because all possible outcomes can be predicted by the model. To strongly support the model it would also have to be difficult for other hypotheses to account for the data, which is not the case.

Alternative explanations of the RB effect

Kanwisher (1987) interpreted all of the above data together as evidence for two separate processes involved in the perception of visual stimuli: type recognition and token individuation. Failure of either results in a failure to "see" the stimulus. In particular, type recognition without token individuation leads to priming of the type (activation of the type) but no perception of an item of that type occurring in the display. The support for this interpretation of why repeated items are often not reported from visual displays is weak (for that reason, the support for the types and tokens theory is also weak). It will be useful for later sections to outline several alternative hypotheses here. We start with a general distinction between on-line and off-line explanations of RB. An on-line explanation of RB holds that the difficulty processing repeated items occurs as the stimuli are being presented; an off-line explanation holds that the difficulty occurs after the display, during retrieval and report.

subjects can choose to "individuate" items at a moments notice -- the time between when the subject realizes that the current item is the pattern mask and when the "type" of the previous item is no longer available because it has been over-written by this mask.

1. Potential on-line accounts:

a. Recognition refractoriness. When a repetition is presented, people have trouble recognizing the identity of the item.

b. Individuation failure. Two versions of the individuation failure account can be proposed. The one proposed by Kanwisher, and the only one to receive attention in the RB literature, is that people have particular difficulty individuating repeated items. The second account supposes that individuation sometimes fails for both repeated and unrepeated items. When individuation fails for unrepeated items subjects can use type information to determine that an item of that type occurred in the display. For repeated items, however, type information does not suffice because it only tells the subject that an item of that type occurred, but not how many. It is unclear why this second hypothesis has never been proposed by RB researchers, but it can account for any current data that can be accounted for with the first form of the individuation hypothesis.

c. Storage failure. When a repetition is presented, people are often not able to store it in memory.

2. Potential off-line accounts:

a. Retrieval failure. Repeated items are properly stored in memory, but when a person attempts to retrieve two items with the same name, they are often unsuccessful retrieving the second one.

b. Bias. Fagot and Pashler (1995) suggested two types of bias. A guessing bias occurs when subjects prefer to guess at items not already reported. This strategy would result in selectively raising the chance of reporting unrepeated items. A censorship bias occurs if subjects withhold the report of an item they actually saw, in this case repeated items. Fagot and Pashler argued that a guessing bias cannot explain RB, but that a censorship bias can.

Where does this leave us?

It should be apparent that there are several potential accounts of the data presented so far, and many more possibilities could probably be generated. On what basis has Kanwisher rejected the various possible accounts in favor of the individuation failure hypothesis (1b)?

In Kanwisher's original article, little consideration was given to off-line accounts. The argument against them rests mainly with the observation that subjects drop repeated items from grammatical sentences. However, as discussed above, this is not convincing.

Recognition refractoriness (1a) was ruled out because subjects have no trouble identifying repetitions, and indeed identify them better than unrepeated items, in the near-threshold recognition task. If the difficulty with repetitions in the other two tasks studied by Kanwisher (1987) was clearly due to an on-line difficulty, this would perhaps be a strong argument. However, this is not the case. In addition, it is probably unwise to cite the data from the near-threshold recognition task as support for the individuation failure account over the recognition refractoriness account, or vice versa, since the results have not been consistently replicated. Finally, Kanwisher (1987) considered the storage failure account (1c) to be a special case of the individuation failure account. This is an odd position to hold, however, since seeing an item as a unique instance of a type would seem to be distinct from storing it in memory (and, the former could occur without the later also occurring). In short, the token individuation account does not have strong empirical support, although it is generally consistent with the data outlined in the previous section.

Evidence for No Perceptual Repetition Deficit

We have described data widely cited as showing that people have difficulty seeing items that are repeated in a rapid display. It was argued that this hypothesis is not strongly supported. Until very recently, most RB research simply assumed that the difficulty is in perception and posed specific questions about the phenomenon (e.g., Does it occur in space as well as time? Is token individuation the same as feature integration?). In the past few years, however, several papers have claimed to provide evidence against Kanwisher's interpretation (Armstrong & Mewhort, 1995; Fagot & Pashler, 1995; Whittlesea et al., 1995; Whittlesea & Podrouzek, 1995) and others have tried to counter this evidence or offer more evidence in favor of a perceptual repetition deficit (Hochhaus & Johnston[2], 1996; Park & Kanwisher, 1994). Note that an alternative to Kanwisher's view can take one of several forms. Indeed, several of these were outlined in the preceding section. Not surprisingly, the most basic arguments against the individuation failure account also apply to other on-line accounts. Thus, in what follows, whenever possible I will first present the arguments in the most general way: addressing all on-line accounts. After that, I will take aim at counter-arguments specific to particular positions.

Fagot and Pashler detection task

As was noted above, the repetition detection paradigm is of limited usefulness

[2] These authors advocate the recognition refractoriness hypothesis (1a).

because there is no way to determine why subjects are inaccurate at detecting repetitions at the faster presentation rates. People may fail to see repeated items, as Kanwisher supposes, or, alternatively, some comparison or storage operation may fail. What is needed is a way to compare performance on the repetition detection task to performance on a control task that imposes equivalent storage and comparison demands, but that does not require identification of repeated items.

Fagot and Pashler (1995, Experiment 2 and 3) attempted to solve this problem by simplifying the paradigm to minimize memory and comparison demands. Subjects saw a different set of two target letters at the start of each trial, and tried to detect occurrences of these targets within the RSVP display. Suppose the target set consisted of the letters A and B. Then it might be that the letter A would occur and shortly after that the letter B would occur (AB condition), or B could occur before A (BA condition). On other trials, two A's would occur (AA condition) or two B's would occur (BB condition). In all cases, there were two targets in the display, whether they had the same identity or not. If subjects have trouble seeing repeated items, they should perform much worse in the AA and BB conditions than in the AB and BA conditions.

When subjects merely had to indicate where in the display the items occurred (the items were presented from left to right)[3] (Experiment 2) they correctly located targets 45% of the time for repetition displays compared to 48% of the time for no-repetition displays; the difference was not significant.

One may be able to devise an explanation of this data that is consistent with Kanwisher's theory. For example, one could suppose that subjects are not in a state in which they have to "individuate" the targets, and they simply note the location of where the targets occurred[4]. Although a post-hoc explanation such as this may be able to explain the data, it substantially limits the generality of the original theory. The experiment shows at the very least that people are able to form on-line representations of repeated items under some circumstances.

We tested the above explanation in a follow-up experiment by requiring subjects to report the identity of the target items in addition to the location. If subjects can report the identity of the targets in repetition displays as accurately as they do in no-repetition displays, then it cannot be said that subjects are not "individuating" the targets. Subjects correctly located the targets 43% and 44% of the time for repetition and no-repetition displays, respectively, replicating the previous

[3]Standard RB has been observed with full-report from such "moving" RSVP streams (Kanwisher & Potter, 1989; Fagot & Pashler, 1995).

[4] Such an explanation was suggested to us by Kanwisher. However, it is not clear why people should be able to determine that a target occurred "here" without "individuating" the item.

experiment. They identified the targets on 58% and 65% of the trials for repetition and no-repetition displays, respectively, and both located and identified the targets on 31% and 34% of the trials. None of these effects were significant. The small (non-significant) effect[5] that was found on the identity task was completely explained by an inflated false alarm rate for responding AB (first target A, second target B). Thus, not even the post-hoc version of the types and tokens model can account for people's performance on this task.

