

Author



Effect of Picture Similarity on Recognition Memory

Mercedes La Voy

Psychology

Mercedes La Voy has always been interested in studying memory, and has recently developed a specific interest in false memories. These interests led her to the lab of Dr. Steyvers and his work on recognition memory. She says that her research experience—including designing the experiment, counterbalancing the variables, and interpreting the results—have helped her to think critically and to solve problems. In Fall 2005, Mercedes moved on to the next step in her education, entering the Washington State University Clinical Psychology Ph.D. Program. When she's away from her work, Mercedes enjoys going to the beach, cooking, and spending time with friends and family.

Key Terms

- ◆ Eyewitness Testimony
- ◆ Police Lineup
- ◆ Recognition Memory
- ◆ Similarity

Abstract

Eyewitness testimony relies on recognition memory of an event. This study examines the relationship between similarity of stimuli and recognition memory as it applies to eyewitness testimony. Participants viewed a study set followed by a test set, in which the similarity of alternatives was continuously manipulated. Items were presented in a lineup format, either side by side or serially. The results show a bias toward the central tendency of similarity, or the picture that is in the middle range of similarity. This causes accuracy to decline under certain conditions. This study has implications for police lineup situations, and suggests ways to address potential bias due to similarity relations among the set of alternatives.

Faculty Mentor



Mercedes La Voy's work is a great example of how basic research in human memory can have real-world applications, such as in police lineups. Typical lineups involve individuals chosen to be somewhat similar to each other to force the witness to be confident about their potential identification. It is not clear however how similarity relations among the alternatives affect identification. Mercedes La Voy investigated this question by testing participants with lineups involving pictures of outdoor scenes, one of which was shown in an earlier study phase. Mercedes showed that participants often based their recognition-memory decisions not on their actual memory but on the similarity relations between alternatives—test items that looked most like other items in the lineup were often chosen as targets.

Mark Steyvers

School of Social Sciences

Introduction

Mistaken eyewitness identification is the most important factor contributing to the conviction of innocent people (Wells and Olson, 2003). Research in human memory can lead to a better understanding of the various biases that arise from specific procedures used in collecting and testing eyewitness testimony.

The notion that similarity plays a critical role in recognition memory is widely accepted (Hintzman, 1988). The more similar the study and test items are, the harder it is to discern a difference between the two. For example, if given two pairs of pictures—two pictures of the ocean versus a picture of a car and a picture of a pencil—it should be easier to discern the difference between the car and pencil. Many experiments suggest that the accuracy of recognition memory is inversely related to the similarity between old and new test items. In other words, as similarity increases, accuracy decreases. Thus, the difficulty of a recognition test is increased by making the distracter pictures more similar to the target pictures.

Tulving, however, proposed that under certain conditions, recognition accuracy has a direct and positive relation to picture similarity, based on two experiments (1981). In Experiment 1, participants studied a series of pictures and took a two-alternative, forced-choice recognition test. Performance was compared in three test conditions: distracter is similar to target picture (A-A'), distracter is dissimilar to target, but similar to other previously studied pictures (A-B'), and distracter is dissimilar to both target and other study pictures (A-X'). The stimuli consisted of pictures that appeared on two adjacent pages in a magazine, where A' was the other half of the target picture, B' was a picture seen in the study list, and X' was a picture that had never been seen before. Similarity between study items and test items was manipulated by pairing different picture combinations. Accuracy in the A-X' condition led to the best performance, suggesting an inverse relation with similarity. However, accuracy was higher for the A-A' than the A-B' condition across all participants, suggesting a positive relation. Tulving later conducted a post hoc analysis, and identified two types of similarity. The first type was the similarity between a test picture and the stored memory of that item (ecphoric similarity). The second type is the similarity between test pictures in a set, which is also referred to as a test pair (perceptual similarity). Thus, ecphoric similarity is high in A-A' and A-B', and low in A-X'. Likewise, perceptual similarity is high in A-A', and low in A-B' and A-X'. The purpose of this analysis was to examine the relation between accuracy and

ecphoric similarity of distracters in A-A' and A-B'. Tulving's results were counterintuitive because the higher similarity condition resulted in higher accuracy, and the lower similarity condition yielded lower accuracy.

