Title

A Novel Approach to Understanding Novelty Effects in Memory

Permalink

https://escholarship.org/uc/item/7799x63f

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN

1069-7977

Authors

Lovett, Marsha C.
Nellen, Stefani

Publication Date

2005

A Novel Approach to Understanding Novelty Effects in Memory

Yasuaki Sakamoto (yasu@psy.utexas.edu)
Bradley C. Love (love@psy.utexas.edu)
Department of Psychology, The University of Texas at Austin
Austin, TX 78712 USA

Abstract

Enhanced memory for oddball items has been long established, but the basis for these effects is not well understood. The present work offers a novel way to think about novelty that clarifies the roles of isolation and differentiation in establishing new memories. According to the isolation account, items that are highly dissimilar to other items are better remembered. In contrast, recent category learning studies suggest that oddball items are better remembered because they must be differentiated from other similar items. The current work pits the isolation and differentiation accounts against each other. The results suggest that differentiation, not isolation, leads to more accurate memory for deviant items. In contrast, gains for isolated items are attributable to reduced confusion with other items, as opposed to preferential storage.

Introduction

Vancouver, Toronto, Montréal, Austin. Given a list of items to remember, people show a memory advantage for an item that differs from the others in some way, such as an American city (Austin) in a list of Canadian cities (Vancouver, Toronto, Montréal). This robust memory phenomenon is known as the von Restorff (1933) effect and has been established in various forms. For example, distinctive faces (Valentine, 1991), behaviors (Hastie & Kumar, 1979), and category members (Palmeri & Nosofsky, 1995) are remembered better than typical items.

Items judged as novel tend to be processed more fully and deeply and may in fact be processed differently (Friedman, 1979). For instance, a distinctive member of a group is judged as more influential and more behaviors of the distinctive member are remembered (Taylor, Fiske, Etcoff, & Ruderman, 1978). Moreover, both the positive and negative characteristics of distinctive individuals are perceived as more extreme (McArthur, 1981).

Novelty detection is the flip side of stimulus generalization and thus likely plays a central role in our mental development. Indeed, infants tend to show preference for a novel stimulus once they habituate to a familiar one (Fantz, 1964) and this ability to respond to novelty is predictive of later intelligence (McCall & Carriger, 1993). Given the importance of detecting novelty in our mental activities, recent work in cognitive neuroscience has focused on identifying the neural circuits underlying novelty processing (e.g., Kishiyama, Yonelinas, & Lazzara, 2004; Ranganath & Rainer, 2003).

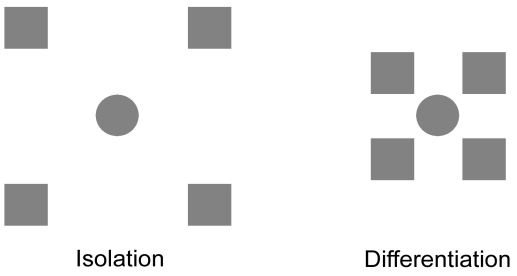

Figure 1: Examples of isolation and differentiation are shown. The left circle is more isolated, whereas the right circle is more differentiated.

Despite the widespread interest in novelty effects, the basis for these effects is not well understood. Most explanations focus on the advantage conferred to items that are relatively distinct or isolated from other stimulus items (e.g., Busey & Tunnicliff, 1999; Valentine, 1991). However, recent category learning research has brought the isolation account into question and has instead suggested that differentiation underlies the memory advantage of oddball items (Sakamoto & Love, 2004). The differentiation account holds that the basis for the oddball advantage is contrasting with highly similar items that establish a context or backdrop.

Figure 1 depicts an oddball item that is either more isolated or differentiated. Inter-item similarity relations play opposing roles in the isolation and differentiation accounts. In the isolation account, items that are highly dissimilar and atypical are best remembered. In the differentiation account, items that are highly similar to other items, yet deviate on a critical property (such as category membership or shape in the case of Figure 1), are best remembered. Previous research has not distinguished these two accounts of novelty effects.

