Title

Binding and Type-Token Problems in Human Vision

Permalink

https://escholarship.org/uc/item/10q8s6pt

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 12(0)

Author

Kanwisher, Nancy G.

Publication Date

1990

Peer reviewed

Binding and Type-Token Problems in Human Vision

Nancy G. Kanwisher
Department of Psychology
University of California, Berkeley
NGK@garnet.berkeley.edu

ABSTRACT

Two computational problems which are trivial for symbol-manipulating systems but which pose serious challenges to connectionist networks are the binding problem and the type-token problem. These difficulties arise because representations in connectionist networks do not automatically i) specify which features go with which object tokens, or ii) distinguish between different tokens of the same type. Nevertheless, these processing shortcomings may constitute advantages when connectionist networks are taken as models of human visual information processing. Perception research shows evidence not only of binding errors, for example in Treisman's illusory conjunctions (Treisman and Schmidt, 1982), but also of type-token errors, as seen in repetition blindness (Kanwisher, 1987) and other phenomena.

INTRODUCTION

There are a number of computational problems which symbol-manipulating systems can handle in a natural and straightforward way, but which pose a special challenge to connectionist networks (Pinker and Prince, 1988; Fodor and Pylyshyn, 1988; Norman, 1986). One of the most notorious of these problems has been variously referred to as the "binding problem" (Hinton, McClelland, and Rumelhart, 1986; Smolensky, 1987), the "indexing problem" (Feldman and Ballard, 1982), and the "tag assignment problem" (Strong and Whitehead, 1989). Roughly speaking, the binding problem refers to the fact that when several things are being represented simultaneously in a network, it is difficult to keep track of what information goes with what. For example, there may be no way to distinguish the visual representation of a red X and a green O from the representation of a red O and a green X.

It is not that the binding problem is unsolvable in connectionist models. Solutions for particular domains have been offered, for example, by Touretzky and Hinton (1986), Smolensky (1987), McClelland (1986), and Mozer (1987). The point is rather that binding is not automatically and naturally accomplished in connectionist models, as it is in symbol-manipulating systems. Thus, a good deal of extra effort is necessary to solve the binding problem in a connectionist model, often at the expense of some of the advantages of parallel processing.

Nevertheless, this processing shortcoming may constitute an advantage when connectionist networks are taken as models of human visual information processing. In particular, there is evidence that the human visual system may suffer from a binding problem of its own. If so, then connectionist networks might provide more accurate models of some stages of human vision than do symbol-manipulating models. Similar points have been made by Norman (1986), Rumelhart and Norman (1982), McClelland (1986), and Mozer (1987).

One example of a human perceptual limitation in binding is the illusory conjunction of visual features described by Treisman and Schmidt (1982). Previously implicated in this context by Hinton and Lang (1985) and Strong and Whitehead (1989), illusory conjunctions are incorrect recombinations of primitive visual features. Treisman and Schmidt showed that when visual attention is diverted from a display containing, for example, a red "X", a blue "T", and a green "O", subjects sometimes confidently reported seeing a green T" or a red "O". This and other findings have led Treisman and Gelade (1980) to hypothesize that early parallel stages of visual processing allow identification of features, but not binding of those features to their locations or to one another. In Treisman's model the binding problem is solved at a subsequent serial stage in which visual attention acts as a gate to allow only one item to be processed at a time (see also Crick, 1984; Koch & Ullman, 1985, and Mozer, 1988; for evidence against temporal binding see McLean, Broadbent, & Broadbent, 1982; Keele, Cohen, Ivry, Liotto, & Yee, 1988).

