How well do police assess the accuracy of eyewitness identification decisions?

Mario Baldassari, Catalina Dau, D. Stephen Lindsay, C. A. Elizabeth Brimacombe

INTRODUCTION

- False lineup choices can lead to innocent people going to jail and guilty people going free.
- Previously in our lab, undergraduate participant investigators (PIs) were as influenced by incorrect participant witnesses (PWs) as by correct PWs (Boyce et al., 2008; Tweedy, 2011).
- How much would police officers outperform undergraduates?

METHOD

- N = 135 undergraduates (UG) (28 m, 107 f)
- Police N = 15 (11 m, 4 f), recruited from local departments, paid $100.
- Each PI interviewed four PWs, each of whom saw a different crime.
- 50% of PWs received a good (GV) view of the criminal, 50% received a poor (PV) view.
- After interviewing, PIs selected a suspect from a police blotter and administered a lineup.
- PIs made a series of confidence ratings before lineup administration, and PIs and PWs both made ratings after lineup administration.

RESULTS

- Compared to undergrad PIs, police PIs produced:
  - Longer interviews
  - Higher recall accuracy for crime video details
  - Slower ID latencies

- More selections of the intended suspect
- 33% of the 55 police interviewers asked about view quality (0 in 64 UG interviews)
- Police less influenced by evidence of innocence (CR and M)
- Police and students largely influenced by hits
- Police and students both somewhat less influenced by FA’s than by hits, but police slightly more often influenced by FA’s than students.

DISCUSSION

- Investigators also rated confidence in each witness after the lineup ID
  - Police did a better job disregarding witnesses who selected foils than did student-investigators
  - Witnesses who inspired more confidence in a CA situation tended to be those who rejected that CA lineup
  - Police also tended to have higher confidence in a witness who rejected than did student-investigators

- Witnesses behaved differently depending on the quality of view they had of the culprit
- Police drew out these behavioral differences more than student-investigators
- Police still put undue weight on lineup decisions made by poor witnesses
- Police did not obviously outperform students in terms of Change in p(S=C)

REFERENCES


Developing a Test of Eyewitness Identification Skill

Mario Baldassari, Justin Kantner, D. Stephen Lindsay

INTRODUCTION

• Jurors and police poorly discern accurate from inaccurate witnesses (Lindsay, Wells, & Rumpel, 1981).
• Liberal response bias on a yes/no recognition test predicted choosing on culprit-absent [CA] lineups, which was consistent with other evidence that recognition response bias is a stable individual difference (Kantor & Lindsay, 2014).
• Bindemann and colleagues (2012) found a relationship between scores on target present trials of the Bruce (1999) 1-10 task and culprit-present (CP) lineup accuracy rates.

METHOD

• Four samples (N’s = 65, 91, 78, 115) of participants watched five crime videos, did a distractor task, completed five CA lineups, and then took our Lineup Skills Test
• Skills test included study phase of 30 faces. test had two components: 30 pairs of two new faces (N/N pairs, the Proclivity to Choose portion) and 30 pairs with one new and one studied face (Old/New pairs, the Face Recognition Skill portion)
• Two samples (N’s = 95, 113) with 50 study faces and 50 test pairs in an attempt to strengthen the predictive utility of the test
• Test pairs description-matched to be more like lineups
• Matching pairs required more ethnic- and age-diverse set, obtained from Park Aging Mind Laboratory at UT-Dallas
• Concurrent samples (N’s = 94, 104, 87, 89) collected with several of these with CP lineups to test predictive utility of Face Recognition Skill portion (i.e., accuracy an old/new pairs) for accuracy on CP lineups.
• Sixth sample also included a second, forced-choice lineup

REFERENCES


RESULTS, DISCUSSION, THEORIES

As shown in the graphs below, there was a mostly consistent relationship between erroneously choosing on New/New face pairs and false ID’s on CA lineups across four samples, but the new stimulus set that was meant to strengthen the NN/CA relationship instead appeared to weaken it in samples 5 and 6.

Why weaker? 4 possible explanations:
1. Easy items leading to restricted range
2. Participants wrongly guessing that list diversity is the variable of interest, demand characteristics
3. Error variance
4. Second, forced-choice lineup changing the way p’s think about following lineups

But also, does that second lineup add any information or predictive utility?

• Proclivity To Choose is reliable across tasks in the Lineup Skills Test, but Face Recognition Skill is not
• Our attempts to widen the range of ON pair accuracy rates and strengthen the correlation with lineup selections have done the opposite
• Next steps will be scaling back to determine the methodological change that resulted in the reduction of the strength of the LST
• The end goal is still to develop the LST into a tool that can be combined with other known predictors of eyewitness identification accuracy in a larger function that can account for as much of the variance among witnesses as possible
The Presence Procedure: Does asking eyewitnesses to confirm/deny presence of the culprit 
before selecting alter patterns of lineup responses?

Baldassari Mario J., mjbldssr@uvic.ca

Lindsay, D. Stephen, slindsay@uvic.ca

Freeman, Jordanna L., lyris@uvic.ca

Corresponding author: Mario J. Baldassari, mjbldssr@uvic.ca, Telephone: 250-721-6165

University of Victoria
PO Box 1700 STN CSC
Victoria BC V8W 2Y2
Abstract
Eyewitness memory can be fallible and can lead to an eventual error in a lineup identification. We tested whether an initial presence/absence judgment would improve eyewitness’ ability to reject a culprit-absent (CA) lineup without impairing their ability to identify the perpetrator from a culprit-present (CP) lineup, a method we have dubbed the Presence Procedure. In a sample of 237 Mechanical Turk workers, those who completed the Presence Procedure tended to reject (both CA and CP) lineups more often than those who responded to a traditional lineup. The relative trade-offs of witnesses adopting a more conservative response criterion in an identification procedure are discussed in the context of the most recent literature.

Keywords: Eyewitness Memory; Eyewitness Identification; Decision Criterion; Alternative Lineup Methods
The Presence Procedure: Does asking eyewitnesses to confirm/deny presence of the culprit before selecting alter patterns of lineup responses?

Psychological research on eyewitness identification has provided police officers and legal experts with much information on how to improve the diagnosticity of lineup responses (Valentine, 2014). In the terminology of signal detection theory (used nicely in the context of eyewitness ID by Clark, 2012), researchers are striving to develop suspect identification procedures that minimize false alarm rates (often called false identification rates) while maximizing hit rates. In the following we briefly review a number of suspect-identification procedures that have been shown to reduce false alarm rates, with or without also lowering hit rates. We then introduce a new procedure and two tests of that procedure.

Many police have already adopted some changes to lineup methods that were originally proposed in psychological research (Bertrand, 2014). For example, most Canadian and American police ensure that all lineup members fit a rough description of the culprit. This practice typically reduces false identifications in culprit-absent (CA) lineups to a greater extent than it reduces correct identifications in culprit-present (CP) lineups (Fitzgerald, Oriet, & Price, 2014; Wells, Lüüs, & Winsdschitl, 1994). Informing participant/witnesses that the culprit might not be present in the lineup usually does the same, as does providing an explicit “not present” response option (Malpass & Devine, 1981; Wells & Olson, 2003). Most police in North America recommend that their officers use unbiased instructions, though practice is sometimes inconsistent (Bertrand, 2014). Also, Weber and Perfect (2012) reported evidence that providing participant witnesses with an explicit “Don’t know” option when presenting a show-up (the culprit or an innocent suspect) reduced false identifications but not hits, hence improving discrimination relative to the
standard show-up procedure without an explicit “Don’t know” option. Our conversations with local police indicate that they do not specifically offer a “don’t know” option, perhaps because they are not aware of or convinced by evidence that it does not reduce hits or because many real-world eyewitness tend to say “I don’t know” even when it is not a specific option (of course, police practice varies widely (Bertrand, 2014).

Sauer, Brewer, and Weber (2008, 2012; Brewer, Weber, Wootton, & Lindsay, 2012) developed a procedure in which participants rate each member of a sequential lineup on similarity to the culprit; those ratings are then used to classify each lineup member as selected or not based on a cutoff rating established by a mathematical model. This method effectively establishes decision criteria for the witnesses rather than depending on the witness to set an optimal threshold. The restrictions placed on the model had a strong effect on the number of correct selections of the criminal. As far as we know this procedure has not been adopted by any law enforcement professionals, perhaps because of the complexity of both the procedure and the interpretation of its results.

Thus, psychologists have gathered evidence that some changes to lineup procedures improve diagnosticity by reducing false alarm rates while having less or no significant effect on hit rates. Some other approaches that were designed to improve diagnosticity appear instead to cause participant/witnesses to adopt a more conservative decision criterion, thereby lowering both hit and false alarm rates when compared to traditional lineup procedures. Depending on the perceived cost of false alarms and the perceived benefit of hits, and on the prior odds that the culprit is in the lineup, arguments can be made in favour of using lineup procedures that lead to adoption of a more strict decision criterion even if they do not enhance diagnosticity (Malpass, 2006; Wells & Olson, 2003).
Presentation of separate voice, face, and body lineups is a clear example of this sort of effect (Pryke, Lindsay, Dysart, & Dupuis, 2004). Sequential as compared to simultaneous presentation of lineups has shown either improved diagnosticity or conservative response criterion adoption, depending on other conditions and on how performance was measured (Gronlund, Wixted, & Mickes, 2014; Lindsay & Wells, 1985; Wells, Steblay, & Dysart, 2015). Zajac and Karageorge’s (2009) wildcard procedure improved diagnosticity for lineup decisions by children, but further testing (Pozzulo, Reed, Pettalia, & Dempsey, 2015) showed that insertion of the wildcard caused adult witnesses to adopt a more conservative response criterion. Pozzulo and Lindsay’s (1991) elimination lineup procedure involves removing lineup members one by one until a single member is left, and then deciding whether that member is the criminal. Critically, witnesses are told to remove the member who looks least like the culprit until the one who looks most like the culprit is left. Then, witnesses are asked whether the remaining member is the culprit or not. The authors found that the procedure was successful in reducing false identifications from child eyewitnesses with no accompanying, significant reduction in correct ID’s. In this study and several since (e.g. Pozzulo et al., 2008; Pozzulo, Reed, Pettalia, & Dempsey, 2015), the same effect did not seem to carry over to adult witnesses. Pozzulo et al. (2015), for example, found very similar reductions in identifications in CP (26% change) and CA (31% change) lineups. The bulk of the evidence suggests that the true effect of the elimination lineup is the adoption of a more conservative response criterion compared to a traditional lineup.