Fagot and Pashler Red-Letter Report Task

One criticism that can be leveled at the detection experiments conducted by Fagot and Pashler (1995) is that the perceptual demands of the task are not the same as those in more traditional repetition blindness experiments. In particular, non-targets do not have to be fully identified, but only rejected as not targets. It has been known for some time that detecting a target may interfere with other mental operations more than simply rejecting an item as not a target (Duncan, 1980). The implications for Kanwisher's explanation of the data are debatable. Our interpretation of these experiments is that they indicate a failure of the model Kanwisher proposed because the model has no prior way of explaining why there is a repetition deficit when all items in the display must be reported but no deficit in the simpler experiments carried out by us. However, one could certainly make additional assumptions that would explain our findings. For example, one could assume that trouble detecting repeated items only occurs when the individuation mechanism is "fatigued" by having to individuate items at rates approaching 10 per second, whereas the detection experiments we carried out only require subjects to individuate the 2 target items.

With this in mind we carried out an experiment that equated the perceptual demands of a full-report task and a much simpler report task. In this experiment subjects viewed an RSVP display of 6 letters in which one of them was always red. On some trials, subjects reported all the items in the display and when done with this they reported the red item. On other trials they reported the red item first and then made the full-report. Subjects did not know which of the two tasks they would have to perform first, so by design the perceptual (and storage) demands of the two tasks were the same. When the full-report task was performed first subjects reported both instances of the repeated item on 8% of the trials. By comparison, subjects reported the letters occupying the same location as the repeated items in no-repetition displays on 46% of the

[5] It seems reasonable to call a 7% effect "small" here given that we typically found 40% RB effects in this paper using full-report.

trials. Thus, a large RB effect was found in the full-report task. On the other hand, when the red-letter was reported first, subjects correctly reported it on 56% of the trials when it was a repetition compared to 58% of the trials when it was not. The data from when the tasks were performed second on that trial were similar (6% vs. 42% and 43% vs. 44%, respectively). Thus, when the on-line processing of the stimuli was equated between these two tasks, there was an RB effect in full-report but not in the report of the red item.

One concern is that when subjects do not see the red item they may guess other items they did see. Thus, some of the responses that were recorded as correct on repetition displays could really be "migration" errors, i.e. the red item is in the C2 location but the subject guesses the letter from the C1 location and gets is right anyway since the letters are the same. One way to check for such a possibility is to use no-repetition displays to compute how often this "migration" error occurs. Such "migration" errors occurred on only 4% of the no-repetition trials. We conducted an additional analysis to correct for guessing in the red report task in a different way by assuming that subjects guess in this task as they do in the full-report task. Using this correction we again found no RB effect. Thus, our attempts to detect a bias for repeated items that covers up a problem "individuating" repeated items were unsuccessful. A natural conclusion to make is that there was no such "individuation" difficulty, and that "repetition blindness" in full-report is due to an off-line report strategy that was not involved in the red-letter report task.

<u>Repetition judgments in a highest-digit RSVP task</u>

The red-report task of Fagot and Pashler shares one feature in common with the majority of work on both sides of the RB debate: performance of a task requiring perception of a repeated C2 is compared to performance of a task requiring perception of an unrepeated C2. One potential problem with this type of experiment is that a bias may contribute to repeated or unrepeated performance, thereby either covering up a real deficit with repeated items or producing an effect of repetition when there is no such deficit. Above, this possibility was countered by testing for potential biases. A different approach, adopted by Fagot and Pashler (in preparation) is to see whether people can identify repeated items more often than they should be able to if repetition difficulties in other experiments are due to failure to form on-line representation, as Kanwisher and colleagues suggest. One novel aspect of this approach is that there is no control condition with which the repetition condition is compared.

For this purpose, Fagot and Pashler (in preparation) modified the highest digit task of Pashler and Badgio (1985). Subjects viewed RSVP displays of digits; they first reported which digit was the highest digit, and then reported how many times it occurred (once or twice). The highest digit repeated half the time, and no other digit ever repeated. This task was chosen both because it is surprisingly easy to perform and because to perform it accurately subjects must fully identify all digits in the display (thus, it cannot be argued that subjects are not individuating all items in the display). In Experiment 1 an item 2 or 3 after the first occurrence of the highest digit was always colored red. On repetition lists, this item was always the repeated item (second occurrence of the highest digit). In other words, C2 was always colored red. This made the subjects task much easier, but other experiments have shown that an item being colored red does not disrupt RB as measured in full-report (Fagot & Pashler, 1995, Experiment 4; Kanwisher, 1991, Experiment 5). Digits were presented at 157 ms/item. Although this is slower than many RB experiments, it is still in the range Kanwisher (1987) claims RB to be found. Thus, on-line accounts of RB predict that repeated items will often not be seen under these conditions. In fact, repetition of the highest digit was detected on 96% of the trials on which the highest digit was repeated, compared to only 1% of the trials on which it did not repeat. When the same subjects had to report all the items in the display, however, there was a 24% RB effect.[6]

Even though there is no <u>a priori</u> reason to suppose that coloring the repeated item red should overcome RB, it is possible that this is just what happened in the experiment described above. One possibility is that when people do not need to view items after a repeated item in a display, RB does not occur. If this were true, then the subjects in the above study could use the fact that they do not need to pay attention to the display after the red item. Thus, to extend the above findings we conducted a second experiment. In this experiment there was no red item and the display was sped up to 143 ms/item. Participants otherwise performed the same task. On blocks 3 and 4, subjects averaged 84% detection of repetitions, with only 6% false repetition reports. After the 4th block of the experiment subjects were told that there was a theory which predicted they could not see repeated items but that we though they could. Participants then continued for another 4 blocks. During the final two blocks of the experiment subjects averaged 92% detection of repetitions, with only 10% false repetition reports. Signal detection methods can be used to transform these numbers into unbiased performance on the detection task. In both

---

[6] It should be noted that subjects were not told in the full-report task that when there was a repeated item it would always be red. It turned out that this instruction resulted in many subjects reporting the red item twice on all trials, even when it was not a repeated item.

cases this corresponds to about 90% unbiased performance (90% and 91%, respectively)[7]. Thus, it would seem that people are able to perform this task at about a 90% level of accuracy and that the effect of the mid-experiment instruction is to correct for a pre-existing bias against reporting that there is a repeated item.

What do these results mean? First of all, that subjects can do this repetition task at about 90% accuracy suggests they have little problem seeing repeated items here. If one were to suppose that 100% of the items in the display were perceived correctly, and that subjects always noticed when the highest digit repeated, then it would be possible that there was as much as a 10% RB effect occurring. Yet, it is typical to find 40% effects in full-report (c.f., Kanwisher, 1987). Thus, one would either have to believe that the RB effect that occurs in full-report does not occur in the repetition detection task considered here because, say, it has different perceptual demands, or that most of the typical RB effect has some different, presumably memory-related, cause. In that case it would be unclear why one believed there was any other type of effect. It also does not make sense to claim that the highest digit task shows no RB simply because it has different perceptual demands, since it is very similar to the repetition detection task originally employed by Kanwisher (1987). Thus, the fact that people are quite good at reporting whether the highest digit repeats in an RSVP display is a serious problem for on-line accounts of RB.