Experiment 2 tested the relationship between four test conditions: A-A', A-A'', A-B' and A-B''. In this notation, single primes indicate high ecphoric similarity of distracters and double primes indicate medium ecphoric similarity. Tulving hypothesized that accuracy might increase when similarity between stimuli increased, specifically when ecphoric similarity of distracters was high, and that accuracy would decline when ecphoric similarity was medium. The results showed that the difference in accuracy in A-A'' versus A-B'' was not statistically significant. These findings support the usual inverse relationship between similarity and accuracy when ecphoric similarity is not high. However, when comparing accuracy in A-A' versus A-B', the reversal of the typical test-item similarity effect was upheld, as in Experiment 1. Therefore, participants were more accurate when the pictures were more similar.

Tulving concluded that, under certain conditions, the typical test-item similarity effect was reversed. Specifically, this reversal can be seen when distracters are more similar to pictures seen earlier in the study, but are not used as target pictures in the memory test. Similarity of distracters to the information stored must be high to observe this positive correlation.

Hintzman (1988) used the MINERVA 2 memory model to explain Tulving's findings. Tulving's experiments have also been replicated to study the inverse relation between confidence and accuracy (Dobbins, Kroll, and Liu, 1998). Additionally, Wells (1993) applied Tulving's findings to the area of eyewitness identification, stating that ecphoric similarity must be considered in the judgment process. However, no experiments have been done to expand upon Tulving's work.

This study seeks to improve upon Tulving's experiments by separating a single picture into six individual subsections, which makes it possible to continuously manipulate the similarity of stimuli (Figure 1). This allows for an enhanced assessment of the relationship between picture similarity and recognition memory performance, analogous to a police lineup in which a decision needs to be made about the presence or absence of a suspect among a set of alternatives. We expect these findings to show the ideal amount of similarity between suspects in a lineup, so that bias is significantly decreased or eliminated, and the eyewitness can

make the most accurate judgment. In addition, where Tulving only considered two choices—analogue to a lineup of only two individuals—these experiments included five choices.

Recent legislation prohibits lineups from being conducted with suspects shown side by side. This law was implemented because eyewitnesses were identifying the suspect by comparison to the rest of the suspects in the lineup (Brulliard, 2005). This strategy works quite well if the perpetrator of the crime is present in the lineup, but if the perpetrator is not present, the eyewitness is more likely to falsely identify another suspect in the lineup. To simulate both side-by-side and sequential lineups, the presentation of the alternatives was varied: in Experiment 1, the alternatives were presented side by side and in Experiment 2, the pictures were presented sequentially. In sequential presentation participants make a judgment on which picture they see based on their memory of the picture, and not by comparing one picture to another in the lineup. Presenting the pictures serially should result in higher accuracy than presenting the pictures side by side.

The central hypothesis of this study is that accuracy will be highest when all of the distracters are dissimilar to the target picture, since participants will only recognize the target picture. The next highest level of accuracy should be seen when distracters are extremely similar to pictures seen earlier in the study, but are not used as target pictures in the memory test. In this condition, it should be difficult to identify the target picture since most of the pictures look familiar; however, the participant should be able to eliminate the dissimilar picture, which increases their probability of choosing the target. The next highest level of accuracy should be seen when participants may be able to eliminate the outermost subsections—those that are not neighboring the target image. Finally, the lowest level of accuracy should be found when all of the subsets are close to the target image, so that it would be difficult for participants to eliminate any extremes.

Methods

The stimuli were created by separating a single picture into six individual subsections (Figure 1). Participants viewed study pictures and their memory was tested in a five-alternative, forced-choice task, with the test items presented in a lineup format. Five test conditions examined similarity (Figure 2).