One limitation of the elimination procedure is that participants must indicate the lineup member who looks most like the culprit before answering whether that person is the culprit or not. If this final remaining member of the lineup were the suspect but the witness ultimately rejected the lineup, a police officer might perceive that the suspect has been judged “most
similar” and therefore use the lineup procedure as inculpatory evidence. This scenario would obviously be worst when the suspect was innocent. Also, that the elimination lineup has not been widely adopted since its initial publication suggests some reluctance to adopt it in the field persists. Although Pozzulo and colleagues’ and Sauer and colleagues’ procedures improve diagnosticity, they may be too complex to be easily adopted and interpreted by police.

We propose a reversal of the decisions in the elimination procedure. We tested the effects of preceding a standard photospread lineup identification task with an initial presence/absence judgment: “Do you see the criminal in the lineup? If you say, “Yes,” you will have a chance to pick him out on the next page.” We speculated that this “Presence Procedure” might heighten the salience of the possibility that the correct answer is “No,” and that it might therefore discourage participants from framing the task as one of finding the lineup member who looks most like the culprit. The Presence Procedure puts the possibility of presence and absence on equal footing, so to speak, rather than the standard procedure in which rejecting a 6-person lineup is merely one of seven decision options. Increasing the salience of the option to reject the option should lead to more rejections overall, which would reflect a witness’s adoption of a more conservative response criterion similar to those found for the wildcard and elimination procedures.

On the other hand, our procedure may encourage witnesses to adopt an absolute judgment strategy and therefore result in an increase in correct rejections with no corresponding reduction in rejections when the culprit does appear in the lineup. Wells (1984) proposed that witnesses could use either an absolute or a relative judgment process when performing an identification test. In an absolute judgment process, an identification is made only if a lineup member attains a degree of resemblance to memory for the culprit that exceeds the witness’s response criterion. In a relative judgment process, the witness compares lineup members to one another, deciding which of them most resembles the culprit. Wells argued that the relative judgment process contributes to errors on culprit-absent lineups in that an
innocent suspect will be falsely chosen if he happens to be the lineup member who most resembles the culprit. As far as we know, compelling experimental evidence for this idea is lacking, but correlational data have been interpreted as supportive of it (e.g., false IDs on CA lineups tend to have long latencies) and Clark, Erickson, and Breneman (2011) reported simulations using the WITNESS model in which a version of the model that used an absolute judgment strategy clearly outperformed versions that used a relative judgment process when lineup foils were suspect-matched (as they often are in the real world and in the lab). If witnesses assume that the lineup probably includes the culprit, then it is rationale for them to use a relative judgment process. Our Presence Procedure might encourage use of an absolute judgment strategy, because it explicitly asks subject/witnesses to frame the task as a yes/no decision as to whether or not the culprit is present (as opposed to implicitly framing it as “which of these looks most like the culprit?”). The witness may infer from this framing that there is a 50/50 chance of the culprit appearing in the lineup, which should discourage use of a relative judgment strategy. If so, then our procedure should reduce false identifications when the culprit is absent more than it reduces hits when the culprit is present.

**Pilot Study**

**Method**

**Participants.** Participants were recruited online via Amazon’s Mechanical Turk (MT) \((N = 82)\) and on the University of Victoria campus through the psychology participation pool \((N = 59)\). MT workers participated for \$0.60 and UVic students participated for bonus course credit. Following exclusionary criteria established before the study, participants who confessed to major distractions or to skipping portions of the study were eliminated from the analysis \((N = 5)\), as were participants who recognized any actors from the videos \((N = 1)\). These exclusions were based on criteria established in previous studies (Baldassari, Kantner, & Lindsay, 2014). Data from the remaining 76 participants were used for analysis. All participants were randomly assigned to the experimental or control group, but a high dropout rate for MT participants in the
control group led to a slightly larger sample in the experimental group (\(N = 76\) combining campus and MT) than the control group (\(N = 59\)).

Participants self-reported demographics. The average reported age was 30.8 years, with a range from 18 to 63. The sample included 42 women, 62 native English-speakers, 50 who reported having earned at least a bachelor’s degree, and 33 who reported having taken no university courses in psychology.

**Materials and procedure.** Participants who accepted the task on Amazon’s work exchange server were linked to a survey hosted on Qualtrics, where they viewed the videos. Participants on campus signed up in exchange for optional bonus points and were given a direct link to the Qualtrics site in an individual testing room where an experimenter sat nearby. The five videos were clipped from British television crime dramas, all of which depicted middle-aged Caucasian male culprits committing crimes. A clip of a man breaking into a home was obtained from “Vincent,” a clip of a man and woman arguing and a clip of a woman’s car exploding as she leaves her home were obtained from “MI-5,” a clip of a man destroying china cabinets with a gun was obtained from “Dalziel and Pascoe,” and a clip of a man shooting another man was obtained from “Murder City.” Clips ranged from 46 to 83 seconds in length and were presented with the original sound tracks. The videos were followed by a short distractor task, after which participants completed five CA lineups. We constructed a CA lineup for each video, each consisting of six photos of men who fit a description of the culprit which we selected from the State of Florida’s online database of criminal mugshots. The mugshots were edited so that every member wore the same clothes and appeared in front of a neutral background. The filler face we thought most resembled the culprit was predesignated the “innocent suspect” in the CA lineup for each crime. Crime order was counterbalanced. When the lineup phase began, participants in
the experimental condition were shown six faces without numbers underneath and asked “Do you see the criminal in the lineup? If you say yes, you will have a chance to pick him out on the next page.” If the witness answered in the affirmative, the computer then displayed numbers under the faces and under the “none” option and asked for a response and a confidence rating from 1% to 100%. If the witness responded in the negative to the initial question, the program asked for a confidence rating and moved on to the next lineup. In the control condition, numbers appeared below the lineup photos alone with the “none” option and standard instructions were used (See Appendix A for a full transcript of the instructions).

**Results**

In the pilot study, on average, the standard lineup procedure produced 61.1% false positive IDs [95% CI: 57%, 65%] while the new procedure produced only 47.7% false positive IDs [95% CI: 44%, 51%]. These percentages were significantly different, $t(133) = 2.66, p < .01$, Cohen’s $d = .46$ [95% CI: 0.11, 0.80]. A Bayesian $t$-test using the prior hypothesis that there would be more rejections with the new procedure showed moderate evidence for the alternative hypothesis, $BF_{10} = 8.901$. These results indicated a difference between the two methods, but without presenting culprit present (CP) lineups we were not able to discern whether the difference was due to an increase in discrimination or merely a shift toward a more conservative decision criterion.

**Current Study**

To test whether requiring an initial presence/absence judgment improved participant’s discrimination or merely led them to be more conservative, we replicated the pilot study with both CA and CP lineups.

**Method**
Participants. Participants were recruited online via MT (N = 381) for $0.60. Following exclusionary criteria established before the study, participants who confessed to major distractions or to skipping portions of the study were eliminated from the analysis (N = 15), as were participants who did not stay on the video pages long enough to watch them (N = 102) and those who responded to at least one lineup in less than 1,000 ms or more than 15,000 ms (N = 27). These exclusions were based on criteria established in previous studies (Baldassari et al., 2014). Data from the remaining 237 participants were used for analysis.1

Participants self-reported demographics. The average reported age was 34.2 years, with a range from 18 to 69. The sample included 145 women, 213 native English-speakers, 144 who reported having earned at least a bachelor’s degree, and 77 who reported having taken no university courses in psychology. Participants were randomly assigned to conditions: 122 in the control and 115 in the experimental condition.

Materials and procedure. Participants accepted the task on Amazon’s work exchange server and were linked to a survey hosted on Qualtrics, where they viewed the videos. Most procedural details were the same as in the pilot study, thus we will only discuss differences. Four of the videos from the pilot study were used in this study: the video from “Vincent,” the video from “Murder City, and the two from “MI-5.” Clips ranged from 47 to 83 seconds in length and were presented with the original sound tracks. The videos were followed by a short distractor task, after which participants completed two CP and two CA lineups. The photo of the culprit in the CP lineup was a still from a portion of the video not included in the clip presented in the study phase. Culprit presence and crime order were counterbalanced.

Results

1An initial sample of 109 participants who passed the exclusionary criteria only produced 20 false identifications of the predesignated innocent suspect. An additional sample of 128 participants was collected. The addition of this second sample did not affect the overall shape of the ROC curves.
ROC curves were constructed to compare conditions based on the method suggested by Gronlund, Wixted, and Mickes (2014). See Figure B.1 for ROC curves treating identifications of any member of a CA lineup as a false alarm. Calculating these values necessitates counting any foil ID as a false positive ID, because these IDs represent breaches of the decision criterion. In Figure B.1, the ROC curves (with the area cut off at a 35% false ID rate) did not have significantly different areas under the curve (partial AUC here because of lineup choices’ restricted range), $D = -0.193$, $p = 0.85$, pAUC experimental $= 0.114$ [95% CI: 0.0917, 0.136], pAUC control $= 0.117$ [95% CI: 0.097, 0.139]. $D$ is the difference between the AUCs divided by the standard error of that same difference, which is estimated here via the bootstrap method with 10,000 bootstraps (Mickes, Flowe, & Wixted, 2012). At lower response criteria, choosing rates were higher in the control than experimental condition. There was a numerically higher sensitivity value in the two-step condition, $d’ = 0.235$ (95% CI: 0.109, 0.36), than in the control condition $d’ = 0.149$ (95% CI: 0.026, 0.273), but the confidence intervals overlap. Post hoc power analyses showed power of 0.74 to detect an effect size of 0.3, and power of 0.98 to detect an effect size of 0.5. As we hypothesized no difference between these values, we conducted Bayesian hypothesis tests because of their ability to find evidence in favour of the null hypothesis. A Bayesian hypothesis test for $d’$ with the prior assumption that the experimental group would have a higher average $d’$ value showed strong evidence in favour of the null hypothesis ($BF_{01} = 21.70$). Taken together, these data suggest that the Presence Procedure did not enhance discrimination.