Because it emphasizes the distinction between parallel processing of primitive features and serial processing of objects and locations, Treisman's model deals most naturally with illusory conjunctions of primitive features in spatial arrays. However, illusory conjunctions also happen for more complex visual categories like words, and affect stimuli displayed in serial lists as well as spatial arrays. For example, Lawrence (1971) showed subjects word lists displayed in rapid serial visual presentation (RSVP). When subjects were asked simply to report the single upper case word in an RSVP list of lower case words, they often reported not the actual capitalized word, but the word that followed it in the sequence. Temporal illusory conjunctions have also been shown for a picture and its frame (Intraub, 1989), and for letters and their colors (McLean, Broadbent, and Broadbent, 1982). Thus, insofar as illusory conjunctions are indicative of a binding problem in visual processing, the problem does not seem to be restricted to the early stages which extract primitive visual features.

A related problem of connectionist networks is the difficulty of representing multiple occurrences of the same category, or multiple "tokens" of the same "type" (Norman, 1986; McClelland, 1986). For spatial arrays, this could be thought of as a subset of the binding problem, since one obvious way of representing the fact that there are two of something would be to bind it to two different locations. This argument has been made by Mozer (1989), who has shown that people often judge a letter array to have fewer items if the array contains repeated letters than if it does not. He argues that the difficulty with repetitions arises because "given the loss of location information, difficulty in detecting repetitions of an object seems certain" (Mozer, 1989, pg. 298). In other words, according to Mozer, the type-token problem can be seen as a *result* of the (more general) binding problem. A similar argument (Keren and Boer, 1985) has been offered to account for the repeated letter inferiority effect (Bjork and Murray, 1977; Santee and Egeth, 1980) in terms of spatial uncertainty.

I will argue here that the human visual system indeed has a serious type-token problem, but that it cannot be accounted for as a simple case of the already-established binding problem. Rather, there seems to be some extra difficulty in encoding two tokens of the same type, over and above the difficulty of binding two nonidentical items to their locations in space or time. To make this point, I will describe some research on a recently-discovered visual phenomenon called "repetition blindness" (Kanwisher, 1987).

Repetition blindness (RB) was first discovered in a pilot observation of Intraub and Potter (Potter, personal communication, 1985), in which subjects were asked to say which picture occurred twice in an RSVP sequence of pictures. At the rapid presentation rates used (about 7 items/second), viewers generally had the impression that they saw most of the pictures but that none of them occurred twice. Further exploration of repetition blindness for words in RSVP (Kanwisher, 1986) showed that the difficulty of detecting repeated words happened even when three words intervened between the two occurrences, and even when the two occurrences differed physically (i.e., one was in upper case and one in lower case).

Repetition blindness was demonstrated most strikingly, however, when repeated words were embedded in sentences and presented in RSVP for immediate verbatim recall (Kanwisher, 1987). When subjects viewed sentences like "When she spilled the ink there was ink all over", the most common response was something like "When she spilled the ink there was all over". In other words, repetition blindness was strong enough to force subjects to selectively omit the second occurrence of the repeated word, sacrificing the meaning and grammaticality of the sentence.

Repetition blindness can not, however, be due to the first occurrence of a repeated word interfering with *recognition* of the second occurrence. When subjects only have to report the last word in an RSVP list performance is helped, not hindered, by an earlier occurrence of the target word in the same list (Kanwisher, 1987). Thus, repetition blindness cannot be a problem in recognizing both occurrences of a repeated word. Rather, what is difficult at rapid presentation rates is individuating the two occurrences as distinct events, or tokens. (Presumably in the name-the-last-word experiment, it was the last word, not the first, which was individuated, or "tokenized", whereas in sentence recall it is typically the first occurrence which is tokenized at the expense of the second.) Thus, RB is a case in which the second occurrence of the repeated word is recognized as a type but not individuated as a new token.