<u>Whittlesea, Dorken, and Podrouzek 70% Repetition Detection with 100% C2 ID</u>

Whittlesea et al. (1995, Experiment 3) replicated Kanwisher's near-threshold recognition task, but with a new twist. As in Kanwisher's experiment, Whittlesea et al. had subjects view an RSVP display and report the last word in the display; however, their subjects were also required to report whether this was a repeated item. Although the authors did not point it out, the high levels of accuracy in these repetition judgments is a problem for an on-line account of RB. When each item except the last was displayed for 117 ms and the last for 67 ms (the same as in Kanwisher's version), subjects correctly reported the last item 99% of the time when it was a repetition and 98% of the time when it was not. When the lag -2 item (two items earlier) was the first occurrence of the repeated item subjects reported a repetition on 71% of the trials, and when the lag -4 item was the critical item they reported a repetition on 52% of the trials.[8]

[7] If the data for one subject is removed who performed substantially worse than the others at both the repetition detection task and the highest digit task, then these numbers are 93% and 94%, respectively.

[8] When no items were repeated, they reported repetitions only 3% of the time, so the above are not inflated due to the 50% chance performance level, i.e.,

The difficulty for an on-line account is subtle but not easily evaded: On those trials where subjects detect the repetition, they must have correctly perceived and stored the first instance of the item. Therefore, according to an on-line account, RB should frequently occur on the last item in the display (the repeated item). However, subjects reported this item 99% of the time, so this cannot be true.

Park and Kanwisher (1994) have suggested that items may sometimes be individuated out of order, so that earlier items in the display are affected by the "blindness" rather than later items. The justification for this belief is that the order in which items are presented does not necessarily dictate the order in which they are processed. One might suppose, then, that subjects adopt a special strategy for the identify and detect repetition task, so that they first individuate the crucial item for the task (the last item in the display), and then attempt to individuate earlier items in the list in order to detect repetitions. By this hypothesis, performance on the last item is so high because it is the first item individuated, and subjects are now "only" 71% accurate at detecting a repetition 2 items back and 52% accurate at detecting repetitions 4 items back because of "blindness" to these repeated items as they review the contents of the display.

This explanation does not survive close analysis, however. First of all, there is little reason to suppose that the recognition process can continue long after a mask has been presented (Averbach & Coriell, 1961; Turvey, 1973). Since each item in an RSVP display is masked many times over by subsequent items in the display, it seems farfetched to suppose that items 2-4 items back in an RSVP display are still being analyzed. Given the types and tokens model, this means that the types are determined and remain in a buffer either already attached to a token or waiting to be so. Suppose that when the first instance of the repeated item does occur it is recognized and its type activated, but it is not bound to a token. Later, when the second instance of the item is presented, the type is activated even more, and bound to a token (because it is the last item in the list). According to the argument, then, the subject would have to be able to determine that the item also occurred earlier in the list from its level of activation alone. That is, the subject can accurately (70% hits, 3% false alarms at lag -2) tell the difference between activation from one presentation and two presentations. If people can do so well judging how many times an item occurred based on activation level alone, it is unclear why RB should ever be observed, including in full-report.

subjects are seeing that there is a repetition about 70% and 50% of the time, respectively.

Other Work Critical to RB

There are other findings that bear critically on an on-line theory of RB which are worthy of discussion, but the scope of this review limits us to only a mention of these studies. Most notable is the work by Whittlesea and colleagues (Whittlesea et al, 1995; Whittlesea & Podrouzek, 1995). Whittlesea et al. argue that full-report performance (both with sentences and with unstructured material) can be explained in terms of a tendency to allow repeated items to "over-write" each other in report. In their view, the subject recalls what item occurred in each position in turn. The retrieval for any given position may result in 0, 1, or more items; in the later case some of the items may be the same. The case in which more than one item is retrieved for a position is the interesting case. If there are no duplicates then the subject reports one of these items, and returns the rest to the set of available items for report. If there are duplicates, however, and one of them is selected for report, the matching item is not returned to the set of available items. According to this theory, one should be able to use the no-repetition data to simulate repetition performance; this has not yet been demonstrated to be the case. If this simulation were performed and produced the right pattern of results, it would be a very compelling demonstration of their model, and would argue against the notion that subjects are not seeing the repeated items.

Whittlesea and Podrouzek (1995) demonstrated that judgments of whether or not an item repeated in an RSVP display are affected by post-display elaborative processing. So, in their Experiment 2, for example, when people reported whether a display contained a repeated item immediately after the display, they reported repetition for 28% of the repetition lists and 14% of the no-repetition lists. When they first recalled the entire list and then reported whether or not there was a repetition in the list they reported repetitions on 49% and 18% of the repetition and no-repetition lists, respectively. These studies demonstrate that at least part of the reason repeated items are not always reported is due to failures in retrieval.

Armstrong and Mewhort (1995) showed people RSVP lists of letters, and then probed them with one of the letters that was presented, asking them to report the letter which followed the probe letter. People were as accurate when the target letter was a repetition as when it was not. The same was true when the task was performed in mixed trials with a full-report task, and the task for the trial was determined by a cue occurring after the display. Thus, like the Fagot and Pashler (1995) red-letter report task, the on-line processing between this task and a full-report task were equated by the design of the experiment.

Armstrong and Mewhort (1995) tested for biases that may have covered up a repetition deficit. That is, it could be that people do have difficulty seeing repeated items, but in certain circumstances are also more likely to guess them too. For example, Kanwisher suggested to Armstrong and Mewhort that according to the types and tokens model, when an item is not individuated it still sometimes activates the type node for that item a second time. It could be that this increased activation, then, results in the guessing of repeated items more often than unrepeated items whose type nodes are activated only once. However, the investigators tested for this type of bias by comparing how often people guess repeated items from the display when an unrepeated item was the target. They found that repetitions were guessed no more than unrepeated items.

Unfortunately, the authors did not test a different -- but much more plausible -- bias. When subjects do not know which item followed the probe, it makes sense that they would guess other items from the display. In the case of no-repetition targets this would not aid performance compared to if the subject did not guess at all. But when the target is a repeated item, it would. Note, however, that there may be other biases that favor no-repetition targets. For example, if subjects were to guess items they did not see, then no-repetition targets would benefit while repetition targets would not. In short, it is unclear in this study if there is a repetition deficit being covered up by a bias favoring repeated items. However, it should be noted that it would not be surprising for an off-line deficit to occur here: the task is very difficult (25% performance) and may require retrieval of the entire sequence. In that case, off-line theories of RB could easily account for a repetition deficit.

Summary

This review describes a number of phenomena that seem difficult if not impossible to explain as an on-line deficit in processing repeated items. Explaining them with the types and tokens model is either impossible or requires hypothesizing new processes and new limitations to account for each experiment. If the data do not establish that the types and tokens model is wrong, and that people see repeated items perfectly well, then it certainly warrants great doubt about the theory.

Response to Critiques

Contrary to the views just expressed, Kanwisher and colleagues (Park & Kanwisher, 1994; Downing & Kanwisher, 1995; Hochhaus & Johnston, 1996) have argued that a perceptual locus for RB can be maintained. Their arguments have not reconciled the data described above with perceptual RB. Rather, rebuttals have focused on providing new positive support for such a theory. However, this evidence is much weaker than is commonly claimed, and often not diagnostic at all. Below we critically examine these arguments.