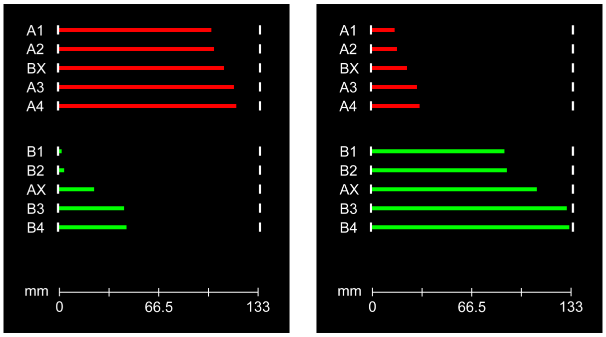

Figure 2: The stimuli used in Experiment 1 are shown. An imperfect rule determined membership in category A or B for all but two items. In this case, red items tend to be in category A, whereas green items tend to be in category B. Each category contains an exception. Item BX is more differentiated than item AX, whereas item AX is more isolated than item BX. To eliminate possible influences of line length on performance (Ono, 1967), subjects were randomly assigned to either the left condition in which the differentiated exception was the longer item or to the right condition in which the isolated exception was the longer item.

In the present work, we evaluate the relative contributions of the isolation and differentiation accounts to enhanced oddball memory. To foreshadow our results, qualitatively different memory advantages are attributable to isolation and differentiation manipulations. Experiment 1 demonstrates that as predicted by the differentiation account, higher-fidelity memory traces result for deviant items that are highly similar to other items. As predicted by the isolation account, deviant items that are dissimilar to other items are easier to identify. Experiment 2 demonstrates that the identification advantage of dissimilar items is attributable to reduced confusion with other items rather than stronger memory traces for those items.

Experiment 1

In Experiment 1, we pit isolation and differentiation against each other in a category learning task. Subjects learned to correctly assign stimuli varying in color (red or green) and length (continuously valued) to one of two contrasting categories through trial by trial classification learning with corrective feedback. As displayed in Figure 2, the membership of all but two items can be correctly determined by applying an imperfect rule (e.g., red items are in category A, whereas green items are in category B). One exception (i.e., oddball) item in each category violates this regularity.

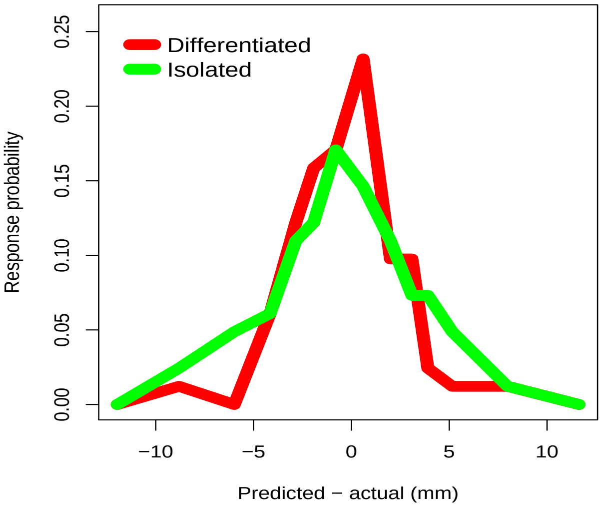

Figure 3: Probability distribution of subjects' responses are shown for the differentiated and isolated exceptions in the reconstruction phase of Experiment 1. The x-axis represents the difference in millimeters between subjects' predicted length and the actual length. Positive values indicate overshoot, whereas negative values indicate undershoot.

The differentiated exception (labeled BX in Figure 2) was highly similar to items belonging to the contrasting category, whereas the isolated exception (labeled AX in Figure 2) was dissimilar to other items. The differentiation account predicts that subjects will develop high-fidelity memory traces for the differentiated exception to reduce confusions with similar items from the opposing category. In contrast, the isolation account predicts that memory should be best for the isolated exception due to its dissimilarity to other items. Memory measures following the learning phase are used to evaluate these accounts of enhanced oddball memory.