Repetition blindness can thus be taken as evidence of a dissociation between type recognition and token individuation in visual information processing. If this dissociation is a general property of human vision, rather than a peculiarity of the way rapid word lists are processed, then we might expect to also find repetition blindness in very different kinds of visual stimuli. First, the effect ought to happen not only for complex visual types like words, but also for simple visual features like colors and shapes. Second, RB ought to happen for items presented simultaneously in spatial arrays, as well as items presented sequentially in temporal lists. Thus, the present experiments looked for repetition blindness in simultaneously-presented arrays of colors and letters.[1]

EXPERIMENT: REPETITION BLINDNESS FOR VISUAL FEATURES IN SPATIAL ARRAYS

Subjects viewed briefly-presented arrays of four letters (Exp A) or four color patches (Exp B), chosen from a set of six. The subjects' task was simply to report the two array items indicated by two cue boxes surrounding their locations. Two items were cued rather than one because RB is an inability to individuate two tokens of the same type, rather than an inability to simply individuate one of two identical stimulus tokens. The cues either appeared just before the array (the Precue condition), or immediately after a 200-ms mask which followed the array (the Postcue condition). Subjects were told to report the two items in a pre-specified order (top before bottom; if both were on the same row, then left before right).

The key question was whether, after correctly reporting the first item, subjects would be less likely to report the second item correctly if it was the same (the Repeated condition) than if it was different (the Unrepeated condition). To diminish the possibility of a response bias favoring report of Unrepeated items, the instructions strongly emphasized the existence of trials containing two identical targets. In order to avoid floor or ceiling effects, a staircase procedure was used to adjust the stimulus duration periodically to keep performance at about 50% correct on Unrepeated trials. ("Correct" means both items were reported correctly in the correct order.) Separate staircases for Precue and Postcue conditions were used. In each case, the adjustments were made on the basis of performance on Unrepeated trials only, but Repeated and Unrepeated trial durations were yoked together.

Method

Subjects. Twenty-one UC Berkeley students participated, 16 in the letter version (A) and 16 in the color version (B). Five subjects were excluded because their performance on Unrepeated trials averaged 15% or less correct in either the letter or the color experiment, indicating that they were not far above chance even for the longest allowable stimulus duration. (Even for these subjects, Repeated scores averaged lower than Unrepeated scores.)

---

[1] This experiment is reported in more detail in Kanwisher (1990).

Materials, Design, and Procedure. In experiment A, the array items were capital letters drawn from the set E,X,T,O,S, and W. These letters were selected to be as likely as possible to differ in terms of primitive shape features. In Experiment B, the array items were small color patches (composed of two adjacent # signs) drawn from the set red, purple, green, white, yellow, and blue. Experiments A and B were isomorphic (that is, A was translated into B by converting each particular letter into a patch of a particular color). In each experiment, the 96 test trials were broken down into 24 trials in each of the four conditions created by crossing Repeatedness by Cue (Pre versus Post). There were two versions each of Experiments A and B. The design counterbalanced for any effects of particular letters or colors, target locations, or any interaction of these. In addition to the test items, there were 48 filler trials which included repeated colors or letters which were not both probed as target items.

The experiment was carried out on an AST AT computer with a NEC Mutisync II screen, using Psychology Software Tools' MEL experimental software (Schneider, 1988). Each trial began when the subject hit the return key on the computer keyboard. A fixation point appeared for 750 ms in the center of an outline square defining the border of the array. Then in the Precue condition the two small cue boxes appeared inside the outline square for 150 ms, surrounding the location where two of the stimulus items would next appear. Next, the stimulus array composed of four letters or color patches flashed on briefly (one in the center of each quadrant of the large square). The stimulus array was displayed for a variable interval determined by the staircase manipulation. Finally, a mask composed of four rectangular white point arrays (Experiment A) or four rectangular color Mondrians (Experiment B) flashed on for 200 ms covering the locations of the four array items. The Postcue condition was identical except that the cues appeared after the mask, not before the stimulus array.