Bias for repetitions covering up a real RB effect

Park and Kanwisher (1994) suggest that some experiments which at first seem inconsistent with the types and tokens model actually are consistent if one assumes that there is a bias favoring the report of repeated items that is covering up a deficit in seeing repeated items. Two types of bias have been suggested. First, it could be that when a single cued item must be reported out of a display of several items, subjects sometimes report the wrong item -- the first instance of the repeated item -- from the display. This is a bias due to a "migration" error. As reported above, Armstrong and Mewhort (1995, Experiments 2 and 3) did not consider this type of bias, so it is possible that those results can be explained in this way. However, it was also noted that this task may require recall of the entire sequence anyway, and thus a repetition deficit would not be surprising to an off-line account. The only other experiment reported above where this type of bias could be at work is the Fagot and Pashler red-letter report task. However, as noted above, supplemental analyses revealed that only a small repetition deficit could be hiding behind a "migration" bias. If one took into account the subject's guessing patterns based on the full-report data, there appears to be little or no overall bias one way or the other (i.e., biases that favor repeated items are balanced out by biases that favor unrepeated items).

The second type of bias that has been suggested concerns what subjects do when they guess. It is possible that subjects guess the item that is in some sense most "activated". Thus, in the types and tokens framework, if the type of a repeated item were recognized but not individuated, the subject would be more likely to guess it because of the higher "type activation" (Park & Kanwisher, 1994). One could test the hypothesis that more highly activated types are more likely to be guessed. What would such a test entail? First, one would want to present to subjects RSVP displays that sometimes contained repeated items, requiring report of a single item from the display. Sometimes this target item would be a repeated item, but on critical trials it would not be. The question would be whether when the target is not a repetition but there is a repeated item in the display, would subjects be more likely to report the repeated item (when they were guessing). It would be useful to have a low overall level of report so that subjects would be guessing much of the time. These specifications for an ideal test exactly describe the experiments of Armstrong and Mewhort (1995). As mentioned above, they found that repeated items were no more likely to be guessed than unrepeated items. If such a bias is not present there, there is little reason to think it would ever occur. Furthermore, for this bias to cover up a repetition deficit, it would have to result in repeated items being guessed a large proportion of the time that subjects were guessing. If this bias were such an effective substitute for "individuation", why would it not be just as effective in full-report?

In summary, although "migration" errors may explain the results of Armstrong and Mewhort (1995), they cannot explain other data discussed in this review. In addition, when the second type of bias is scrutinized it is not a plausible bias, and there is data to suggest that it does not occur in a situation where it would most likely. Finally, no type of bias I am aware of could, even in principle, explain the location report data in the detection experiments of Fagot and Pashler (1995, Experiments 2 and 3), the Whittlesea et al. (1995) observations of subjects correctly reporting the final item in an RSVP display 99% of the time while also accurately reporting whether it was a repetition, or, finally, the repetition detection responses in the highest digit displays.

Memory Load

Park and Kanwisher (1994) claim that RB is unaffected by memory load, and that this is inconsistent with retrieval-based explanations of RB[9]. Both of these claims can be challenged. Start with the second. The argument (though not stated explicitly) appears to be that if a problem is in output, then making output more difficult will increase the size of the problem. Although this sounds reasonable, it is not necessarily true. For one thing, the difficulty with repeated items could occur before or after memory load has its effect, but still be an output problem[10]. Even if memory load and repetition had their effect on the same aspect or phase of retrieval, it is not clear whether the presence of a higher memory load should make RB larger, smaller, or the same size. For example, it could be larger because each memory trace is weaker and thus more vulnerable to whatever causes RB. It may be smaller because there is less room for RB to have an effect on weak memory traces.

Now consider the claim that RB does not interact with memory load. This claim is not as transparent as it at first seems. One might assume it means that memory load has no effect on the raw RB effect, R-U (where R is the proportion of repetition trials on which both critical items are reported and U is the proportion of no-repetition trials on which both critical items are reported). This is not, however, what is meant; RB, defined in this way, decreases as memory load increases.

[9] Specifically, Park and Kanwisher (1994, pp. 511) write: "...RB does not interact with total memory load...[and this is] inconsistent with the hypothesis that RB results from a selective loss of repeated items from memory."

[10] This would be entirely consistent with the censorship bias suggested by Fagot and Pashler (1995).

What Park and Kanwisher mean instead is that what they call the RB index, defined as R/(R+U), is the same across different memory loads. The RB index is not an unreasonable way to normalize an effect across different levels of performance, but Park and Kanwisher wish to assert that it is a preferred measure of how much RB is occurring in a given condition. However, they present no concrete argument for why this is so[11]. Below, I attempt to supply such an argument so that its assumptions can be scrutinized.

[11] What they argue is that RB is like a filter, removing some proportion of repetitions from the processing stream, and so a measure of RB should involve a ratio not a subtraction.

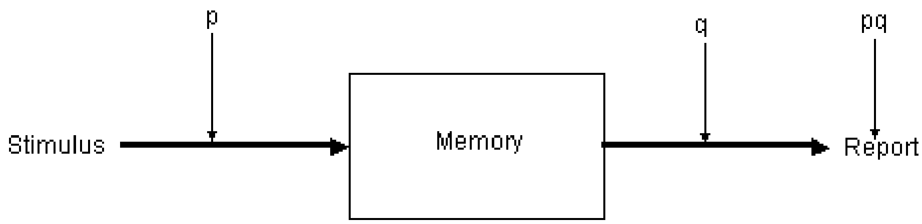


Figure 1: A simple model of full-report: each stimulus is entered into memory with probability p and is later reported with probability q; the overall probability of report is pq.

Figure 1 depicts a simplified model of the processes involved in full-report. Each item is stored in memory with a probability p, and is later retrieved with a probability q, assuming it was stored in the first place. The probability the item is ultimately reported, then, is pq. Now suppose, as implied by a perceptual theory of RB, that repetition affects p and memory load affects q. Let p_R and p_U refer to the probabilities that the item is stored in memory for repetition and no-repetition stimuli, respectively, and let R_L and R_H (U_L and U_H)refer to the probabilities that the item is reported for high and low memory loads in the repetition (no-repetition) condition, respectively. It follows that:

(1) $$R_L/U_L = R_H/U_H = p_R/p_U$$

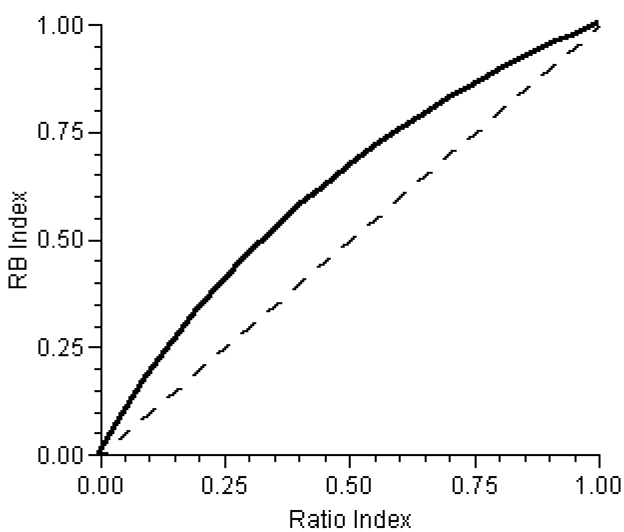
Furthermore, the rightmost expression in equation 1 corresponds to the conditional probability that a repeated item is stored in memory given that it would also have been stored had it been an unrepeated item. Thus, given all the assumptions that entered into the above model, a perceptual theory of RB predicts that memory load should have no effect on the ratio of repeated to unrepeated performance. We will call this the ratio index[12].