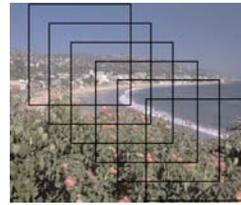


Figure 1
Example of stimuli creation from larger image

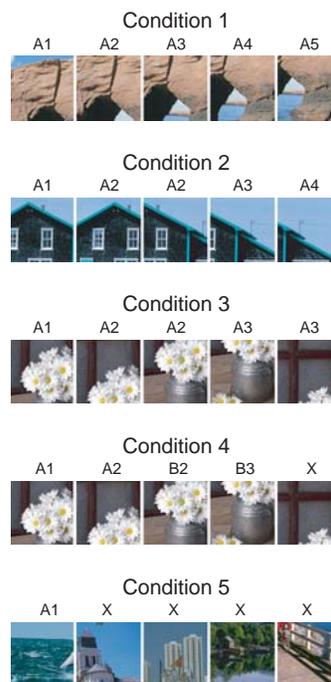


Figure 2
Example stimuli in the five test conditions

In Figure 2, A1 is the target picture shown in the study blocks and A2 through A5 are neighboring subsections of the same picture (either to the immediate left or the immediate right of the target picture). The Bs are dissimilar to the target, but similar to other previously studied pictures. B1 never appears as a target in a test trial; rather, neighboring subsections of B1 viewed in a study trial were used as distracters (B2 and B3). X is a random picture that has never been seen, and is dissimilar to any picture seen in any of the study or test blocks.

After the first study and test blocks, participants were shown the study pictures again (the same pictures, but displayed in a different, random order) and were given the second half of the memory test. During the second memory test, participants were asked about different pictures from those in the first test. An overview of the design is shown in Table 1. The study was approved by the Institutional Review Board (IRB) of UCI under protocol #2002-2699.

Table 1
Basic experimental design

Study Phase 1	5 buffer items + 96 target items + 5 buffer items = 106 items
Distracter Task	"Where's Waldo"
Test Phase 1	36 randomly selected items
Study Phase 2	Same 106 study items in a different random order
Distracter Task	Different "Where's Waldo"
Test Phase 2	36 remaining items in a random order

Participants

Undergraduate students from the University of California, Irvine served as participants in exchange for course credit. Experiment 1 included 50 participants and Experiment 2 included 58.

Apparatus

The experiment was implemented in Microsoft PowerPoint. The stimuli consisted of pictures that were presented on a desktop personal computer equipped with Windows 98 and a color monitor. To advance to the next trial in a study block, participants clicked an arrow on the screen. A paper answer sheet was used to record responses.

Design

In this one-way experiment, the independent variable of similarity of alternatives to the study item (A1-A2/A1-A3/A1-A4/A1-A5/A1-B2/A1-B3) was studied within subjects. The A1-X condition was added as a control. The dependent variable was accuracy, determined by the number of correct responses on the recognition memory test.

During the study, each picture appeared alone on the screen; during the test, the target picture was one of five pictures arranged horizontally on the screen. In other words, the target picture, which was viewed in the study, always appeared in the test along with four other distracter pictures. In each test trial, the position of the target picture was randomly assigned to position 1–5 in the lineup. This tactic was used to ensure that participants did not detect any patterns or cues in the position of the target picture. Each study picture was only used once to reduce a buildup of confusion from previous trials. Pictures of landscapes were selected for ease of manipulation.

There were four blocks: blocks 1 and 3 were study blocks, and blocks 2 and 4 were test blocks. Each study block consisted of 96 study pictures and 10 buffer items (5 at the beginning and 5 at the end) that were discarded at test to control for recency effects. The same 106 pictures appeared in random order for both study blocks. Each test block comprised 36 trials of five pictures, totaling 72 test trials. One picture was always the target, and the other four pictures in the lineup served as distracters. In total there were 288 distracters and 394 different pictures. With this design, recognition memory performance was tested in all five conditions. Order effects were counterbalanced by randomizing the conditions, the study images, and the test trials. The subsection chosen as A1 was counterbalanced. A distracter task of “Where’s Waldo” puzzles was inserted between each block. The study and test blocks were randomized across trials and participants to counterbalance for learning and order effects.