Though we found little evidence that an initial presence/absence question enhances discrimination, there were more hits and more false alarms in the control condition compared to

---

2 Appendix C includes information regarding ROC curves using only selections of the pre-designated innocent suspect as false alarms.
the experimental condition, which suggested a possible difference in decision criterion, as can be seen in Table B.1. After substantial correction for extreme values,\(^3\) participants in the control condition displayed essentially neutral bias, \(C = -0.036\) (95% CI: -0.099, 0.027), while those in the Presence Procedure showed a slightly conservative bias, \(C = 0.10\) (95% CI: 0.038, 0.161). These two values of \(C\) were significantly different, \(t(245) = 3.05, p = .003,\) Cohen’s \(d = .39\) (95% CI: 0.134, 0.642). A Bayesian hypothesis test of the difference in \(C\) using the prior hypothesis that bias would be more conservative for the Presence Procedure showed moderate to strong evidence for the alternative hypothesis (\(BF_{10} = 21.703\)).

**Discussion**

The data show a significant difference in \(C\) between the two conditions, which when combined with the lack of difference in areas under the ROC curves and the lack of difference in \(d'\) indicates that participants who completed the Presence Procedure adopted a more conservative decision criterion than did their counterparts who were given standard lineup instructions. The Presence Procedure reduced all positive identifications, similar to the multiple lineups, elimination, and wildcard procedures.

In a conference presentation, Guerin and Weber (2015) reported several studies in which participants studied photos of faces and were later tested on mini-lineups, for which some participants made an initial yes/no judgment as to whether or not a studied face was present. Guerin and Weber reported that discrimination was higher among participants who made an initial presence/absence judgment compared to those who made no such judgment. There were

\(^3\) Sensitivity (\(d'\)) and criterion (\(c\)) cannot be calculated from values of 0% or 100% accuracy. Typical corrections are to replace these values with a number halfway between the extreme value and the next closest possible value. In this case, 0% accuracy was changed to 25% (experimental \(N = 81\), control \(N = 61\)) and 100% was changed to 75% (experimental \(N = 47\), control \(N = 74\)) due to the three possible outcomes on each type of lineup (CP/CA), zero correct (0%), one correct (50%), or both correct (100%).
many differences between our procedures and theirs, and further research will be needed to determine the conditions under which the Presence Procedure leads to adoption of a more conservative response criterion versus those under which it enhances discrimination.

If the Presence Procedure indeed induces an absolute judgment strategy, then the data showing a shift in criterion provide evidence against the sharp distinction between absolute and relative judgments. It is likely the fact that these two judgment types act in concert with a host of other cognitive processes such as motivation to find the criminal, task demands present at identification, and individual proclivity to choose (Baldassari et al., 2014). These results, along with those produced by the WITNESS model (Clark et al., 2011), lead us to believe that it is a critical next step for psychologists to find a way to directly test the differences between these two judgment strategies.

**Practical Application.** Our study indicates that adding an initial presence/absence judgment procedure leads to adoption of a more strict decision criterion with no change in discrimination. Thus, using this Presence Procedure would be advisable when the prior odds of the suspect being the culprit are low. For example, in a situation in which a police officer constructed a lineup very early in an investigation with little evidence that the suspect is the culprit, this two-step procedure could help avoid unnecessary false IDs. Research indicates that police may not be enthusiastically willing to change lineup presentation methods and some procedural limitations may disable them from changing procedures quickly (Bertrand, 2014), but such a small and easy change may be easier to implement when an officer has reason to believe that it may be useful.
Conflict of Interest Statement

The authors declare that they have no conflicts of interest.

Funding

NSERC provided funding that paid M.Turk workers for participation in this study.
References


out those who say they don’t know. Law and Human Behavior, 36, 28-36.


Appendix A

Lineup explanation, from both conditions:
You will now be shown a set of six photographs of faces and asked to attempt to identify the perpetrator of the videotaped crime in which [crime explanation]. The person who committed the crime may or may not be in the lineup—that is, sometimes police mistakenly suspect the wrong person and build a lineup around someone who had nothing to do with the crime (and in some such cases a witness mistakenly identifies the innocent suspect, sometimes ultimately leading to a tragic miscarriage of justice in which an innocent person is imprisoned).

Imagine that you are making your identification decision as part of a police investigation and that your decision is likely to have serious consequences. You don’t want to let the culprit get away, but you also don’t want to falsely accuse an innocent person and you don’t know for sure whether or not the criminal is in the lineup.

If you can identify the person who committed the crime, please select the number of that person’s photo underneath the lineup. If you think that the person that you saw commit the crime is not in the line-up, then please select "None."

You will then be asked to rate your confidence in your decision and to rate how difficult it was to make that decision. Please indicate your answer by sliding the scale on the screen to the appropriate number between 1 and 100.

With the lineup on-screen:

Two-step condition:
Do you see the criminal in the lineup? If you say yes, you will have a chance to pick him out on the next page.
Traditional condition:

If you can identify the person who committed the crime, please select the number of that person’s photo. If you think that the person that you saw commit the crime is not in the line-up, then please select "None."
## Appendix B

### Control

<table>
<thead>
<tr>
<th></th>
<th>Susp ID</th>
<th>Foil ID</th>
<th>Rejection</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP</td>
<td>0.705</td>
<td>0.147</td>
<td>0.147</td>
</tr>
<tr>
<td>CA</td>
<td>0.087</td>
<td>0.230</td>
<td>0.683</td>
</tr>
</tbody>
</table>

### 2-Step

<table>
<thead>
<tr>
<th></th>
<th>Susp ID</th>
<th>Foil ID</th>
<th>Rejection</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP</td>
<td>0.673</td>
<td>0.114</td>
<td>0.213</td>
</tr>
<tr>
<td>CA</td>
<td>0.045</td>
<td>0.208</td>
<td>0.747</td>
</tr>
</tbody>
</table>

*Table B.1.* All ID rates in each condition.
Figure B.1. ROC curves of the experimental condition that included the initial yes/no question and the control condition with standard lineup instructions, all foil choices were counted as false positive identifications. There were 11 confidence bins (0, 10, … 100), but no participants who made IDs used the confidence level of 0. Thus, only ten points appear in the ROC curves because of an overlap at the most inclusive end.
Appendix C

Partial AUC for the ROC’s in Figure 4 were significantly different, \( D = 2.12, p = 0.034 \) pAUC exp = 0.031 [95% CI: 0.023-0.039], pAUC control = 0.019 [95% CI: 0.011-0.028], but this difference should be viewed with suspicion due to the very limited range of the curves used to make the comparison (an 8% false ID rate). The ranges were so limited because it would be inappropriate to extend the ends of the ROC curve so far past the highest false alarm rate in the data, which was very low in Figure 3, without making major assumptions. This difference was additionally inconclusive because it did not take into account foil IDs, which are breaches of the decision criterion. When participants gave the highest confidence rating, there was no effect of condition.

![Lineup ROC: Initial Yes/No Question, Innocent Suspect IDs](image)

*Figure C.1.* ROC from combined samples counting only false identifications of the pre-designated innocent suspect as false alarms. There were 11 confidence bins (0, 10, … 100), but
no participants who made IDs used the confidence level of 0. Thus, only ten points appeared in the ROC curve for those who received standard instructions because of overlap at the most inclusive end. Likewise, no participants who made correct IDs or false IDs of the predesignated innocent suspect used the confidence levels of 0 or 10. Thus, ROC for the two-step instructions appeared to only have 9 data points because three points overlap at the most inclusive end.
A History of the Society for Applied Research in Memory and Cognition

William E. Crozier*
John Jay College of Criminal Justice, United States

Ella K. Moeck
Flinders University, Australia

Camille C. Weinsheimer
Simon Fraser University, Canada

Dawn-Leah L. McDonald
Victoria University of Wellington, New Zealand

Mario J. Baldassari
University of Victoria, Canada

To contribute to the 2015 Conference Retrospective, we chronicled the 22-year history of the Society for Applied Research in Memory and Cognition (SARMAC). Throughout the 70s, 80s, and early 90s, there was no society dedicated to applied cognitive work. That changed in 1994, when researchers at the third iteration of the Practical Aspects of Memory conference agreed to establish SARMAC. Since then, SARMAC membership has grown tenfold and its conferences have traveled the globe. Recently, the society established this very journal for the dissemination of applied research in memory and cognition. SARMAC has a promising future marked by increased student involvement, the announcement of JARMAC’s inaugural impact factor, and a visit to Australia for the society’s 2017 conference.

**General Audience Summary**

Founded in 1994, the Society for Applied Research in Memory and Cognition (SARMAC) is an organization dedicated to promoting and showcasing the best applied cognitive research. In this article, we have recorded the society’s 22-year history by outlining the need for SARMAC based on the state of the applied cognitive field at the time, and how SARMAC has grown since its creation. We conducted interviews with some of SARMAC’s past and present prominent members, reviewed publications (including those resulting from SARMAC’s predecessor, the Practical Aspects of Memory conferences), and compilations of SARMAC’s own archived membership and conference records. Together, our research shows that SARMAC was formed to support a diverse and rapidly growing community of applied cognitive researchers. SARMAC has excelled in meeting this goal; as evidenced by its diverse and committed membership, extremely successful biennial conferences, an influential new journal (JARMAC), and ambitious plans for the future. Simply put, the state of the society is strong.