The RB index used by Park and Kanwisher is not the same, nor linearly related to, the ratio index derived above. It is, however, monotonically related to it, so a real effect on one of them implies a real effect on the other. Thus, Park and Kanwisher's analysis might seem reasonably justified (to the extent that the assumptions of the above derivation are justified). For statistical reasons, however, the RB index is not a particularly good choice. To see why,

[12] This index is related to a more general strategy suggested by Schweickert (1985), which involves looking for additive patterns of effects on log accuracy. Schweickert viewed this as an extension of the additive factors method of analyzing reaction time (Sternberg, 1969). The discussion in the text of the underlying assumptions of the ratio index borrows from Schweickert.

consider the relationship between the two measures. Figure 2 shows the RB index times two (doubled to be between 0 and 1 like the ratio index) as a function of the ratio index. When the ratio index is 0, .25, .5, .75, and 1.0, the RB index times two is 0, .4, .66, .86, and 1.0. So, when the ratio index is over .5 (as is the case in their study), the RB index used by Park and Kanwisher has the effect of squashing values closer together in a non-linear way, potentially making it harder to detect an effect.

Figure 2: Relationship between the RB index used by Park and Kanwisher (1994) and the ratio index described in the text (solid line plots twice RB index against the ratio index; dashed line is the main diagonal).



What of the assumptions one makes by using the RB (or ratio) index? At the very least these indices rely on two strong claims about full-report: First, there are at least two stages of processing that are applied to each item, where failure in any stage results in failure to report the item. Second, a factor that affects the probability of success for one stage does not affect the output of this stage. This second assumption is akin to assumptions underlying the application of additive factors logic to reaction time experiments with suprathreshold stimuli (Sternberg, 1969). Although there may be grounds to believe the assumption in that case, there seems to be no particular reason to believe it in an unspeeded task such as this[13]. Alternatives abound. For example, increasing memory load may lead to dropping poorly perceived items more than well perceived items, perhaps for strategic reasons. The items dropped may be just those items most susceptible to a bias against repetitions. If perception of an item and its later retrieval were related in this way, then it could be that memory load has no effect on the RB index (or the ratio index), even if RB were an output problem.

These points notwithstanding, one could argue that all theories except the one proposed by Park and Kanwisher would make the lack of an effect of memory load on the RB index (or ratio index) a mere coincidence, while their theory directly predicts it. We

[13] The strongest support for the additive factors method with reaction time data is the detailed patterns of additivity and interaction that are consistent across experiments and can be sensibly interpreted (see, Sanders, 1980). That the related analysis on accuracy data leads to such consistency has not been looked at in the same depth.

need to ask, however: Just how much of a coincidence is their finding? As noted earlier, the use of the RB index reduces the power of any experiment to detect an effect. In Park and Kanwisher's data, for the case where the load followed the repeated item, loads of 1, 2, and 3 produced ratio indices of .65, .73, and .81, respectively[14]. Is a .16 effect sizable? If the use of the ratio (or RB) index has any validity, and if we assume that a perceptual account of RB is correct, then the ratio index corresponds to the proportion of repetitions that make it into memory that would have still made it into memory had they not been repetitions. Thus, this .16 effect on the ratio index corresponds to RB prohibiting 16% more items from making it into memory. Though this may not be statistically significant, a 16% difference here can hardly be called a powerful demonstration of no effect.

In summary, the argument that RB cannot be due to retrieval problems because RB does not increase with memory load misses the mark for three reasons. First, an output explanation of RB need not predict an interaction of RB and memory load. Second, even supposing that it does, the analysis carried out by Park and Kanwisher relies on dubious assumptions. Third, and most important, the data are in fact consistent with the existence of a sizable effect of memory load on the amount of RB as measured by the RB or ratio indices.

<u>RB-like effects</u>

There have been several reports of effects similar to RB in that subjects have difficulty reporting rapidly presented repeated items (Bjork & Murray, 1977; Humphreys Besner, & Quinlan, 1988; Mozer, 1989; Hochhaus & Marohn, 1991). For example, Bjork and Murray (1977) showed subjects displays in which one or two letters were embedded in a 4 X 4 matrix of pound signs (#). The matrix was briefly presented (25-50 ms) and immediately replaced by a matrix of masking characters, with the position of one of the two letters cued for report. The investigators found that subjects were less accurate reporting the target when it was identical to the uncued letter than when it was different.

These studies have been claimed to support RB in two different ways. First, they demonstrate that people have trouble reporting repeated items in a number of different situations. It has further been argued that these effects cannot be due to bias or retrieval difficulties, and therefore the same can be said about RB. Second, in many cases these studies do not require subjects to report both of the repetitions. Therefore, it has been argued, they show that one need not retrieve

[14] These numbers were computed using the mean repeated and unrepeated scores from Park and Kanwisher's (1994) Table 12. A more appropriate analysis would be to compute the ratio index for each subject separately.

both instances of the repeated item for the difficulty to occur, implying that the problem cannot be a bias against repeated items (Park & Kanwisher, 1994).

The first argument depends, of course, on the assertion that these effects are not due to a bias or retrieval difficulty themselves. Arguments to this effect have been made, but they are not strong arguments. For example, Santee and Egeth (1980) used the paradigm of Bjork and Murray (1977). Subjects were told that the target letter would be either an A or an E. On some trials, however, the cued letter was actually A and E both superimposed along with the masking character (#). On these trials subjects reported the uncued letter as much as the potential target that was not in the display, i.e., they guessed repetition as much as no-repetition. This experiment argues against certain bias explanations -- in particular, the mechanism Fagot and Pashler (1995) termed a guessing bias. However, it does not address other biases, including the bias that Fagot and Pashler termed a censorship bias.

What of the fact that subjects need not retrieve both instances of repeated items for a bias to occur? This argues against the problem being one of output interference, but does not address the issue of a bias in any way.

Signal Detection Theory and Repetition Blindness

Some investigators have argued against bias explanations of RB based on measures derived from signal detection theory (SDT). These measures are functions of two parameters: hit rate (H), the proportion of trials on which a "signal" is present and the subject responds that it is present, and false alarm rate (F), the proportion of trials with no signal (called "noise" trials) on which the subject responds that a signal is present. For example, Park and Kanwisher (1995) had subjects report the number of vowels (1 or 2) in RSVP displays of letters. They compared sensitivity for detecting that two of the same vowel were present (repetition lists) to sensitivity for detecting that two different vowels were present (no-repetition lists). A hit occurred when two vowels were presented and the subject responded "2". A false alarm occurred when only one vowel was present and the subject responded "2". Since the experimenters had no way to determine whether a given false alarm was a false alarm for two identical vowels versus two different vowels, the same false alarm rate was used in both cases. If most of the false alarms were really false alarms for two different vowels (which could have been determined by asking subjects which vowels they thought they saw), the effect on sensitivity (as defined by SDT) could very well disappear. This is not the only problem with this experiment. Hochhaus and

Johnston (1996) pointed out that performance in the task was quite poor, raising the possibility that subjects are not detecting the vowels as the letters are presented, but instead store the sequence and check for vowels only after the display. In this case "sensitivity" differences could reflect effects on memory processes.