Method and Results

Learning phase Eighty-two University of Texas undergraduates learned to classify 10 line stimuli. They completed either 20 blocks of learning trials or two consecutive error-free blocks, whichever occurred first. A block was the presentation of each learning item in a random order. Subjects were provided with the imperfect rule because our main interest was their memory for the exceptions (cf. Palmeri & Nosofsky, 1995; Sakamoto & Love, 2004). As predicted, subjects classified the isolated exception more accurately (.66 vs. .60) than the differentiated one, $t(81) = 2.89$, $p < .01$.

Reconstruction phase Following learning, subjects answered three arithmetic problems to prevent rehearsal of information from the learning phase.

Subjects then reconstructed the lengths of the differentiated and isolated exceptions. The color and category assignment of the exception were given. The exception's initial length was set 66.5 mm, amid the two exceptions' actual lengths. Subjects reconstructed each exception three times, alternating between each on successive trials. Reconstruction error was measured as the absolute difference between the actual and predicted lengths. Consistent with the differentiation account, the mean reconstruction error (averaged across three trials) was significantly smaller (2.1 mm vs. 2.7 mm) for the differentiated than for the isolated exception, $t(81) = 3.37$, $p < .01$. As shown in Figure 3, more responses centered around the actual value (i.e., distance of 0) for the differentiated than for the isolated exception.

Transfer phase Following the reconstruction phase, subjects answered another set of three arithmetic problems. Finally, subjects completed two transfer blocks in which they classified the 10 items presented in the learning phase without corrective feedback. Although subjects reconstructed the differentiated exception more accurately, their transfer classification performance was significantly better (.91 vs. .77) for the isolated than for the differentiated exception, $t(81) = 3.89$, $p < .01$.

Individual analyses In Experiment 1, more accurate memory did not result in better identification. An item is easier to identify when it is more separated from its nearest confusable item. The differentiated exception was not easy to identify because although it was represented more accurately, its representation was still not as separated from the nearest confusable item from the opposing category as the representation of the isolated exception was. The isolated exception was highly dissimilar (19.95 mm to the nearest neighbor on each side), resulting in a larger difference between the reconstructed length and the length of the nearest confusable neighbor than the differentiated exception (6.65 mm to the nearest neighbor on each side).

We advance that identification is largely determined by how separated an item is in a representational space rather than memory accuracy itself. In accord with this idea, there was a significant negative correlation between individual subjects' reconstruction errors and transfer classification accuracies for the differentiated exception, $r(80) = -.27, p < .05$. Smaller reconstruction errors on the differentiated exception, and thus larger differences between the predicted length and the actual length of the nearest neighbor, were associated with better identification of the differentiated item. For the isolated exception, there was no correlation ($r = 0$) between the reconstruction errors and the transfer classification accuracies. The isolated exception was highly dissimilar to its near neighbors, and the difference between the reconstructed length of the isolated exception and the actual length of the nearest confusable item was relatively large regardless of the magnitude of the reconstruction error.

Experiment 2

Experiment 1 demonstrated that items that are highly differentiated from similar items result in more accurate memory, whereas items that are highly separated from other items are easier to identify. In Experiment 2, similarity relations are equated for the isolated and differentiated exceptions as shown in Figure 4. Instead, the differentiated item (BX) is more frequently contrasted with similar items by presenting its adjacent rule-following items from the opposing category (A2 and A3) more often (90% vs. 10%) than its distant rule-following items (A1 and A4) during learning. In contrast, the isolated item (AX) is more frequently contrasted with dissimilar items by presenting its distant rule-following items from the opposing category (B1 and B4) more often (90% vs. 10%) than its adjacent rule-following items (B2 and B3). Other than these changes in similarity relations and presentation frequency of the rule-following items, Experiment 2 is identical to Experiment 1.

As in Experiment 1, the differentiated exception had more opportunities for confusion with members of the opposing category and resulted in a finer-grained memory representation. In the learning phase, 52 University of Texas undergraduates classified the isolated exception more accurately (.51 vs. .44) than the differentiated one, $t(52) = 3.32, p < .01$. Following learning, the differentiated exception resulted in smaller reconstruction errors (2.6 mm vs. 4.2 mm) than the isolated one, $t(52) = 2.62$, $p < .05$. More reconstruction responses centered around the actual value for the differentiated exception as displayed in Figure 5.