**Keywords:** SARMAC, JARMAC, Applied, Memory, Cognition, History

* Correspondence concerning this article should be addressed to William E. Crozier, Psychology – Rm 10.61.01, 524W 59th St., New York, NY 10019, United States. Contact: wcrozier@jjay.cuny.edu.

Author Note

We sincerely thank Bob Belli, Michael Gruneberg, Paula Hertel, Rod Lindsay, Steve Lindsay, Christian Meissner, Kazou Mori, Kathy Pezdek, Don Read, Deryn Strange, and John Turtle for their assistance during the research and preparation of this article. We especially extend our gratitude to Maryanne Garry for her patience, guidance, and unwavering support on this project.
The Society for Applied Research in Memory and Cognition (SARMAC) was founded in 1994 for the purposes of creating a professional home to promote the emerging field of applied cognitive psychology. Despite its 22-year existence, the history of the society has been largely undocumented. In this paper, we first chronicle SARMAC’s beginnings as an outgrowth of the Practical Aspects of Memory conferences. Next, we provide an account of how SARMAC has grown by reviewing the past conferences, including the research topics addressed at each, and the membership of the society over the past two decades. Finally, we close with a look to the future of SARMAC. The information contained in this article is derived from a collection of personal interviews, records kept by SARMAC, and publications; that is, it is based on a combination of assorted documents and memory, rather than a thoroughly recorded historical account. Accordingly, it is worth mentioning something that will be familiar to many cognitive psychologists: the reconstructive nature of this history will likely contain minor departures from reality. Nonetheless, we believe this account of SARMAC’s history might interest anyone with an enthusiasm for applied cognitive psychology.

Establishment

In 1978, the University of Cardiff hosted the International Conference on Practical Aspects of Memory (PAM). PAM was the first conference dedicated to applied memory research (D. Read, personal communication, January 18, 2016), a topic that was only beginning to reach the forefront of psychology (Hoffman & Deffenbacher, 1992; Neisser, 1982; Pezdek, Deffenbacher, Lam, & Hoffman, 2006). At this first meeting, Ulrich Neisser began the conference with his famous complaint that “If X is an interesting or socially important aspect of memory, then psychologists have hardly ever studied X” (Neisser, 1978, p. 2). Over the course of the conference, researchers attempted to challenge Neisser’s statement by presenting work on the topics of memory aids, the effects of stress on memory, individual differences in memory, educational applications, and clinical implications of memory (Grunenberg, Morris, & Sykes, 1978, 1988). Nearly a decade later, PAM II was held in 1987 at the University of Wales, Swansea. The growth of the applied memory field was evident: The number of presented papers doubled that of PAM I. Papers addressed 24 different topics, including adult and child witnesses, metamemory, ecological perspectives, neurological memory deficits, drugs and memory, dyslexia, student learning, clinical considerations for memory, and general educational implications (Grunenberg et al., 1988). Indeed, in his closing address, Neisser declared that his experience at the PAM II conference was “...like being a child turned loose in a candy store; on every side there has been something ecologically delicious” (p. 545). The seeds for a new society were sown at PAM II; it was here that members began to discuss the fact that the small and infrequent conferences did not sufficiently bring together a rapidly growing community of researchers.

In 1994, the University of Maryland hosted PAM III, which included 320 papers and posters (Payne & Conrad, 1997). The paper topics included mood-congruent memory, timing of autobiographical memories, laypeople’s beliefs about memory, individual differences in memory ability, clinical applications of memory, collaborative memory, prospective memory, eyewitness testimony, and memory for faces (Herrmann, McEvoy, Hertzog, Hertel, & Johnson, 1994; Payne & Conrad, 1997). The overarching goal of the PAM conferences was for psychologists to apply knowledge of memory and appropriate research methodologies to explain real-world phenomena, and to discuss methods for solving problems for real-world practitioners in memory-related fields (Grunenberg et al., 1978). Indeed, these PAM meetings were held during a period when, according to Michael Toglia (a founding member of SARMAC and former Executive Director), “basic cognitive research had much to offer that was not being sufficiently tapped” (personal communication, December 18, 2015). Don Read, another founding member, notes the “...separation between those primarily interested in application of research as compared to those primarily interested in answering basic questions about cognition and memory” (personal communication, December 20, 2015).

The PAM conferences also highlighted some of the differences within the applied memory field. One growing trend was the everyday-memory movement. The primary aim of this approach was to examine real-world functions of memory—as Neisser (1978) had suggested—through more naturalistic, ecologically valid paradigms instead of lab-based manipulations (Ericsson & Simon, 1980). In the preface for their collection of papers from PAM III, Payne and Conrad (1997) suggested that PAM I and II sparked a growth in the field of research on everyday memory and that texts from the first two PAM meetings had a major impact on the field. But many scientists had reservations about what the everyday-memory movement offered, particularly Banaji and Crowder. In a provocative and influential paper in the American Psychologist, Banaji and Crowder (1989) proffered that researchers should be extremely wary of sacrificing experimental control for ecological validity. They proposed a $2 \times 2$ array for research designs, with high- and low-generalizability on one axis, and high- and low-ecological validity of method on the other axis. They reasoned that the ideal research would be high on both axes; conversely, researchers would avoid designing research that is low on both metrics. When faced with the remaining two cells, however, Banaji and Crowder preferred a study high in generalizability with low ecologically valid or “contrived” methods to a study with high ecologically valid method and low generalizability. They argued that when a researcher sacrifices experimental control in favor of studying memory in an everyday context, little is contributed to our general understanding of memory. Banaji and Crowder’s (1989) criticism of everyday memory initiated a heated discussion, and led to a number of rebuttals published over the next several years (e.g., Baddeley, 1993; Bruce, 1991; Ceci & Bronfenbrenner, 1991; Conway, 1991; Gruneberg, Morris, & Sykes, 1991; Loftus, 1991; Roediger, 1991), including Klatzky’s (1991) “Let’s Be Friends.” Although respondents agreed that Banaji and Crowder (1989) had minimized the importance of everyday-memory research, they disagreed amongst themselves on how to strike an appropriate balance between the
directly applicable findings offered by naturalistic observation and the ability to isolate memory processes offered by controlled research in the laboratory. Of course, disagreements can be valuable to scientific progress. However, this disagreement illustrates an important point in the history of applied cognitive psychology: The everyday memory debate was waged inefficiently over the course of several years via published commentaries; a regular conference might have better facilitated these discussions.

Disagreement over methods was not the only challenge that applied cognitive researchers faced. Founding SARMAC members Herrmann and Gruneberg (1993) noted that basic and applied researchers disagreed on an assortment of issues that hindered communication between the two groups, ranging from publication customs (p. 559) to the proper way to draw scientific inferences (p. 559). Herrmann and Gruneberg also noted that applied researchers sometimes faced challenges in publishing their work, a perception shared by others in the field (M. Toglia, personal communication, December 18, 2015). Because of these challenges, Herrmann and Gruneberg made two recommendations just before PAM III. First, they suggested that applied researchers should be diligent about connecting their research questions to general theoretical frameworks and issues. Second, journals devoted to basic research should adjust their acceptance criteria to accommodate applied research methods and evaluate the applied research accordingly. Herrmann and Gruneberg’s concerns about the consideration of theory, the evaluation of applied research on its own merits, and the debate over the balance between experimental control and ecological validity illustrate the importance of regular scientific discussion. While not all cognitive psychologists shared Herrmann and Gruneberg’s views, there was clearly an opportunity for applied cognitive researchers to create a community for such discussion (M. Gruneberg, personal communication, March 18, 2016).

The PAM conferences were the first attempt at a forum to address the disagreements over the best way to conduct applied cognitive research. But PAM conferences soon outgrew their utility because they were a conference without an accompanying society. In the absence of an upcoming conference, there were few opportunities and no organizational context for pursuing these interests. As Steve Lindsay put it, “...the times were ripe for scientifically rigorous, theoretically grounded approaches to cognition in everyday life so there was a strong demand for a professional society that valued such work” (personal communication, December 8, 2015).

Talk of forming a society such as SARMAC began in 1993 at an informal gathering of applied researchers attending the meeting of Psychonomic Society in Washington, DC. The following year, SARMAC was officially founded with the society’s first membership drive organized by Douglas Herrmann, Chuck Thompson, Graham Davies, David Burrows, David Payne, Ron Okada, Roy Malpass, Michael Gruneberg, Darryl Bruce, Michael Toglia, and Don Read at PAM III. Many of the individuals who were involved in the PAM conferences would become SARMAC members, although, interestingly, no one seems to remember who came up with either the name or the acronym.

The aim of SARMAC was, and still is, to “promote applied memory and cognitive research, and to promote collaboration and communication between basic and applied research” (Gruneberg, 1997, p. 1). According to Toglia, SARMAC was unique in that it “began to shine a spotlight on applications of cognitive and cognition-related research, and, perhaps more importantly, shone a spotlight on the potential of such work” (personal communication, December 18, 2015). Paula Hertel, another early member of the society, said that SARMAC provided “…an organization for cognitive researchers—mainly memory researchers at the time—who were interested in problems with both theoretical and applied angles” (personal communication, December 21, 2015). In the first chapter of the published volume that detailed the presentations of SARMAC I, Bruce and Read (1998) also characterized the society as a place for researchers of eyewitness and autobiographical memory to bridge the gap between their topics, and use the combined findings to discuss implications for basic theories of memory.