Participants were tested in five conditions, all of which were presented in a lineup. Condition 1 contained pictures: A1-A2-A3-A4-A5; Condition 2 contained pictures: A1-A2-A2-A3-A4; Condition 3 contained pictures: A1-A2-A2-A3-A3; Condition 4 contained pictures: A1-A2-B2-B3-X; and Condition 5 contained pictures: A1-X-X-X-X. (Figure 2).

Difficulty levels were achieved by dividing each picture into 6 subsections, with each subsection overlapping the last. The pictures were divided into subsections vertically, horizontally, and diagonally, with a 70% overlap between each subsection.

Multiple versions of the experiment were created; each version included different pictures, reducing the bias of individual participant. For example, in one version a participant may have seen a picture of a flower, whereas in another version a participant may have seen a picture of a house. If the first participant were particularly interested in flowers, he or she would be more likely to have an accurate memory of the flower picture, but such individual bias was accounted for by randomly assigning different pictures to different versions of the experiment. To account for ethical concerns, no graphic, offensive, insulting, or disrespectful pictures were used.

A second experiment was created in which the pictures were displayed serially. The design and apparatus were identical to those of the first experiment.

Results

A univariate Analysis of Variance (ANOVA) revealed a significant difference between the two experiments. When collapsing over stimulus condition and focusing on the overall accuracy in each experiment, it can be seen that the manner in which the pictures are presented has a significant effect on participant’s accuracy ($F(1,7774) = 37.7, p < 0.000$). Tables 2 and 3 provide the frequencies of responses in each condition, where A1 is always the target.

Table 2

Experiment 1: Frequency of responses in each condition, rounded to the nearest whole number

Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
A1 = 13	A1 = 13	A1 = 26	A1 = 38	A1 = 64
A2 = 21	A2 = 21	A2 = 17	A2 = 18	X = 9
A3 = 26	A2 = 21	A2 = 17	B2 = 24	X = 9
A4 = 28	A3 = 28	A3 = 20	B3 = 12	X = 9
A5 = 12	A4 = 16	A3 = 20	X = 9	X = 9

Table 3
Experiment 2: Frequency of responses in each condition, rounded to the nearest whole number

Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
A1 = 10	A1 = 15	A1 = 29	A1 = 48	A1 = 74
A2 = 23	A2 = 21	A2 = 17	A2 = 15	X = 7
A3 = 24	A2 = 21	A2 = 17	B2 = 23	X = 7
A4 = 32	A3 = 26	A3 = 19	B3 = 8	X = 7
A5 = 11	A4 = 18	A3 = 19	X = 6	X = 7

According to the ANOVA, the mean accuracy in Experiment 1 was significantly different from the mean accuracy in Experiment 2 within Condition 4 ($F_{(1,2590)} = 27.2, p < 0.000$) and Condition 5 ($F_{(1,2590)} = 31.0, p < 0.000$). Table 4 illustrates mean accuracy in both experiments and in all five conditions (the additional numbers flanking the table are average values).

Table 4
Mean accuracy across experiment and condition

	Conditions					Average
	1	2	3	4	5	
Experiment 1	0.13	0.13	0.25	0.38	0.64	0.40
Experiment 2	0.09	0.15	0.29	0.48	0.74	0.47
Average	0.11	0.14	0.27	0.43	0.70	0.43

Collapsing over experiment type, an omnibus ANOVA of condition showed a significant effect on correct response ($F_{(4,7771)} = 456.8, p < 0.000$). Subsequent condition contrast tests revealed significant differences in mean accuracy between all conditions ($p < 0.000$) except Conditions 1 and 2, which were not significantly different. Omnibus tests within both Experiments 1 and 2 showed significant differences in mean accuracy of condition $F_{(4,3595)} = 169.8, p < 0.000$ and $F_{(4,4171)} = 298.3, p < 0.000$, respectively. A further analysis of condition contrast pairs within both experiments revealed an identical pattern of differences in mean accuracy between all conditions ($p < 0.000$) except Conditions 1 and 2, which were not significantly different.