Subjects were instructed to fixate on the point, look at the array and the cue boxes, and report the two items appearing in the location surrounded by the cue boxes. The subject typed the response into the computer keyboard. This was either the two letters they thought they saw (Exp. A) or the first letters of the names of the two colors they thought they saw (Exp. B). If they had no idea what color or letter was presented in either or both positions, they typed corresponding question marks. Subjects were told three times during the instructions that the experiment contained trials with repeated letters (or colors), and that if they thought both target items were the same they should type the corresponding letter in twice. Before the experimental session, subjects learned the stimulus set by going through 8 trials with feedback which were just like the experimental trials (including both Repeated and Unrepeated trials), except that the stimulus array was displayed for a full second. This served to train them on the color names (and letter set) and to make sure they understood how to respond correctly to Repeated trials. Then they did 24 faster practice trials (without feedback) before the experiment began. Stimulus durations were adjusted periodically throughout the practice test and experimental trials, to keep Unrepeated performance at about 50% correct for both Pre and Post Unrepeated conditions (there were separate staircases for Pre and Post trials).

Results

The results are shown in Table 1. Individual target items were only scored as correct if they were reported in the correct location (indicated by report order). The data were then scored in terms of the conditional probability of getting the second item correct given that the first item was reported correctly. (Performance on the first item alone was fairly constant across conditions--averaging 77% for the letter experiment and 68% for the color experiment--so this technique does not differ much from simply reporting the percent of trials in which subjects got both target items correct.) A correction was used to discount each subject's raw Unrepeated score by the expected number of correct Unrepeated responses due to guessing, based on an analysis of that subject's errors.[2]

	Precues			Postcues		
	Rep.	Unrep.	Duration	Rep.	Unrep.	Duration
Letters (Exp. A)	.41	.60	62 ms	.44	.48	120 ms
Colors (Exp. B)	.24	.49	88 ms	.45	.46	129 ms

Figure 1. The probability of getting the second item correct, given correct report of the first, as a function of Cue (pre vs. post) and whether the two items are the same (Rep.) or different (Unrep.). The Duration column gives the average stimulus duration for that condition. Unrepeated scores have been corrected (downward) for guessing, as described in footnote 2.

Analysis of variance by subjects on the (corrected) probability of getting the second item correct given correct report of the first showed a significant main effect of Repetition, $\underline{F}(1,15) = 10.0$, $\underline{p} < .01$, and significant interactions of Cue x Repetition, $\underline{F}(1,15) = 16.4$, $\underline{p} < .01$, and Cue x Color/Letter, $\underline{F}(1,15) = 8.3$, $\underline{p} < .05$. However, there was no significant interaction of Repetition x Color/Letter, $\underline{F} = 0.1$. No other main effects or interactions reached significance.

Discussion

These data show substantial repetition blindness for both colors and letters, but only in the Precue condition. That is, in the Precue condition, given correct report of the first target item, subjects were significantly less likely to get the second item correct when it was the same as the first (e.g., "XX" or "red red"), compared to when it was different (e.g., "TX" or "blue red"). Thus, repetition blindness--and the type-token problem it exemplifies--generalize to spatial displays of simple visual stimuli.

The fact that RB seems to occur primarily in the Precue condition is a bit mysterious. At first glance, one might have predicted the opposite--i.e. that directing attention to the target locations ahead of time (as in the Precue condition) might have diminished RB. However this didn't happen; RB is evidently robust even when subjects know the target locations before the stimulus appears. As for the lack of RB in the Postcue condition, the most straightforward account would be that the stimulus durations necessary to obtain equivalent performance were longer. Alternatively, Precues might somehow pre-empt grouping processes (available in the Postcue condition) which organize the display into "chunks" according to similarity, making repetitions evident by their emergent orientation or other features. More research is needed to explain the sensitivity of RB to cueing strategy.

---

[2] Corrections were made for Unrepeated Pre and Unrepeated Post conditions separately. For each, the number of incorrect responses in which the subject got the first item correct, but reported a different item from the array in the place of the second item was tallied. This count (N) includes 2/3 of all trials in which the subject guessed the second item from a pool of recognized-but-not-located items, 1/3 of which would be expected to generate correct responses. Thus each raw Unrepeated score, i.e. the number in which both the first and second target were correctly reported in the correct order, was discounted by N/2 before dividing by the total number of trials in which the first item was correctly reported. This correction also takes into account second-target outright guesses of items that were neither recognized nor located, since the number of these which should be discounted (i.e. the number of responses in which the first target is correct, but the second is an item not anywhere in the array) happens to equal the number of these responses which get spuriously included in the location-guess (N) count. Conservatively, no correction was made for the Repeated condition data.