Unlike in Experiment 1, the identification advantage for the isolated exception was eliminated when the isolated and differentiated exceptions shared the same inter-item similarity relations. Subjects' transfer classification accuracy without corrective feedback for the isolated (.81) and the differentiated (.80) exceptions showed no significant difference, $t < 1$. Although the isolated exception resulted in fewer classification errors than the differentiated exception during learning, they were both highly confusable with other items as indicated by their low accuracies in the learning phase.

These results suggest that the frequency manipulation in Experiment 2 plays an important role in memory accuracies but has little impact on identification abilities. Identification performance is largely determined by the inter-item similarity relations. As in Experiment 1, there was a significant negative correlation between individual subjects' reconstruction errors and transfer classification for the differentiated exception, $r(50) = -.46, p < .01$. Unlike in Ex-

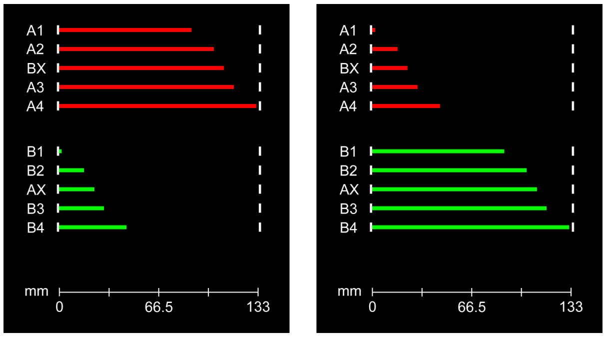

Figure 4: The stimuli used in Experiment 2 are shown. Unlike Experiment 1, the differentiated (BX) and isolated (AX) exceptions share the same inter-item similarity relations. The differentiated exception is more "differentiated" because its near neighbors from the contrasting category (A2 and A3) are presented more frequently than its distant neighbors (A1 and A4) during learning. The isolated exception is more "isolated" because its distant neighbors (B1 and B4) are presented more often than its near neighbors (B2 and B3).

periment 1, the same pattern was found for the isolated exception in Experiment 2, $r(50) = -.63, p < .01$. The difference between the predicted length and the nearest confusable item's actual length in Experiment 2 was similar for both the isolated and differentiated exceptions. This was not the case in Experiment 1 as the isolated item was more dissimilar to other items than was the differentiated item.

General Discussion

Experiments 1 and 2 evaluated two explanations for enhanced oddball memory. The isolation account holds that most dissimilar items should be remembered best. In contrast, the differentiation account holds that items should be remembered best that are similar to other items, yet differ in some critical property that brings the contrast into focus. In most laboratory experiments, both of these explanations are usually operable and their relative influences are indeterminate. In the initial example in this paper, for example, Austin is isolated in that it has unique properties that Vancouver, Toronto, Montréal do not have, but it is also differentiated in that it shares many properties with these other cities, yet varies in the key property of nationhood. Which explanation is operable depends on how one construes the list of city names. To address this conundrum, we pit isolation and differentiation accounts against each other in a classification learning task utilizing rule-plus-exception category structures in which one category's exception was relatively isolated and the other's was relatively differentiated.

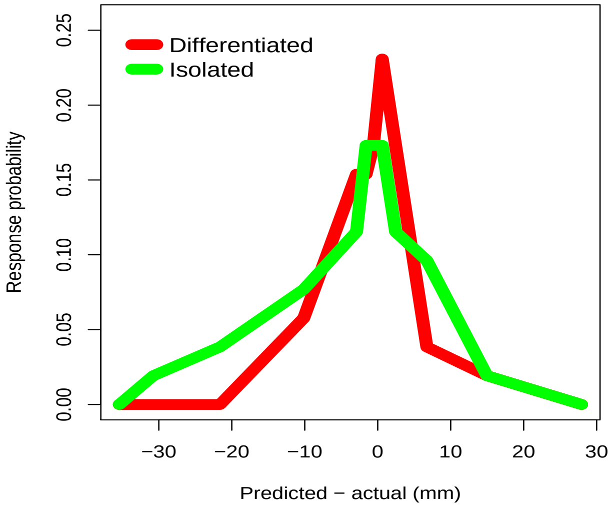

Figure 5: Probability distribution of subjects' responses are shown for the differentiated and isolated exceptions in the reconstruction phase of Experiment 2. The x-axis represents the difference (in mm) between the predicted and actual lengths.