SARMAC’s inception improved upon PAM conferences by creating an organization devoted to applied research in memory and cognition. The society began to host conferences every other year, and published regular newsletters that highlighted members’ newest publications and accomplishments—practices that continue today. SARMAC also appointed a Press Officer to reach out to non-members who could use applied research in non-academic pursuits. The society grew rapidly with over 200 members by 1997, demonstrating the appeal of its aims. SARMAC’s first Executive Officers were Doug Herrmann (Chair), Chuck Thompson (Chair-Elect), and David Burrows (Executive Director). The society’s first Board of Governors included Graham Davies, Marcia Johnson, David Rubin, Don Read, Kathy Pezdek, and Gene Winograd. Michael Toglia served as SARMAC’s first Press Officer, Darryl Bruce as the Membership Secretary, and Leslie Miller as the Newsletter Editor.

Membership

A review of SARMAC’s membership since 2010 (the earliest year available to us) makes it obvious that the society’s membership has grown in both numbers (increasing from 300 members in 2010 to 500 members in 2015) and diversity. For example, since 2013, approximately 20% of the society’s members have come from non-English speaking countries. Moreover, the number of nations represented in SARMAC has increased. In 2010, SARMAC members hailed from 21 different nations, a number that increased to 26 in 2013 and to 29 in 2015. Although the majority of members are from the United States, Canada, and the United Kingdom, countries throughout the Asian, European, and Australian continent are represented. (For a complete record of the membership data, see Table 1.) Generally, at any given time, the number of members from particular countries reflects the location of SARMAC conferences. For example, in 2013 Dutch membership peaked at 15%, likely due to the Rotterdam conference. Similarly, in 2015 when SARMAC was in Victoria, Canada, 23% of all members were Canadian. In order to meet its goal of being a truly international organization, SARMAC aims to increase representation and foster relationships with researchers around the world (K. Mori, personal communication, December 16, 2015).
development and education of SARMAC’s student members. The Board will delegate much of the responsibility for supporting student travel and awarding student research grants to the Caucus, and will work with the Caucus to develop additional programs and opportunities.

A hallmark of SARMAC is that even as the society grows, the camaraderie among its members remains. Researchers in the SARMAC community spend considerable time with one another as colleagues, collaborators, and friends over the course of their careers. Indeed, some members joined SARMAC as students; to them, SARMAC conferences can seem like reunions. Such camaraderie is an ideal way for students in particular to engage with their peers and established experts, potentially forming long-term collaborative relationships. As Kazuo Mori reflects, “SARMAC makes it easy to know and find people in related fields. But most importantly, I like SARMAC because of the members and the friendliness of the members. It’s such a nice society to be a part of” (personal communication, December 16, 2015).

Table 1
Percentage of Membership (Count) Distributed by Continent and Nation During the Past Six Years

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of members</td>
<td>312</td>
<td>392</td>
<td>500</td>
</tr>
<tr>
<td>North America</td>
<td>59.0% (184)</td>
<td>44.3% (174)</td>
<td>59.6% (298)</td>
</tr>
<tr>
<td>United States</td>
<td>47.4% (148)</td>
<td>37.0% (145)</td>
<td>36.2% (181)</td>
</tr>
<tr>
<td>Canada</td>
<td>11.5% (36)</td>
<td>7.4% (29)</td>
<td>23.4% (117)</td>
</tr>
<tr>
<td>Trinidad and Tobago</td>
<td>3% (1)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Europe</td>
<td>26.0% (81)</td>
<td>36.0% (141)</td>
<td>24.6% (123)</td>
</tr>
<tr>
<td>UK</td>
<td>14.1% (44)</td>
<td>18.6% (73)</td>
<td>12.2% (61)</td>
</tr>
<tr>
<td>Denmark</td>
<td>3.2% (10)</td>
<td>1.8% (7)</td>
<td>3.0% (15)</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>1.9% (6)</td>
<td>3.8% (15)</td>
<td>1.4% (7)</td>
</tr>
<tr>
<td>Other</td>
<td>6.7% (21)</td>
<td>11.2% (44)</td>
<td>7.8% (39)</td>
</tr>
<tr>
<td>Australia</td>
<td>6.7% (21)</td>
<td>11.2% (44)</td>
<td>8.8% (44)</td>
</tr>
<tr>
<td>New Zealand</td>
<td>4.5% (14)</td>
<td>3.3% (13)</td>
<td>2.6% (11)</td>
</tr>
<tr>
<td>Asia</td>
<td>3.5% (11)</td>
<td>5.1% (20)</td>
<td>4.6% (23)</td>
</tr>
<tr>
<td>Japan</td>
<td>2.8% (9)</td>
<td>4.3% (17)</td>
<td>3.8% (19)</td>
</tr>
<tr>
<td>Other</td>
<td>.006% (2)</td>
<td>.8% (3)</td>
<td>.8% (4)</td>
</tr>
<tr>
<td>South America</td>
<td>.002% (1)</td>
<td>.005% (2)</td>
<td>–</td>
</tr>
<tr>
<td>Africa</td>
<td>.003% (1)</td>
<td>–</td>
<td>.002% (1)</td>
</tr>
</tbody>
</table>

Approximately one-third of current SARMAC members are students, a fact that reflects the high value the society places on young scholars. To promote student involvement, students receive a heavily discounted membership fee and are encouraged to attend and present at SARMAC conferences. In 2015 at SARMAC XI, the SARMAC Executive Board recognized the large representation of student members by inviting student attendees to meet and form the first SARMAC Student Caucus. At that meeting, students elected a Founding Student Committee to serve until the next meeting in 2017. The primary aim of the Student Caucus is to enhance the professional

Table 2
Organization of Biennial Conferences

<table>
<thead>
<tr>
<th>Name</th>
<th>Dates</th>
<th>Host and location</th>
<th>Attendance</th>
<th>Organizing chair(s)</th>
<th>Keynote speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARMAC I</td>
<td>July 15–17, 1995</td>
<td>U of British Columbia, Vancouver, BC, Canada</td>
<td>150</td>
<td>Don Read</td>
<td>S. Ceci, D. Rubin</td>
</tr>
<tr>
<td>SARMAC II</td>
<td>July 11–13, 1997</td>
<td>Ryerson Polytechnic U, Toronto, ON, Canada</td>
<td>Unknown</td>
<td>John Turtle</td>
<td>D. Herrmann &amp; C. Yoder, M. Gruneberg, E. Tulving</td>
</tr>
<tr>
<td>SARMAC III</td>
<td>July 9–11, 1999</td>
<td>U of Colorado, Boulder, CO, USA</td>
<td>Unknown</td>
<td>Kathy Pezdek &amp; Chuck Thompson</td>
<td>G. Wells</td>
</tr>
<tr>
<td>SARMAC IV</td>
<td>June 12–14, 2001</td>
<td>Queen’s University, Kingston, ON, Canada</td>
<td>~135</td>
<td>Rod Lindsay</td>
<td>D. Payne, G. Wells</td>
</tr>
<tr>
<td>SARMAC V</td>
<td>July 2–6, 2003</td>
<td>Aberdeen University, Aberdeen, Scotland</td>
<td>~250</td>
<td>Fiona Gabbert, Rhiannon Ellis, Amina Memon, &amp; Lauren Shapiro</td>
<td>V. Bruce, R. Lindsay, E. Loftus, P. Ornstein, D. Rubin</td>
</tr>
<tr>
<td>SARMAC VIII</td>
<td>July 26–30, 2009</td>
<td>Kyoto, Japan</td>
<td>300</td>
<td>Yukio Itsukushima &amp; Kazuo Mori</td>
<td>R. Engle, C. Fine, P. Hertel, R. Kawashima, T. Matsuzawa</td>
</tr>
<tr>
<td>SARMAC X</td>
<td>June 26–29, 2013</td>
<td>Erasmus University, Rotterdam, NL</td>
<td>304</td>
<td>Katinka Dijkstra</td>
<td>D. Berntsen, G. Cumming, J. Joormann, J. Hibbing, E. Marsh, J. Wicherts</td>
</tr>
<tr>
<td>SARMAC XI</td>
<td>June 24–27, 2015</td>
<td>University of Victoria, Victoria, BC, Canada</td>
<td>&gt;400</td>
<td>Steve Lindsay &amp; Don Read</td>
<td>I. Dror, M. Johnson, H. Pashler, D. Simons</td>
</tr>
</tbody>
</table>
The first SARMAC conference was hosted by the University of British Columbia in 1995. This meeting had 150 attendees, 66 presentations, and two invited speakers. Since 1995, the SARMAC community has grown in size and influence. Conferences have been hosted around the globe in 11 cities across six countries, including Canada (1995, 1997, 2001, 2015), the United States (1999, 2007, 2011), Scotland (2003), New Zealand (2005), Japan (2009), the Netherlands (2013), and soon Australia (2017). SARMAC V, hosted by the University of Aberdeen in 2003, was the first meeting outside North America. It featured five keynote speakers, 27 symposia, 16 paper sessions, and 61 posters—considerable growth in eight short years and four conferences. SARMAC VIII was a milestone that marked the society’s international expansion—the first conference held in a non-English speaking country. SARMAC IV was also unique in that it was scheduled to be held in Spain, but the society was forced to reschedule after the host university backed out less than a year before the conference. Fortunately, Rod Lindsay, who was preparing to host SARMAC V in Kingston, Canada, at Queens University, agreed to host the conference two years earlier than planned. Despite the difficulties associated with planning a conference in nine months, SARMAC IV included 83 presentations and approximately 135 attendees (personal communication, R. Lindsay, March 17, 2016). This abrupt rescheduling also explains why SARMAC IV departed from trends in attendance, number of presentations, and topic diversity (see Figure 1).