CONCLUSIONS

The available evidence suggests that both repetition blindness and illusory conjunctions reflect general limitations of visual information processing. Both phenomena occur for a broad range of visual stimuli - from simple visual features like colors to complex visual entities like words abstracted across case. Both occur for stimulus items distributed across either space or time, and both can be described (Kahneman and Treisman, 1984; Kanwisher, 1987, 1990) as errors in binding visual types (colors, letters, words) to visual tokens (objects or events). That is, RB can be thought of as an error binding one type to two tokens, whereas illusory conjunctions can be thought of as errors binding two different types to one token. Further, Kanwisher (1990) has argued that the tokens necessary to detect visual repetitions are the same mental entities as the tokens necessary to conjoin visual features. Yet, despite all these similarities, the present experiment demonstrates an important asymmetry in the two phenomena.

While illusory conjunctions can be explained in terms of an overall lack of location information, repetition blindness cannot. In the current experiment, items were only counted as correct when they were reported in the correct location, so both Repeated and Unrepeated trials required subjects to bind item identities to their locations. Thus, a general binding problem or an overall lack of positional information (as implicated by Mozer, 1989 and Keren & Boer, 1985) can not explain the observed difference in performance in the Repeated and Unrepeated conditions. (If anything, a general binding problem would favor the Repeated condition, because switches between the two target locations would not be detected, whereas they would be counted as incorrect in the Unrepeated condition.) Instead, there seems to be a particular difficulty in binding one type to two different tokens, above and beyond the difficulty of binding two different types to two different tokens (or, it would seem, in binding two different types to one token, as in feature conjunction). The early experiments on RSVP word lists demonstrate the analogous situation for temporal tokens: temporal RB was robust even though the serial order of unrepeated words was reported fairly accurately.

Thus, human visual perception seems to suffer from two distinct problems: a binding problem and a type-token problem.3 Both problems are typical of connectionist networks but not of symbol-manipulating systems, which suggests that connectionist networks may be better able to model some stages of human visual information processing. If so, then it is perhaps noteworthy that repetition blindness does not generalize indefinitely. For example, there is no RB at the level of meaning: when two synonyms are embedded in a sentence and presented in RSVP for immediate recall, both are reported with ease (Kanwisher and Potter, 1990).[4] This might suggest an upper bound for the utility of connectionist models of visual processing, somewhere after lexical entries are activated, yet before meanings are retrieved.

ACKNOWLEDGEMENTS

The research reported here was supported by NIMH grant MH45245-02 to the author. I thank Macy Ng for research assistance, and Molly Potter, John Rubin, Michael Jordan, and Michael Mozer for heroic last-minute commentary. Correspondence should be sent to Nancy Kanwisher, Department of Psychology, University of California, Berkeley, Ca 94720.

---

3 An alternative perspective suggested to me by M. Mozer (personal communication, March 1990) is also consistent with the present data and explains the type-token problem as a by-product of the sequential "readout" system which does binding: if an item's identity is suppressed after it is read out, the difficulty with repetitions will result.

4 Also, when sentences containing repeated words are presented auditorily in compressed speech, no "repetition deafness" is found (Kanwisher and Potter, 1989).

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

Fodor, J.A. & Pylyshyn, Z. W. (1988). Connectionism & cognitive architecture: A critical analysis. Cognition, 28, 3-71.

Feldman, J. A. & Ballard, D. H. (1982). Connectionist models and their properties. Cognitive Science, 6, 205-254.

Hinton, G. E. & Lang, K. J. (1985). Shape recognition and illusory conjunctions. Proceedings of the International Joint Conference on Artificial Intelligence, Vancouver, B.C., Canada.