An additional ANOVA was conducted to examine accuracy in Test Phase 1 and Test Phase 2. When collapsing over experiment, this analysis showed a significant difference in accuracy between Test Phase 1 and Test Phase 2 in Condition 1, Condition 3, and Condition 4 ($p < 0.000$), with Condition 2 and Condition 5 not reaching significance. When the analysis was conducted within Experiment 1, a significant difference was found in Condition 1, Condition 3, and Condition 4 ($p < 0.000$), with Condition 2 and Condition 5 not reaching significance. When the analysis was conducted within Experiment 2, a significant difference was found in Condition 1, Condition 2, Condition 3, and

Condition 4 ($p < 0.000$), with Condition 5 not reaching significance. Table 5 and Table 6 illustrate the frequency of responses in both experiments and in all conditions in Test Phase 1 and Test Phase 2, respectively. This data is also shown graphically in Figure 3 and Figure 4, which are separated by experiment and compare the accuracy of each condition in Test Phase 1 and Test Phase 2.

Table 5
Test phase 1: Frequency of responses in each condition, rounded to the nearest whole number

	Conditions				
	1	2	3	4	5
Experiment 1	A1 = 19	A1 = 15	A1 = 20	A1 = 35	A1 = 67
	A2 = 22	A2 = 20	A2 = 15	A2 = 19	X = 8
	A3 = 25	A2 = 20	A2 = 15	B2 = 23	X = 8
	A4 = 23	A3 = 33	A3 = 25	B3 = 13	X = 8
	A5 = 9	A4 = 12	A3 = 25	X = 10	X = 8
Experiment 2	A1 = 18	A1 = 19	A1 = 18	A1 = 42	A1 = 75
	A2 = 17	A2 = 22	A2 = 18	A2 = 18	X = 6
	A3 = 28	A2 = 22	A2 = 18	B2 = 25	X = 6
	A4 = 26	A3 = 29	A3 = 23	B3 = 8	X = 6
	A5 = 9	A4 = 7	A3 = 23	X = 6	X = 6

Table 6
Test phase 2: Frequency of responses in each condition, rounded to the nearest whole number

	Conditions				
	1	2	3	4	5
Experiment 1	A1 = 10	A1 = 12	A1 = 31	A1 = 41	A1 = 62
	A2 = 20	A2 = 23	A2 = 18	A2 = 15	X = 9
	A3 = 26	A2 = 23	A2 = 18	B2 = 25	X = 9
	A4 = 31	A3 = 24	A3 = 15	B3 = 10	X = 9
	A5 = 13	A4 = 19	A3 = 15	X = 9	X = 9
Experiment 2	A1 = 5	A1 = 10	A1 = 41	A1 = 55	A1 = 74
	A2 = 26	A2 = 19	A2 = 16	A2 = 12	X = 6
	A3 = 21	A2 = 19	A2 = 16	B2 = 20	X = 6
	A4 = 36	A3 = 23	A3 = 14	B3 = 8	X = 6
	A5 = 12	A4 = 29	A3 = 14	X = 5	X = 6