Experiment 1 shows that isolation and differentiation lead to qualitatively different memory enhancements. The differentiated item that was contrasted with highly similar items led to more accurate memory than did the isolated items, whereas the isolated item that was dissimilar to other items was easier to identify (i.e., classify as an exception) than the differentiated item. Experiment 2 shows that the isolation advantage is attributable to reduced confusion with other items rather than preferential storage. The isolation advantage was eliminated when the inter-item similarity relations were equated for the isolated and differentiated items. These different memory enhancements can be predicted by the same mental representations. In Experiments 1 and 2, subjects whose reconstruction of the oddball item was more separated from the nearest confusable item of the opposing category showed better transfer classification performance on that item.

One alternative view is that the number of errors during the learning phase drove these differences as subjects in both experiments made more errors classifying the differentiated than the isolated exception during learning. Indeed, similarity and confusability are the catalysts of differentiation and also beget classification errors. However, errors and differentiation are not synonymous. By manipulating the feedback associated with an item, Sakamoto and Love (2004) dissociated violating a regularity and committing an error during learning, and found that enhanced memory is attributable to structure violation and not errors per se.

Methodological Implications

The present results may help resolve the apparent conflict between studies that do and do not find isolation advantages. The isolation advantage in transfer classification of Experiment 1 can be attributable to the isolated exception being less confusable with members of the opposing category than the differentiated exception is. Thus, the isolation advantage in transfer is likely due to the nature of the other test items, rather than due to a stronger memory trace for the isolated exception.

Analogously, studies that have found an isolation advantage in old/new recognition judgments did not include foils that are similar to isolated items (e.g., Busey & Tunnicliff, 1999; Valentine, 1991). Studies that include foils equally similar to all studied items do not find an isolation advantage (e.g., Davidenko & Ramscar, 2004; though see Nosofsky & Zaki, 2003). In contrast, the advantage for the differentiated exception in Experiments 1 and 2's reconstruction task is not attributable to other items included in the test set and instead indicates that subjects developed finer-grained representations for the differentiated exception.

As demonstrated in the present studies, one determinant of whether an isolation advantage or disadvantage is observed is the nature of the task. For instance, item confusability constrains performance for tasks that yield an isolation advantage, whereas confusability is not harmful or even beneficial for tasks not favoring isolation. Future work that employs multiple memory measures and distinguishes between isolation and differentiation will be necessary to fully resolve these issues.

Theoretical Implications

A number of category learning and memory models utilize novelty detection mechanisms to gate storage (e.g, Nosofsky, Palmeri, & McKinley, 1994; Love, Medin, & Gureckis, 2004). For example, Love et al.'s SUSTAIN clustering model forms new clusters in memory in response to surprising events, such as learning that a bat is a mammal and not a bird. This mechanism allows SUSTAIN to correctly predict enhanced recognition memory for stimuli that violate salient regularities as observed in Palmeri and Nosofsky's (1995). Similarly, Nosofsky et al.'s RULEX hypothesis-testing model correctly predicts enhanced memory for exceptions by explicitly storing items that violate inferred rules.

Sakamoto and Love (2004) modified Palmeri and Nosofsky's design to tease apart the predictions of cluster- and rule-based accounts of category representations and to test the differentiation hypothesis. Sakamoto and Love introduced an asymmetry in the category structures in which one category contained more rule-following items than the contrasting category. The exception violating the more frequent regularity had more opportunities for confusion with members of the opposing category and should lead to a finer-grained memory representation according to the differentiation account. This result held and was predicted by SUSTAIN but could not be accounted for by RULEX. The result seems at odds with rule-based representations of regularities in general (e.g., Pinker, 1991) and instead favors a clustering account that is more schematic in nature.