Table 1

Critical conditions from Hochhaus and Johnston's (1996) paradigm

	Event Element		
Trial Type	Precue = C1	Target (signal / noise) = C2	Probe
Identity	COIN	COIN / JOIN	COIN
Unrelated	POST	COIN / JOIN	COIN

Hochhaus and Johnston (1996) applied SDT in a different paradigm (after Johnston and Hale, 1984) which would seem to get around some of the problems of the above experiment. These investigators showed subjects a word as a precue (C1) for 250 ms which was then masked for 115 ms. A second word was presented as a target (C2) for 50 ms, and also masked. Finally, a probe word was displayed and subjects indicated whether or not the target (C2) matched the probe. Table 1 shows the crucial conditions employed by Hochhaus and Johnston (1996). In the identity condition C1 and the probe word matched, while in the unrelated condition C1 and the probe word did not match. On signal trials C2 matched the probe word, while on noise trials it differed in exactly one letter (all words were 4 letters long). Hit rates and false alarm rates for repeated C2 detection were derived from the signal and noise identity conditions, respectively. Hit and false alarm rates for no-repetition detection were derived from the signal and noise unrelated conditions, respectively. Across all five experiments, the average hit rate in the no-repetition condition was 86% and the average false alarm rate was 11%. In the repetition condition, the average hit rate was 75% and the average false alarm rate was 24%. Since hit rate goes down and false alarm rate goes up in the repetition condition, a reasonable conclusion would be that subjects are perceptually less sensitive to repeated items in these experiments.

However, there are reasons to doubt that Hochhaus and Johnston's analyses truly demonstrate that people have trouble perceiving repeated items even in the quick displays they used. It must be kept in mind that using SDT measures of sensitivity does not guarantee that what is being measured is perceptual sensitivity. In SDT one hypothesizes a model of the underlying decision process, and then draws inferences based on this model. The truth of any inference therefore depend on the model being correct. Thus, the implications of these results cannot be fully understood without first considering the implicit decision model.

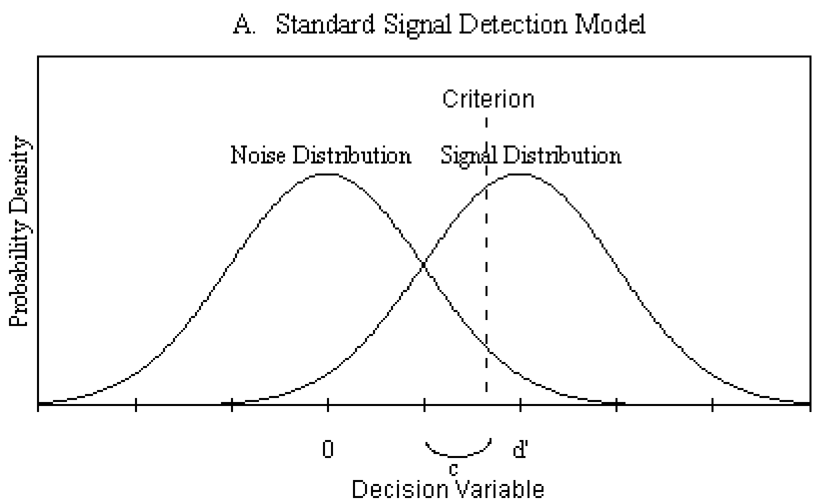


Figure 3: The standard signal detection theory (SDT) decision model (A)

Figure 3A shows the standard SDT model for a detection experiment[15]. This model assumes that there is a one-dimensional decision variable which reflects the judged probability that a signal is present, and the subject decides that a signal is present whenever this variable exceeds some moveable criterion. The distribution of the decision variable on noise and signal trials (the left and right distributions in Figure 3A, respectively) are assumed to be gaussian with equal variance and means separated by d' standard deviations. The placement of the criterion is represented by the variable c: the criterion lies c standard deviations above the mid-point between the two distributions (or -c standard deviations below the mid-point). A hit occurs when a value from the signal distribution exceeds the criterion; a false alarm occurs when a value from the noise distribution exceeds the criterion. It follows that:

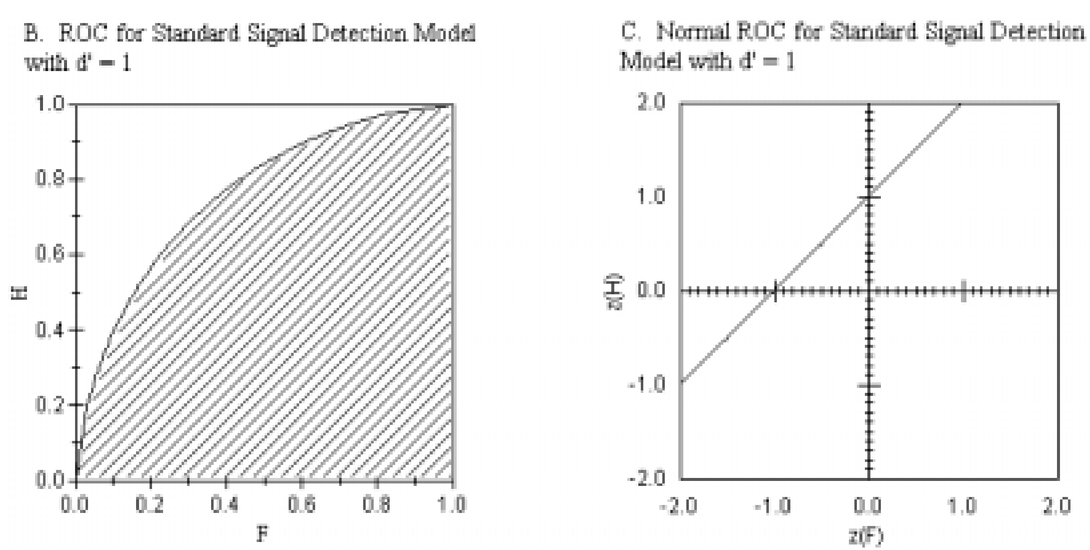
(2) $d' = z(H) - z(F)$, and

(3) $c = -1/2 * (z(H) + z(F))$,

[15] In the next few paragraphs I briefly present several concepts from signal detection theory. For a thorough review, see MacMillan and Creelman (1991).

where z is the inverse of the standard normal distribution. If a manipulation has an effect on the observed d’ then it is said to affect the sensitivity to the signal; if it has an effect on c it is said to affect a bias in responding. This can be generalized to a two-choice discrimination task, where the subject’s job is to report whether stimulus A or stimulus B occurred on each trial; one needs only to regard one stimulus as signal and the other as noise.

Furthermore, one can dispense with the assumption that signal and noise distributions have equal variance if the receiver operating characteristic (ROC) is determined. The ROC represents how hit and false alarm rates vary as different response criteria are employed. The typical way of determining the ROC is to have the subject give a confidence rating with each response. For example, Hochhaus and Johnston (1996) had subjects first make a response, and then give a confidence rating of 1 (low confidence) to 3 (high confidence), for a total of 6 possible ratings. Thus, one criterion would call a response a hit or false alarm only if the subject responded “yes” with confidence 2 or higher. A less stringent criterion would call a response a hit or false alarm if the subjects responded “yes” at all, or “no” with a confidence rating of only 1. Each possible choice of criterion yields different hit and false alarm rates. The ROC is formed by plotting each of these hit and false alarm rate pairs, and connecting the points with a curve. Figure 3B shows the ROC for the standard decision model with d’ = 1. It is often useful to transform the hit and false alarm rates by the inverse of the standard normal distribution before plotting. This is called a normal ROC (“normal” because of the inverse normal transformation that is applied). The normal ROC for the standard decision model with d’ = 1 is shown in Figure 3C. This is a line with slope 1. If the variance of the noise distribution is greater than the signal distribution, then a line with a slope greater than 1 will result. If the variance of the signal distribution is greater, then less than unit slope will result. One consequence of this is that if the variances are unequal, d’ will depend on the particular criterion that is chosen (see equation 2). Thus, d’ is not a robust sensitivity measure when the signal and noise distributions have unequal variance. It turns



out that the area under the ROC curve (the stippled region in Figure 3B) is a measure of sensitivity that does not depend on whether or not the signal and noise distributions have equal variance, and is therefore a more robust sensitivity measure.