Hinton, G. E., McClelland, J. L., & Rumelhart, D.E. (1986). Distributed Representations. In Parallel distributed processing: Explorations in the microstructure of cognition. Vol 1:Foundations, ed. D. E. Rumelhart, J. L. McClelland, & the PDP Research Group. MIT Press/ Bradford Books.

Intraub, H (1989). Illusory conjunctions of forms, objects, and scenes during rapid serial visual presentation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 98-109.

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

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

Kanwisher, N. (1990). Repetition blindness and illusory conjunctions: Errors in linking visual types with visual tokens. Manuscript submitted for publication.

Kanwisher, N. & Potter, M. (1989). Repetition blindness: The effects of stimulus modality and spatial displacement. Memory and 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.

Kahneman, D & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), Varieties of Attention. New York: Academic Press.

Keele, S.W., Cohen, A., Ivry, R., Liotto, M., and Yee, P. (1988). Tests of a temporal theory of attentional binding. Journal of Experimental Psychology: Human Perception and Performance, 14, 444-452.

Keren, G. & Boer, L. C. (1985). Necessary and sufficient conditions for repeated-letter inferiority: The role of positional uncertainty. In M. I. Posner & O. S. M. Marin (Eds.), Attention and Performance XI, 601-612, Hillsdale, NJ: Erlbaum.

Koch, C & and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology, 4, 219-227.

Lawrence, D. H. (1971). Two studies of visual search for word targets with controlled rates of presentation. Perception and Psychophysics, 10, 85-9.

McLean, J. P., Broadbent, D. E., & Broadbent, M. H. P. (1982). Combining attributes in rapid serial visual presentation tasks. Quarterly Journal of Experimental Psychology, 35A, 171-186.

McClelland, J. L. (1986). The programmable blackboard model of reading. In Parallel distributed processing: Explorations in the microstructure of cognition. Vol 2: Psychological and Biological Models, ed. J. L. McClelland, D. E. Rumelhart, & the PDP Research Group. MIT Press/ Bradford Books.

Mozer, M. (1987). Early parallel processing in reading: A connectionist approach. In M. Coltheart, Attention and Performance XII, Hillsdale, NJ: Lawrence Erlbaum Associates.

Mozer, M. (1988). A connectionist model of selective attention in visual perception. Proceedings of the Tenth Annual Conference of the Cognitive Science Society (pp. 195-201).

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

Norman, D. A. (1986). Reflections on cognition and parallel distributed processing. In Parallel distributed processing: Explorations in the microstructure of cognition. Vol 2: Psychological and Biological Models, ed. J. L. McClelland, D. E. Rumelhart, & the PDP Research Group. MIT Press/ Bradford Books.

Pinker, S. & Prince, A (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. Cognition, 28, 73-193.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. Cognition, 32, 65-97.

Rumelhart, D. E. & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. Cognitive Science, 6, 1-36.

Santee, J. L. & Egeth, H. E. (1980). Interference in letter identification: A test of feature-specific inhibition. Perception and Psychophysics, 27, 321-330.

Schneider, W. (1988). Micro experimental laboratory: An integrated system for IBM PC compatibles. Behavior Research Methods, Instruments, and Computers, 20, 206-217.

Smolensky, P. (1987). A method for connectionist variable binding. Technical Report CU-CS-356-87, Department of Computer Science, University of Colorado at Boulder.

Strong, G. W. & Whitehead, B. A. (1989). A solution to the tag-assignment problem for neural networks. Behavioral and Brain Sciences, 12, 381-433.

Touretzky, D. S. & Hinton, G. E. (1986). A distributed connectionist production system. Technical Report CMU-CS-86-172, Computer Science Department, Carnegie Mellon University.

Treisman, A. & Schmidt, H. (1982). Illusory conjunctions and the perception of objects. Cognitive Psychology, 14: 107-41.

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