Conference attendance reached its peak to date when over 400 people attended SARMAC XI in Victoria, British Columbia, Canada in 2015. Comparing these figures with SARMAC I shows the society’s growth in 20 years. Attendance has nearly tripled from 150 to over 400, and the number of presentations has grown fourfold, from 75 papers in 1995 to 348 papers, symposia, and posters in 2015. According to Read, who was involved in organizing both conferences, “this growth speaks to the considerable interest in the myriad topics subsumed by applied cognition and memory” (personal communication, December 20, 2015).

To understand how the type of research presented at SARMAC conferences has developed over the years, we reviewed the titles and abstracts (when available) of symposia and papers at each conference. We found several trends worth noting. First, every conference has included presentations on eyewitness and forensic research, autobiographical memory, and research on memory distortion; these are the standard topics at SARMAC. Other topics have varied, but they represent a wide variety of topics relevant to memory and cognition—such as developmental, linguistic and social processes, metacognition, mood and clinical phenomena, and methodological issues. Second, recent conferences have included a better balance between eyewitness and other forensic topics and all other topics; Figure 1 shows this trend. SARMAC IV’s disproportionately large representation of eyewitness and forensic-memory research was due to the short-notice change in conference location and organization, described previously.

This shift in the distribution of topics reflects the society’s aims to appeal to the wider community of researchers working in applied memory and cognition and is further evidenced by the increased diversity of SARMAC VI’s keynote topics. In Wellington, keynotes covered topics such as “The Power of Testing in Improving Educational Performance” (by Henry Roediger), “Mind Bugs” (by Mahzarin Banaji), and “Society does not do factor analysis: Limitations of the new WISC” (by Jim Flynn). At SARMAC VII, Craig Anderson spoke on “Violent Video Game Effects on Cognition, Affect and Behavior,” and even at SARMAC IX (hosted by the John Jay College of Criminal Justice), four of the five keynotes focused on topics other than legal issues, such as Carol Tavris’s keynote, “The Mysteries of Communicating Science.”

Third, the number of presentations at each conference has steadily increased. Posters in particular have seen a growth in popularity since they became a permanent fixture of conferences at SARMAC V (see Figure 2). As the conferences have grown larger, poster presentations have been integral in allowing researchers, particularly students, to showcase their work and participate in scientific discussion at SARMAC conferences.

![Figure 1](image1.png)  
**Figure 1.** Proportion of presentations (in symposia and individual sessions) in each general category.

![Figure 2](image2.png)  
**Figure 2.** Number of presentations at each SARMAC conference.
These trends demonstrate the importance of SARMAC conferences. According to Read, an increasing number of researchers and universities are bridging the gap between basic and applied research due to “a spreading belief that we are better able to reap the rewards of information and neuroscience technologies in methodology and analysis” (personal communication, December 20, 2015). As Read noted, there is “more energy directed toward application of research today than ever before.” Indeed, at SARMAC XI, Dan Simons returned conference-goers to this question: “If you can’t see how something can be applied, then what’s the point in researching it?” There is no doubt that SARMAC has been influential in channeling the energy Read referenced by providing conferences for sharing applied research.

Journal

Previously, two journals represented SARMAC: Applied Cognitive Psychology (ACP) and Cognitive Technology. These two journals served the society well for several years. As the society’s official journal, ACP included one Editor for SARMAC that was chosen by the society, a position held by Kathy Pezdek (1995–1999), Don Read (2000–2004), and Bob Belli (2005–2009). But as SARMAC continued to grow, the Executive Board became increasingly aware of the need for SARMAC to have its own official journal, allowing the Board to control its editorial format and bolster the financial health of the society. At first, several Board members sought a negotiated settlement that would preserve ACP as an official journal of SARMAC, but ultimately they were unsuccessful. With a blueprint for a new journal, Kathy Pezdek, Steve Lindsay, and then-President Maryanne Garry worked with the Association for Psychological Science Executive Director, Alan Kraut, and publisher Larry Erlbaum to develop a proposal. After multiple presentations to publishers pitching a journal that “presented applied cognitive research in a fresh new way” (K. Pezdek, personal communication, January 10, 2016), the group received an offer in 2009 from Elsevier. When Garry signed the contract on December 17, 2010, the Journal of Applied Research in Memory and Cognition (JARMAC) was born.

Although it has been just four years since its first issue in 2012, JARMAC is one of the society’s most visible successes, thanks in no small measure to the journal’s founding Editor, Ron Fisher. JARMAC offers several benefits to the society, beyond the obvious benefit of providing an outlet for communication. First, as the official publication of SARMAC, the journal increases the society’s visibility. Second, due to more favorable membership subscriptions, the journal has already produced greater financial stability for the society, and with revenue sharing to begin in 2017, SARMAC will be in a position to help cover conference expenses, stipends, and travel awards for students. Third, because JARMAC now requires a General Audience Summary for all empirical reports, it should be easier to share research findings and their implications with non-academic audiences.

Future Goals

As successful as SARMAC has been over the past 22 years, the Executive Board, led by Christian Meissner, has initiatives planned for strategic growth (personal communication, December 19, 2015). One such plan is to focus on recognizing membership, with an emphasis on student members. For example, Don Read’s retirement spurred the Board to establish an award in his honor, to recognize his long-standing support of the society and his commitment to help emerging researchers. The inaugural J. Don Read Early-Career Award, which recognizes outstanding early-career achievement, was awarded to Jason C. K. Chan. Chan’s research shows that initial retrieval can influence subsequent retrieval of information in positive and negative ways (e.g., Chan, Thomas, & Bulevich, 2009). His research reflects a core value of SARMAC in that it is theoretically rich, uses solid experimental designs, and ensures a context for application of the findings (C. Meissner, personal communication, February 16, 2015). Another example that illustrates the focus on membership is the establishment of a SARMAC Student Caucus. Other plans include recognition of outstanding student papers and posters. Meissner says that the ultimate goal is to make sure SARMAC continues to provide opportunities for its members, whether by celebrating its members’ successes, making it easier for students to attend SARMAC conferences, or by encouraging mentorship between young and established scholars (personal communication, December 19, 2015).

Due to the society’s successful conferences, quality research, and presence in the research community, we agree that “SARMAC really punches above its weight” (M. Garry, personal communication, December 11, 2015). Over the past two decades, SARMAC has established a viable organization for the exchange of research on applied aspects of memory and cognition that should continue to benefit cognitive research in the future.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Author Contributions

This article was written by the SARMAC Student Caucus Founding Executive Committee: William Crozier (John Jay College), President; Mario Baldassari (University of Victoria), Vice President; Ella Moek (Flinders University), Social Chair; Dawn-Leah McDonald (Victoria University of Wellington), Secretary; and Camille Weinsheimer (Simon Fraser University), Treasurer. All members of the committee contributed to conducting interviews, researching history, and preparing the manuscript. For more information about the SARMAC Student Caucus, you can find us on Facebook (www.facebook.com/sarmacstudents) or via email (sarmacstudentcaucus@gmail.com).

References


Make ’em Laugh? The Mnemonic Effect of Humor in a Speech
Mario J. Baldassari, University of Victoria
Matthew Kelley, Lake Forest College

ABSTRACT. The present study investigated the mnemonic effects of using a joke and the influence of the location of the joke within a speech. In Experiment 1, participants heard 2 passages—one beginning with a humorous limerick and the other with a nonhumorous one. In Experiment 2, the limericks were presented at the end of the passages. Across both experiments, humor enhanced memory only for the humorous limericks, $t(44) = 3.22, p < .01$; $t(35) = 2.59, p < .002$, and did not influence memory for the rest of the passage. These results were discussed in context of the current literature, and future directions for study were outlined.

An elderly husband and wife visit their doctor when they begin forgetting little things. Their doctor tells them that many people find it useful to write themselves little notes. When they get home, the wife says, “Dear, will you please go to the kitchen and get me a dish of ice cream? And maybe write that down so you won’t forget?”

“Nonsense,” says the husband, “I can remember a dish of ice cream.”

“Well,” says the wife, “I’d also like strawberries and whipped cream on it.”

“My memory’s not all that bad,” says the husband. “No problem—a dish of ice cream with strawberries and whipped cream. I don’t need to write it down.” He goes into the kitchen for a while; his wife eventually hears pots and pans banging around. The husband finally emerges from the kitchen and presents his wife with a plate of bacon and eggs.

She looks at the plate and asks, “Hey, where’s the toast I asked for?”

Many guides designed to enhance presentation skills suggest that speakers should start their speeches with a joke in an effort to gain the attention of the audience and to make the presentation more memorable (e.g., Jeary & Cottrell, 2008; Kosslyn, 2007). The present study was designed to assess whether the presence of humor in a speech impacted memory for both the humorous stimulus (i.e., joke) as well as the surrounding content (i.e., speech).

Past research has shown that the cognitive processes involved in humor generally enhance memory for humorous events or material (e.g., Schmidt, 1994; Worthen & Deschamps, 2008). Most commonly, researchers have developed humorous stimuli in the form of sentences, pictures, and lectures (e.g., Kaplan & Pascoe, 1977; Schmidt & Williams, 2001). For instance, Schmidt (1994) assembled a number of humorous sayings and created nonhumorous counterparts (controls) for them. As an example, consider the humorous quote, “The only way to keep your good health is to eat what you don’t want, drink what you don’t like, and do what you’d rather not,” and its nonhumorous revision, “The only way to keep your good health is to eat good food, drink healthy drinks, and do healthy activities” (Schmidt, 1994, p. 954). More complex manipulations have integrated humor into hour-length lectures while manipulating the degree to which the humor related to the topic at hand (Kaplan & Pascoe, 1977). Although it seemed straightforward, the manipulation of the humor was quite difficult because humor is not universal—what is funny to one person may not be funny to the next (Kuiper et al., 2010). Humorous stimuli, therefore,
were carefully normed and pretested to ensure wide-ranging humor responsiveness before being used experimentally.