Figure 3: The receiver operating characteristic (ROC) predicted by this model (B), and the normal ROC predicted by this model (C) (see text for further explanation).

In a wide variety of experimental paradigms including detection experiments and discrimination experiments employing two stimuli, there is good reason to believe the standard SDT decision model. The strongest support comes from the fact that normal ROC curves tend to be linear like the one in Figure 3C (Swets, 1986b). Other decision models that have been proposed do not predict this (Swets, 1986a; see also, MacMillan & Creelman, 1991, ch. 4). However, the Hochhaus and Johnston experiments are not "simple" detection experiments. First of all, there are more than two potential stimuli that can be presented on any given trial, even though the subject responds only "yes" or "no". It is not clear that the stimulus/decision space is adequately captured with a two distribution model. This will be elaborated upon below. Second, the paradigm employed is more complicated than that of most discrimination experiments where the same two stimuli are used over and over: In Hochhaus and Johnston's experiments subjects are given a precue (C1) which they know is not informative, a second item (C2) is quickly presented, and then a probe word is presented, which on half the trials matches C1. They are likely to be puzzled by what must appear to be a pointlessly complicated procedure. In this context, subjects may well come to confuse inputs and doubt their perceptions. Fagot and Pashler (1995) suggested that a censorship bias may explain the standard RB effect. According to such a bias, subjects may doubt their perceptions of repeated items, and therefore choose not to report them. Something similar may occur here. The question then becomes whether a censorship bias would produce an effect on SDT measures of sensitivity. Neither Park and Kanwisher or Hochhaus and Johnston discussed a consider a censorship bias.

In order to determine the effect of a censorship bias on SDT sensitivity measures, the following Monte Carlo simulation was performed. The standard decision model (Figure 3A) with d' = 3 was used to generate the normal ROC curve labeled "control" in Figure 4A (based on 1000 simulated trials). A second data set was generated by applying a censorship bias to this data. This could have been done in several ways; the method chosen was to "censor" 10% of all values exceeding the third criterion by replacing their value with zero. The normal ROC curve for this data is labeled "censorship bias" in Figure 4A.

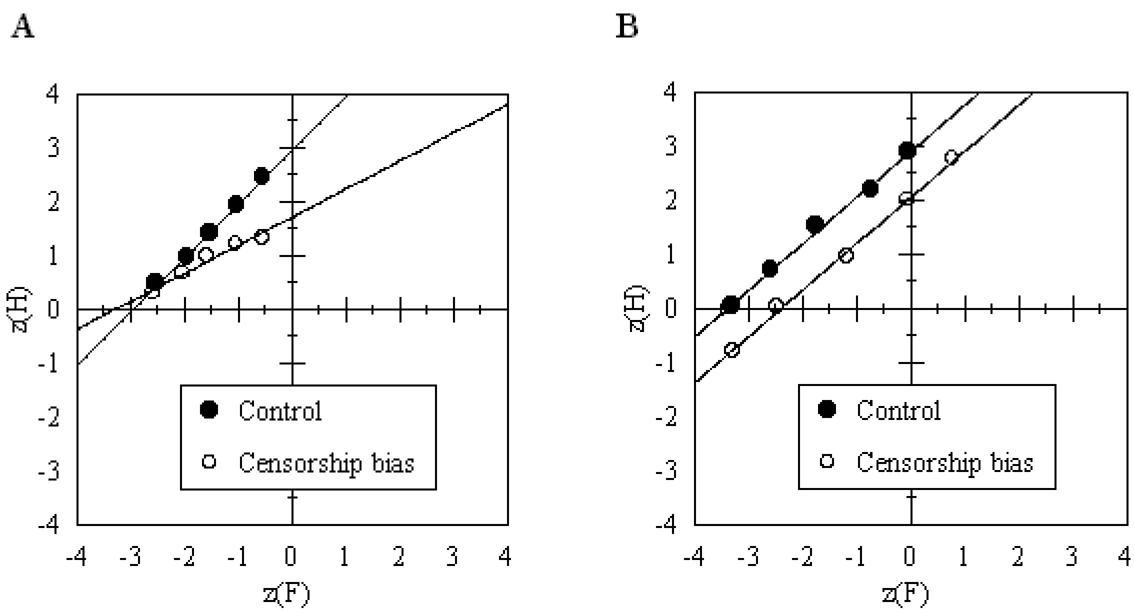


Figure 4: Normal ROC curves generated with two different decision models before (control) and after a censorship bias has been applied (A: standard signal detection theory decision model in which a censorship bias is implemented by "censoring" 10% of the items exceeding a fixed threshold, B: decision model based on the logogen model of word recognition in which a censorship bias is implemented by raising the criterion for the to-be-censored word by .75).

As can be seen, the censorship bias flattens the ROC, hence resulting in an effect on SDT sensitivity measures. In addition, the normal ROC curve is approximately linear, consistent with the standard signal detection model with unequal variance. As it happens, this version of a censorship bias would be inconsistent with parallel normal ROC curves between the repetition and no-repetition conditions. However, Hochhaus and Johnston (and Park and Kanwisher) merely report sensitivity measures; this simulation demonstrates that an effect on sensitivity does not rule out a censorship bias.

The above simulation assumes that the standard decision model is correct. However, this need not be the case. In a typical detection experiment subjects must consider two potential stimulus possibilities (signal and noise). In the current experiment there are many potential stimuli (any four letter word is a potential stimulus in the unrelated condition). It is not clear that a two distribution model suffices. Consider the well-known logogen model of word recognition

(Morton, 1969). According to this model, for every word there is a different unit, and a given word is recognized when the activation of its unit exceeds some criterion. Norris (1995) pointed out that this implies a non-standard decision model. In order to determine the shape of the ROC function for this model and to determine how a censorship bias would effect the ROC, another simulation was performed.

Four logogen units were simulated on each of 1000 trials. One unit corresponded to the presented word and the other three units corresponded to potentially confusable words. So in this simulation it is assumed that only three words are confusable with the presented word on each trial; this assumption could obviously be relaxed. When the stimulus for a given unit was presented, its activation was chosen from a normal distribution with mean 3 and variance 1. When a confusable word was presented its activation was chosen from a normal distribution with mean 0 and variance 1. In the control condition, whenever a unit's activation exceeded a criterion of 1.5, it fired. If only one unit fired on a given trial (the typical case), then its corresponding word was chosen as the response. If more than one unit fired, the one whose activation exceeded the criterion by the most was selected. If none of the units fired a guess was made (yes or no with equal probability). Confidence ratings were based on the amount of activation above the criterion. If the unit exceeded the criterion by 1.5 or more the response was a high confidence response; if it exceeded the criterion by between .75 and 1.5 the response was a medium confidence response; otherwise (including guesses) the response was a low confidence response. A censorship bias was implemented by raising the criterion for the censored word by 0.75.