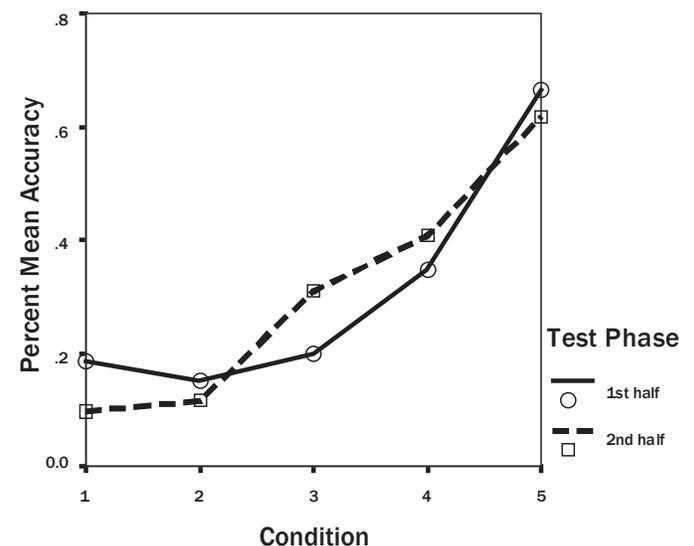


Figure 3
Experiment 1: Mean accuracy across conditions in Test Phase 1 versus Test Phase 2

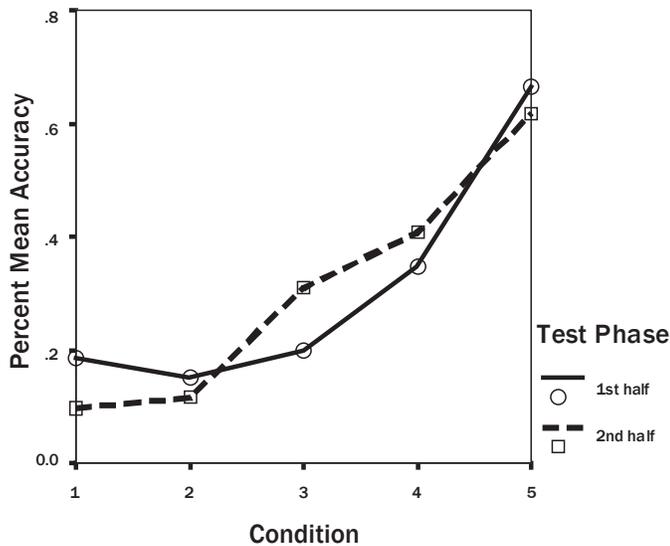


Figure 4
Experiment 2: Mean accuracy across conditions in Test Phase 1 versus Test Phase 2

Discussion

By examining response frequencies across stimuli in each condition, a strong bias was uncovered. The majority of responses fall within the middle range of similarity (Tables 2, 3, 5, and 6). Thus, it appears that participants tend to choose the picture that is the central tendency of similarity in the lineup, which often has nothing to do with their memory trace. This theory holds in Condition 3, in which participants were unexpectedly accurate. Their bias to choose the central tendency in the lineup led to a more accurate response, since the central tendency corresponded to the target picture. In other words, the lineup was constructed from the images A3-A2-A1-A2-A3 (randomly ordered across the screen), so the extremes were A3 and A2, and the central tendency was A1, the target.

As expected, accuracy was highest in Condition 5, since participants only recognized A1. The next highest outcome was seen in Condition 4, which opposes the usual test-item similarity effect, and shows that, with a greater degree of similarity between stimuli, participants become more accurate. Precision followed in Condition 3, since the bias toward the central tendency caused accuracy to increase. Participant's performance was similar in Conditions 1 and 2, since the similarity of stimuli in these two conditions was nearly identical, and participants were choosing the central tendency of similarity in both conditions.

The significant difference between Experiment 1 and Experiment 2 offers support for the recent legislation to

present suspects serially. In Experiment 1, participants had lower levels of accuracy. This was assumed to be because they were comparing the pictures to one another, rather than referring to their memory trace. In Experiment 2, participants were unable to compare the pictures to each other, and were more likely to choose the picture based on their memory trace, which led to higher levels of accuracy.

A global statement cannot be made regarding practice and learning effects from the first to second test block. It was hypothesized that accuracy would increase in Test Phase 2, which only occurred in select conditions. It seems as though there was a crossover because participants' biases likely grew stronger over the course of Test Phase 2 (Figures 3 and 4).