Researchers interested in the mnemonic effects of humor considered a wide variety of design issues, such as: (a) will participants be informed of the impending memory test (informed/intentional learning vs. uninformed/incidental learning) and (b) should humor be manipulated between-subjects (unmixed list) or within-subjects (mixed list)? For instance, Schmidt (2002) and Takahashi and Inoue (2009) explored these design issues and found that humor was more likely to elicit an effect with incidental learning, but that effects were also seen with intentional tasks. Although early work (e.g., Schmidt, 2002) seemed to suggest that humor effects were only present in mixed-list designs, where funny material could stand out from the context of unfunny material, more recent work has shown mnemonic benefits to humor with both mixed and unmixed lists (Takahashi & Inoue, 2009). That said, the standard design exploring humor and memory employs an incidental learning task with mixed lists (e.g., Schmidt, 1994; Schmidt, 2002).

Researchers have attempted to explain the mnemonic benefit of humor by appealing to a variety of related mechanisms. One explanation is that the surprise felt by encountering humorous, distinctive, or even bizarre material might lead to deeper (more meaningful) processing, greater elaboration or connection with long-term memory, additional rehearsal (or retelling) of the material, or even the storage of additional contextual information with the unique stimulus—each of which is sufficient to improve memory for such information (Schmidt, 2002). Other explanations have suggested that the mnemonic benefit of humor stems from the greater emotional or physiological arousal that accompanies humor.

In particular, the encoding-based explanations suggested that humorous materials were given additional rehearsal, elaboration, or privileged storage during encoding (e.g., Schmidt, 2002; Takahashi & Inoue, 2009). One implication of these explanations was that the extra rehearsal or elaboration for humorous stimuli would likely be at the expense of surrounding material. In terms of the present research question, this theory might suggest the presence of humor in a speech would negatively impact memory for speech itself. Evidence from Kaplan and Pascoe (1977) caused doubt in this explanation as they found that the material following humorous stimuli was remembered no better or worse than material following nonhumorous stimuli. However, one study does not provide a definitive test of a theory.

In contrast, retrieval-based explanations of the humor effect suggested that humorous materials were easier to distinguish from nonhumorous material in memory—that is, they were more distinctive (Schmidt, 2002). One could argue that the prevalence of humor benefits with mixed-list designs clearly support this theory—presumably, the humorous material stood out from the background of nonhumorous material. However, the presence of humor effects with unmixed designs was potentially problematic, unless one viewed the humorous material as standing out from all other information in memory, most of which is not likely to be humorous. Returning to the present question of interest, if humor influenced memory at the point of retrieval through distinctiveness, one would expect the mnemonic benefit to remain solely with the jokes and not with the surrounding speech information. It was difficult to tell from the current literature whether a retrieval-based account would predict a decline in memory for the speech information, as predicted by the encoding-based accounts.

The goal of the present study was to explore whether there was any truth to the old adage that starting a speech with humor would render the speech more memorable. The major explanations seem to predict that, while memory for the joke would be enhanced, the memory for the surrounding material would be either unaffected or impaired by the presence of humor. Findings in this direction would call into question the utility of starting with a joke.

**Experiment 1**

In Experiment 1, participants listened to two different passages of factual information. One passage began with a humorous limerick, whereas the other began with a nonhumorous limerick. To better simulate the conditions under which one might hear a speech, participants were given a brief orientation statement that asked them to imagine they were about to listen to a keynote speaker at a relevant conference before they actually listened to the limerick and passage. Following presentation, participants completed a brief distracting activity and then completed a recall and recognition test for all of the presented information. Based on previous research, we predicted that the humorous limerick would be remembered better than the nonhumorous limerick—the standard humor
benefit. Further, based on the encoding-based explanations of the humor effect, we predicted that memory for the passage might be impaired, especially the content that directly follows the limerick.

Method
Participants. Forty-five (6 men, 39 women) undergraduates were recruited from an Introduction to Psychology course at Lake Forest College and were compensated with extra credit in the course. Participants were run in groups of one to four people and completed the task at separate computers.

Materials and Design. Prior to the experiment, we created 16 limericks with ideas from Internet sources (i.e., Jokes.com, Laughoutloud.net) and modified them to create four types of limericks: funny and taboo, funny and not taboo, not funny and taboo, and not funny and not taboo. For example, a funny-taboo limerick began, “There once was a man named Dave. He had a dead whore in a cave,” whereas its nonfunny-taboo equivalent would start, “There once was a corporate whore named Dave. His investment plan was brave.”

In order to assess whether the categories were valid (humorous limericks were actually humorous; taboo limericks were taboo), 16 first-year students at Lake Forest College rated how funny and how taboo each limerick was on a simple 4 point Likert-type scale (1 = funny/taboo, 4 = not funny/not taboo). From 16 limericks, we selected two funny (mean rating: 3.29, taboo: 3.26) and two not funny (mean rating: 1.17, taboo: 1.06) versions that were matched on humor and taboo ratings (see Table 1). The four limericks were coded 1F (1 Funny), 1NF (1 Not Funny), 2F, and 2NF and appear in Table 1.

We found that limericks with taboo or inappropriate language were the only limericks rated as funny, so the second original intended manipulation (presence of possibly taboo words in the limerick) was instead held constant by using the same taboo words in both the funny and the not funny limericks. Due to limerick 1NF’s low rating on the taboo scale, the second line was changed from “Who felt like an old corporate whore” and the fifth line was changed from “Would give her a job at the store” to “Would hire her to cut meat at the store” to add the same taboo features as Limerick 1F. Due to limerick 2NF’s low rating on the taboo scale, the word “shit” was added to the fourth line to create the same taboo effect as Limerick 2F.

Two content passages were created: one passage discussing the Spanish economy and the other outlining the recent popularity of Hollywood celebrities to work as voice actors for large commercial campaigns (see Table 2). The subject matter was decided based on simplicity of the topics and potential relation to the limericks. Each was designed to relate to the subject matter of a different limerick. The “Woman from Moher” limericks (1F; 1NF) flowed with the Spanish economy passage because both discuss issues surrounding job availability, whereas the “Man named Dave” limericks (2F; 2NF) related to the money theme in the Hollywood celebrity endorsements passage. The relationships are purposefully weak so that limerick memory and passage memory could not be confused at test. Both passages were of similar length (73 and 80 words, 4 and 5 sentences, respectively).

Across participants, each limerick was paired equally often with each passage and was placed at the beginning, before the passage content. Four audio recordings were created for the experiment. Each recording contained one limerick and one passage read in the same voice with natural inflection; the same male speaker was used in each recording.

Each sentence in each passage was converted into a series of propositions—singular idea units with truth value—for the purposes of scoring participants’ recalls. Overall, each passage contained 12 idea units. The limericks were also separated into idea units to ensure consistent scoring.

We created a two-alternative, forced-choice recognition test for the limericks and passages. Distractor sentences were created for the recognition test. Small details were changed in each sentence that altered its meaning. For example, “Many unemployed citizens have turned to the Spanish under-market, taking jobs that are ‘off the books’

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limericks With Pilot Ratings</td>
</tr>
<tr>
<td>1F (Funny rating: 3.39, original Taboo rating: 3.17)</td>
</tr>
<tr>
<td>There once was a woman from Moher, who realized she could work no more. She took to the street. A guy flashed her his meat and decided to be a whore.</td>
</tr>
<tr>
<td>1NF (Funny rating: 1.28, original Taboo rating: 1.06)</td>
</tr>
<tr>
<td>There once was a woman from Moher, who felt like an old corporate whore. She had no idea her older friend Mia would hire her to cut meat at the store.</td>
</tr>
</tbody>
</table>

Baldassari and Kelley
TABLE 2

<table>
<thead>
<tr>
<th>Table 2: Passages for Experiments 1 and 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Passage 1</strong></td>
</tr>
<tr>
<td>As a result of the economic downturn in Spain, small businesses have closed and large companies have laid off workers across the country. Many unemployed citizens have turned to the Spanish under-market, taking jobs that are off the books and are paid ‘under the table.’ After a series of bank reforms, Spain appears poised to rise out of the recession. However, the growth of the under-market will likely slow the economic recovery process.</td>
</tr>
<tr>
<td><strong>Passage 2</strong></td>
</tr>
<tr>
<td>After the recent changes of the culture in Hollywood, many famous actors have felt no shame in doing large commercial campaigns for big corporations. The majority has gone to voice-over work, but a few actually appear in the commercials. Of course, they are paid more to appear than for voice-overs. Up until recently, it was viewed as “selling out” to take a paycheck for doing nothing more than using fame to sell a product. Nowadays, it’s viewed as good marketing.</td>
</tr>
</tbody>
</table>

Results and Discussion

As seen in Figure 1, the mean proportions of the propositions accurately recalled in the free recall portion of the experiment for the funny and nonfunny limericks were .60 (SD = .32) and .43 (SD = .23), respectively. A paired-samples t test showed that the difference between the means was statistically significant, t(44) = 3.22, p < .002, Cohen’s d = 0.61. The funny limericks, therefore, were remembered better than the nonfunny limericks. A separate paired-samples t test showed no significant difference between recall performance...
for the passages, $t(44) = -0.97, p = .34$.

Figure 2 displays the recognition accuracy for the funny and nonfunny limericks and passages. Paired-samples $t$ tests found no significant differences between the proportions of sentences correctly recognized for limericks, $t(44) = -0.01, p = .99$, or passages; $t(44) = -0.35, p = .73$. Performance appeared to be near ceiling.

Although, overall, the type of limerick did not appear to influence recall or recognition of the passage, it is possible that the influence of the limerick might have been restricted to just the first few propositions of the passage. Figure 3 shows recall accuracy for the first three propositions following each type of limerick. A series of paired-samples $t$ tests showed no significant differences between any adjacent propositions, suggesting that the limerick (funny or nonfunny) had no influence on memory for the adjacent material, $t(44) = 0.65, p = .52$; $t(44) = .53, p = .60$; $t(44) = .53, p = .60$ for first three propositions after the limerick, respectively.