These results are also inconsistent with exemplar accounts (though see Sakamoto, Matsuka, & Love, 2004). Exemplar models cannot account for the oddball recognition advantage because they store every training instance in memory rather than using novelty gated storage and do not accord special status to oddball items. To determine recognition strength, exemplar models sum the similarity of the probe item to all exemplars stored in memory, which does not predict an advantage for oddball items. Exemplar models do not favor items that are odd by virtue of being isolated because the summed similarity recognition calculation favors typical items. For this reason, correct identification is modeled as the inverse of summed similarity, thus favoring isolated items (Busey & Tunnicliff, 1999). The idea is that the most dissimilar item is least confusable and best remembered. Of course, this account is at odds with the current findings suggesting finer-grained representations for differentiated than for isolated exceptions. The critical problem with exemplar models is that storage is not dependent on what other items are already stored in memory.

Consideration of the current results, coupled with those from Sakamoto and Love (2004), strongly favor non-rule-based representations of regularities or patterns. Factors such as frequency, expectation violation, and similarity to other items are crucial factors driving performance in these tasks, suggesting that storage is gated by novelty and mental representations are cluster- or schema-like and are engaged through similarity-based processing.

Final Note

Novelty effects have been examined in various domains, including the study of schemas/stereotypes (Rojahn & Pettigrew, 1992), basic memory phenomena Hunt & Lamb, 2001, face recognition (Valentine, 1991), the neurobiological basis of memory (Kishiyama et al., 2004), and category learning (Sakamoto & Love, 2004). The relative contributions of isolation and differentiation are not well understood in these studies. The present work offers a novel way to think about novelty that clarifies the roles of isolation and differentiation in establishing new memories.

Acknowledgements

This work was supported by AFOSR grant FA9550-04-1-0226 and NSF CAREER grant 0349101 to B. C. Love.

References

Busey, T. A., & Tunnicliff, J. L. (1999). Accounts of blending, distinctiveness, and typicality in the false recognition of faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1210–1235.

Davidenko, N., & Ramscar, M. (2004). Reexamining the "distinctiveness effect": Poorer recognition of distinctive face silhouettes. In K. D. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (p. 1550). Mahwah, NJ: Lawrence Erlbaum Associates.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*, 668–670.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*, 316–355.

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, *37*, 25–38.

Hunt, R. R., & Lamb, C. A. (2001). What causes the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1359–1366.

Kishiyama, M. M., Yonelinas, A. P., & Lazzara, M. M. (2004). The von Restorff in amnesia: The contribution of the hippocampal system to novelty-related memory enhancements. *Journal of Cognitive Neuroscience*, *16*, 15–23.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of human category learning. *Psychological Review*, *111*, 309–332.

McArthur, L. Z. (1981). What grabs you? The role of attention in impression formation and causal attribution. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium* (Vol. 1, pp. 201–246). Hillsdale, NJ: Erlbaum.

McCall, R. B., & Carriger, M. S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development*, *64*(1), 57–59.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79.

Nosofsky, R. M., & Zaki, S. F. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1194–1209.

Ono, H. (1967). Difference threshold for stimulus length under simultaneous and nonsimultaneous viewing conditions. *Perception & Psychophysics*, *2*, 201–207.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 548–568.

Pinker, S. (1991). Rules of language. *Science*, *253*, 530–535.

Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews: Neuroscience*, *4*, 193–202.

Rojahn, K., & Pettigrew, T. F. (1992). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology*, *31*(2), 81–109.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, *133*, 534–553.

Sakamoto, Y., Matsuka, T., & Love, B. C. (2004). Dimension-wide vs. exemplar-specific attention in category learning and recognition. In M. Lovett, C. Schunn, C. Lebiere, & P. Munro (Eds.), *Proceedings of the 6th International Conference of Cognitive Modeling* (pp. 261–266). Mahwah, NJ: Lawrence Erlbaum Associates.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*, 778–793.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, *43A*, 161–204.

von Restorff, H. (1933). Analyse von vorgängen in spurenfeld. I. Über die wirkung von bereichsbildungen im spurenfeld [Analysis of processes in the memory trace. I. On the effect of group formations on the memory trace]. *Psychologische Forschung*, *18*, 299–342.