The bias in a participant's memory is crucial to police lineups and eyewitness testimony. In a situation in which an eyewitness is viewing a lineup, their choice of the perpetrator should be purely memory driven. Memory researchers, especially Gary Wells and Elizabeth Loftus, have helped uncover memory biases and have applied their findings to lineups. Tulving examined biases through a two-alternative forced-choice recognition task, but the use of a five-alternative, forced-choice recognition task has uncovered biases that Tulving did not find.

Future work in this area is critical to evaluating the accuracy of eyewitness testimony. The results of this study could be applied to recognition memory for pictures of faces presented in a lineup. Asking participants to rate their confidence for each test trial could reveal the inaccuracy of eyewitness testimony. Also, researchers could modify this study's experimental design, showing the target picture in only half of the test trials and giving participants the option of saying the target is not present. There is a practical application to eyewitness testimony inherent in the design that only shows the target picture in half of the test trials, since not all police lineups contain the actual perpetrator.

The results of this study indicate that the test-item similarity effect is countered under certain conditions. A bias toward the central tendency similarity was clearly indicated, which decreased accuracy in some conditions. It is possible to infer the ideal amount of similarity between suspects in a lineup, so that bias is significantly decreased or eliminated, and the eyewitness can make an accurate judgment. Specifically, when suspects in a lineup are similar to the perpetrator in the lineup, the typical test-item similarity effect is reversed, and eyewitnesses make more accurate judgments. If suspects in a lineup are radically different from one another, eyewitness accuracy will likely be high if the per-

petrator is present, but low if the perpetrator is not. The former situation indicates that the perpetrator is too obvious (e.g., a lineup composed of primarily Hispanic males, with one Caucasian male). The latter situation indicates that the eyewitness will likely fail to accurately access his or her memory trace for the event, and rather choose the suspect who most closely resembles the perpetrator. Based on these results, suspects in a lineup should be presented serially to prevent eyewitnesses from identifying the suspect by means of comparison to the rest of the suspects in the lineup.

Acknowledgements

I would like to thank Drs. Mark Steyvers and Scott Brown for their support and guidance. They challenged me to go above and beyond my expectations of myself, and it was through this encouragement that I realized my true potential. I would also like to thank Miles Munro for his assistance throughout the project. The project was supported by the Undergraduate Research Opportunities Program at UCI and the Summer Undergraduate Research Program.

Works Cited

- Brulliard, K. (2005, March 6). "Revamping Va.'s police lineups: New methods urged to curb mistakes." Washington Post, p. C01.
- Dobbins, I.G., N.E.A. Kroll, and Q. Liu. (1998). Confidence-accuracy inversions in scene recognition: A remember-know analysis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(5), 1306–1315.
- Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. Psychological Review, 95(4), 528–551.
- Loftus, E.F. (1996). Eyewitness testimony. Cambridge, MA: Harvard University Press. (Original work published 1979).
- Loftus, E.F. (2002). Memory faults and fixes. Issues in Science and Technology, Summer, 2002, 41–50.
- Loftus, E.F. (2003). Our changeable memories: Legal and practical implications. Nature Reviews: Neuroscience, 4, 231–234.
- Tulving, E. (1981). Similarity relations in recognition. Journal of Verbal Learning and Verbal Behavior, 20, 479–496.
- Wells, G.L. (1993). What do we know about eyewitness identification? American Psychologist, 48(5), 553–571.
- Wells, G. L. and E.F. Loftus. (2003). Eyewitness memory for people and events. A. M. Goldstein (Ed.) Handbook of Psychology. Vol 11 Forensic Psychology (I.B. Weiner, Editor-in-Chief). New York: John Wiley and Sons, pp. 149–160.
- Wells, G.L. and E.A. Olson. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. Journal of Experimental Psychology: Applied, 8(3), 155–167.
- Wells, G.L. and E.A. Olson. (2003). Eyewitness testimony. Annual Review of Psychology, 54, 277–295.