Figure 4B shows the results of this simulation. First of all, it will be noted that the normal ROC curves are lines with slope 1. Thus, it is not possible to argue from the ROC that this model is incorrect. Second, the censorship bias results in a downward shift of the curve, so an effect on SDT sensitivity measures is found, just as in the previous simulation.

In summary, the RB experiments employing SDT do not challenge the conclusion of this review, that people have no problem detecting or identifying repeated items in RSVP displays. In particular, they cannot rule out the censorship bias suggested by Fagot and Pashler (1995) it predicts an effect on SDT measures of sensitivity. The bottom line is that SDT is not suited to confirming or rejecting a perceptual account of RB, and that the conclusions reached earlier in this review should not be doubted because of existing experiments using SDT measures.

Conclusion and summary

It has been claimed that people have difficulty seeing repeated items in RSVP displays. We have reviewed the original evidence for this claim and concluded that it is weak. We have also presented more recent evidence that suggests people do not have difficulty seeing repeated items. In light of the repetition detection results of Fagot and Pashler (1995), to maintain the token individuation failure hypothesis, one would have to suppose that the difficulty with repeated items only occurs when as many as 10 items a second are individuated. Even if one makes this assumption, the theory cannot explain the red-letter report task of Fagot and Pashler, or the repetition report data of Whittlesea, et al. (1995). Furthermore, it is not clear how the theory could be reconciled with people's high level performance when reporting the number of occurrences of the highest digit from RSVP displays (Fagot and Pashler, in preparation).

Rebuttals to these findings have focused on providing new positive support for a perceptual locus of RB. Foremost among these new studies is Park and Kanwisher's (1994) demonstration that repetition and memory load do not interact on the RB index and experiments showing effects of repetition on SDT measures of sensitivity (Park & Kanwisher, 1994; Hochhaus & Johnston, 1996). However, these experiments employ analyses which depend on dubious models, and in the first case mentioned, a repetition by memory load interaction may well exist. Analyses that depend on a set of unsupported assumptions cannot properly overturn conclusions arrived at using more direct empirical tests, such as those referred to in the previous paragraph. Before a perceptual locus of RB should be accepted (whether it be the token individuation account first suggested by Kanwisher or a recognition refractoriness account such as Hochhaus and Johnston have suggested) the data described in this review must be accounted for. Furthermore, there must be data that are either inconsistent with alternative accounts such as a censorship bias or are more parsimoniously explained with a perceptual locus. To date, neither of these has been done.

References

Armstrong, I.T., & Mewhort, D.J.K (1995). Repetition deficit in rapid-serial-visual-presentation displays: Encoding failure or retrieval failure? Journal of Experimental Psychology: Human Perception and Performance, 21, 1044-1052.

Averbach, E., & Coriell, A.S. (1961). Short-term memory in vision. Bell Systems Technical Journal, 40, 309-328.

Bjork, E.L., & Murray, J.T. (1977). On the nature of input channels in visual processing. Psychological Review, 84, 472-484.

Downing, P., & Kanwisher, N. (1995). Types and tokens unscathed: A reply to Whittlesea, Dorken, and Podrouzek (1995) and Whittlesea and Podrouzek (1995). Journal of Experimental Psychology: Human Perception and Performance, 21, 1698-1702.

Duncan, J. (1980). Selective attention and the organization of visual information. Journal of Experimental Psychology: General, 113, 501-517.

Fagot, C., & Pashler, H. (1995). Repetition blindness: Perception or memory failure? Journal of Experimental Psychology: Human Perception and Performance, 21, 275-292.

Forster, K.I. (1970). Visual perception of rapidly presented word sequences of varying complexity. Perception & Psychophysics, 8, 215-221.

Hochhaus, L., & Johnston, J.C. (1996). Perceptual repetition blindness effects. Journal of Experimental Psychology: Human Perception and Performance, 22, 355-366.

Hochhaus, L., & Marohn, K.M. (1991). Repetition blindness depends on perceptual capture and token individuation failure. Journal of Experimental Psychology: Human Perception and Performance, 17, 422-432.

Humphreys, G.W., Besner, D., & Quinlan, P.T. (1988). Event perception and the word repetition effect. Journal of Experimental Psychology: General, 117, 51-67.

Johnston, J.C., & Hale, B.L (1984). The influence of prior context on word identification: Bias and sensitivity effects. In H. Bouma & D.G. Bouwhuis (Eds.), Attention and performance X: Control of language processes (pp. 243-255). Hillsdale, NJ: Erlbaum.

Kanwisher, N. (1986). Repetition blindness: Type recognition without token individuation. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge.

Kanwisher, N. (1987). Repetition blindness: Type recognition without token individuation. Cognition, 27, 117-143.

Kanwisher, N., & Potter, M. (1989). Repetition blindness: The effects of stimulus modalitiy and spatial displacement. Memory & Cognition, 17, 117-124.

Kanwisher, N., & Potter, M (1990). Repetition blindness: Levels of processing. Journal of Experimental Psychology: Human Perception and Performance, 16, 30-47.

Macmillan, N.A., & Creelman, C.D. (1991). Detection theory: A user's guide. Cambridge, England: Cambridge University Press.

Mitchell, D.C. (1979). The locus of the experimental effects in the rapid serial visual presentation (RSVP) task. Perception & Psychophysics, 25, 143-149.

Morton, J. (1969). Interaction of information in word recognition. Psychological Review, 76, 165-178.

Mozer, M.C. (1989). Types and tokens in visual letter perception. Journal of Experimental Psychology: Human Perception and Performance, 15, 287-303.

Norris, D. (1995). Signal detection theory and modularity: On being sensitive to the power of bias models of semantic priming. Journal of Experimental Psychology: Human Perception and Performance, 21, 935-939.

Park, J., & Kanwisher, N. (1994). Determinants of repetition blindness. Journal of Experimental Psychology: Human Perception and Performance, 20, 500-519.

Potter, M.C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. In D. Kieras & M. Just (Eds.), New methods in reading comprehension research (pp. 91-118). Hillsdale, NJ: Erlbaum.

Potter, M.C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. Journal of Memory and Language, 29, 633-654.

Sanders, A.F. (1980). Stage analysis of reaction processes. In G. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 331-354). Amsterdam: North-Holland.

Santee, J.L., & Egeth H.E. (1980). Selective attention in the speeded classification and comparison of multidimensional stimuli. Perception & Psychophysics, 28, 191-204.

Schweickert, R. (1985). Separable effects of factors on speed and accuracy: memory scanning, lexical decision, and choice tasks. Psychological Bulletin, 97, 530-546.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W.g. Koster (Ed.), Attention and performance II (pp. 276-315). Amsterdam: Elsevier-North Holland.

Swets, J.A. (1986A). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. Psychological Bulletin, 99, 100-117.

Swets, J.A. (1986B). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. Psychological Bulletin, 99, 181-198.

Turvey, M.T. (1973). On peripheral and central processes in vision: Inferences from an information processing analysis of masking with patterned stimuli. Psychological Review, 80, 1-52.

Whittlesea, B.W.A., Dorken, M.D., & Podrouzek, K.W. (1995). Repeated events in rapid lists: Part 1. Encoding and representation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 1670-1688.

Whittlesea, B.W.A., & Podrouzek, K.W. (1995). Repeated events in rapid lists: Part 2. Remembering Repetitions. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 1689-1697.