Participants were asked to rate the relationship between the limerick and passage, whether the limerick increased their interest in the passage, and whether the limerick distracted them from the passage on a 1–5 scale. None of the ratings differed significantly between funny and nonfunny conditions, $t(44) = -1.61, p = .12$; $t(44) = .18, p = .86$; $t(44) = .98, p = .33$, relation, interested, and distraction listed respectively.

In Experiment 1, we successfully replicated the standard humor benefit by showing that funny limericks were remembered via free recall significantly better than nonfunny limericks. Further, the humor of the limericks in this case had no visible effects on memory for the rest of the passages. Participants also did not answer the follow-up questions significantly different across humor conditions. The results, then, do not support the old adage that starting a speech with a joke will render the speech more memorable because the memory enhancement was restricted to the joke.

**Experiment 2**

Experiment 2 was designed to assess whether the effect of the humorous limerick depended on its location within the passage, and whether the primacy location was the only reason for the humor effect seen in Experiment 1. Experiment 2 followed the same basic format as Experiment 1 except that the limerick was presented at the end of the passage. Based on the first experiment’s results, we expected to see the same effects shown
in Experiment 2—that is, a humor benefit for limerick content only and no effects on adjacent material or any of the rest of the passage material.

**Method**

**Participants.** Thirty-seven (27 women, 10 men) undergraduates were recruited from an Introduction to Psychology course at Lake Forest College and were compensated with extra credit in the course. Participants were run in groups of one to four people and completed the task at separate computers.

**Materials and Design.** The same limericks and passages used in Experiment 1 were used for Experiment 2; they appear in Tables 1 and 2. The recordings from Experiment 1 were edited such that the limerick followed the passage rather than leading into it in both the audio recording and the visual display. The same propositions, free recall and recognition instructions, were used in Experiment 2 as well. To ensure that the experiment was completely automated, the experiment was created using E-Prime and conducted on IBM-compatible computers.

**Procedure.** The procedure used was the same as that in Experiment 1.

**Data Analysis.** A rater assessed recall performance by determining which of the propositions in each limerick/passage combination were accurately remembered. The rater used lenient scoring criteria in which participants were awarded credit if they recalled the general idea of the proposition. The rater discussed any uncertain cases with a fellow researcher until consensus was reached and kept a running list of “acceptable phrases” to ensure consistent scoring of propositions throughout the process. For analysis, the mean proportion of propositions correctly recalled was calculated for accurate comparison between limericks and passages and among the different limericks.

**Results and Discussion**

One participant’s data set was removed from the analysis because of a procedural mistake—the participant received only one level of the manipulation. Hence, all subsequent analyses were performed using 36 participants. As seen in Figure 4, the mean proportions of the propositions accurately remembered in the free recall portion of the experiment for the funny and nonfunny limericks were .51 (SD = .12) and .36 (SD = .14), respectively. A paired-samples t test showed that the difference between the means was statistically significant, t(35) = 2.59, p = .014, Cohen’s d = 1.15.

The funny limericks were again remembered better than the nonfunny limericks. A paired-samples t test showed no significant difference between the mean proportion correct for the passages, t(35) = .03, p = .98.

Figure 5 displays the recognition accuracy for the funny and not-funny limericks and passages. Consistent with Experiment 1, paired-samples t tests found no significant differences between the proportions of sentences correctly recognized for limericks, t(35) = -1.59, p = .12, or passages, t(35) = -1.19, p = .24.

Again, although the type of limerick did not appear to influence recall or recognition of the passage overall, it is possible that the influence of the limerick might be restricted to just the last few propositions of the passage in this condition. Figure 6 shows recall accuracy for the last three propositions preceding each type of limerick. A series of paired-samples t tests showed no significant differences between any adjacent propositions, suggesting that the limerick was the only part of the stimulus affected by its humor or lack thereof for the last 3 propositions, in order: t(35) = 0.00, p = 1.0; t(35) = 1.44, p = .15; t(35) = -0.77, p = .45.

Only the comparison between conditions of the question, “Did the limerick distract you from the passage that preceded it?” showed a significant difference, t(35) = 2.19, p = .04, Cohen’s d = .50. The limericks were rated as significantly more distracting in the funny condition (M = 4.22, SD = .84) than in the nonfunny condition (M = 3.69, SD = 1.25). None of the other ratings differ significantly between funny and nonfunny conditions, t(35) = -0.82, p = .42; t(35) = -1.07, p = .29.

In Experiment 2, the standard humor benefit was successfully replicated by showing that funny limericks were remembered via free recall significantly better than nonfunny limericks. Consistent with Experiment 1, the presence of humor had no visible effects on memory for the rest of the passages. Participants did not answer significantly different across humor conditions with answers to the follow-up questions, although they did report feeling that the funny limericks were significantly more distracting. The results, again, do not support the old adage that starting with a joke will enhance the memorability of the speech.

**General Discussion**

Experiment 1 revealed a significant positive effect of humor on memory only when free recall was tested. Any residual effects (positive or negative) of
humor on the rest of the passage were not statistically significant. The humor benefit was not evident with the recognition test, which likely occurred due to a ceiling on performance levels. Additionally, the presence of humor at the beginning of a speech did not influence participants' interest in the passage and the presence of humor did not seem to distract participants. These results were replicated in Experiment 2, even when the location of the limerick was changed to the end of the speech.

These experiments demonstrate that the basic humor benefit is replicable and robust, even with a novel procedure and stimuli. Though there was no influence of humor on the passages themselves, they were not manipulated for humorous content. They are still, however, an important part of the stimulus, as the lack of difference between memory for the passages in the humorous and nonhumorous conditions adds interesting information to the presence of the effect in this study. Our results are consistent with those reported by Kaplan and Pascoe (1977), who showed a humor effect for humorous examples given within a lecture but only for those specific examples and not for the actual lecture material.

Some other strengths of the current study are that it used a less typical methodology in that participants learned intentionally (the memory test was not a surprise) and another more typical method in that the humor condition was manipulated within-subjects and produced the effect. The current study also used the propositional method of scoring data, which is a more objective and operationalized way to score recall data than by number of words within a sentence (Kintsch, 1974). Moreover, one could also argue that the nonfunny limericks were quite bizarre, and the lack of memory enhancement for them provides evidence of the strength of the humor effect over the bizarreness effect (and any taboo effects on memory due to the counterbalancing of taboo words across limerick type).

The results of this study provide a challenge to the encoding-based explanations of the humor benefit which seemed to predict that the extra attention (or rehearsal) paid to the humorous material would be at the expense of the immediately surrounding material (e.g., Schmidt, 2002; Takahashi & Inoue, 2009). The results, however, may be consistent with a retrieval-based explanation of the humor effect, although these theories do not directly address memory for surrounding information.

One possible weakness of this study is the

FIGURE 4
Proportion Correct Free Recall as a Function of Humor and Stimulus Type From Experiment 2.

FIGURE 5
Proportion Correct Recognition as a Function of Humor and Stimulus Type From Experiment 2.

FIGURE 6
Proportion Correct in Free Recall as a Function of Humor and Proposition From Experiment 2.
lack of a manipulation check on the humorous material for each participant. In other words, the study did not include a check to see whether each limerick was funny or nonfunny as intended for each participant. Such a check was avoided in order to keep the manipulation unknown to participants until after the experiment was over. However, the pilot study pretested the humor of the limericks, so the manipulation was nonetheless expected to work and did work, as a humor effect was shown.

Another potential limitation of the current study is the lack of a strong content relationship between the limerick and its passage. While leading with information couched in a joke might only help memory for basic concepts, such enhancement would be very helpful for the most critical information in a presentation or lecture situation, and so a more closely related limerick might improve general concept memory in the desired manner. Similarly, the use of a shorter joke might also make the nature of the experiment less obvious and add to the strength of the effect. The short length of the stimuli may also limit the generalizability of the study to presentations or lectures made in a classroom or at a convention. In such real-life situations, a speech would last much longer than one paragraph, so the manipulation necessary to show a humor effect might need to be stronger if it is even, in fact, possible for a joke to affect memory for content of a speech or paragraph in a different study or situation.

Further, the type of joke employed here could also have limited generalization, because limericks are a very specific and obvious type of joke. We chose limericks because of the ease of matching them to each other and the restrictions put on this study by time and resources. However, the length and type of the joke did not appear to be a major issue because the effect was produced twice. Finally, the taboo words, though repeated in all limericks, may not have maintained taboo meanings in different contexts (specifically in 2F/2NF) and were not retested for taboo status after the addition of new words. Despite these limitations, the content remembered in free recall, that the humor of the limerick, was the important manipulation in all conditions, not its content of taboo words.

Future Directions

The present experiments could potentially open up some new possibilities for the field of research on the humor benefit. First, it might be interesting to control for individual difference in humor perception by creating stimuli that are tailored to each individual’s sense of humor. Although difficult, such a study could add much information to our knowledge of the effect as researchers could be more sure that each participant would find the stimuli humorous. It would also address a potential limitation of the current study, which is that even though the stimuli were rated as funny, they were not rated as highly funny. Despite the possibility of a more robust effect (or an effect on the passage) with funnier stimuli, the current study did replicate the effect quite successfully.

Next, it would be nice to examine the effect of humor in more real world situations (e.g., Kaplan and Pascoe, 1977). If researchers are to answer the question of how to truly remember things better, they must first test the theories in situations like lectures and speeches. For instance, one might show participants videos of lectures or presentations. Indeed, initially this research began with a pilot test of a funny and a nonfunny video of a short speech, but the experiment never got past the piloting stage because the funny stimulus was not rated as sufficiently funny. However, future research using the recall methods here and the propositional method of scoring might lead to further validation of the effect and suggestions for ways to use it in the real world.

References


