

# Confidence and accuracy of lineup selections and rejections: Postdicting rejection accuracy with confidence

Citation for published version (APA):

Lindsay, R. C. L., Kalmet, N., Leung, J., Bertrand, M. I., Sauer, J. D., & Sauerland, M. (2013). Confidence and accuracy of lineup selections and rejections: Postdicting rejection accuracy with confidence. *Journal of Applied Research in Memory and Cognition*, 2(3), 179-184.  
<https://doi.org/10.1016/j.jarmac.2013.06.002>

## Document status and date:

Published: 01/09/2013

## DOI:

[10.1016/j.jarmac.2013.06.002](https://doi.org/10.1016/j.jarmac.2013.06.002)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.



## Target article

## Confidence and accuracy of lineup selections and rejections: Postdicting rejection accuracy with confidence



R.C.L. Lindsay<sup>a,\*</sup>, Natalie Kalmet<sup>a</sup>, Jaime Leung<sup>a</sup>, Michelle I. Bertrand<sup>b</sup>, James D. Sauer<sup>c</sup>, Melanie Sauerland<sup>d</sup>

<sup>a</sup> Queen's University, Canada

<sup>b</sup> University of Winnipeg, Canada

<sup>c</sup> University of Portsmouth, United Kingdom

<sup>d</sup> Maastricht University, Netherlands

## ARTICLE INFO

## Article history:

Received 14 September 2011

Received in revised form 31 May 2013

Accepted 18 June 2013

Available online 28 June 2013

## Keywords:

Confidence-accuracy

Eyewitness identification

Showup

Choosing

Rejection

## ABSTRACT

Correlation and calibration approaches show meaningful, positive confidence-accuracy relations for witnesses making selections from lineups, but rarely for rejections (Brewer & Wells, 2006; Sauerland & Sporer, 2009). This disparity may reflect the difference between selecting a single photo versus rejecting a set of photos. Participants ( $N = 101$ ) in two experiments made selections from and rejections of lineups in situations requiring either a single confidence rating about a single face (typical of “choosers”) or a single confidence rating about multiple faces (typical of “nonchoosers”). Mean confidence ratings were significantly higher for accurate versus inaccurate decisions for both selections and rejections when decisions were based on single faces. Single decisions about multiple faces produced no significant difference in confidence between correct and incorrect rejections but a significant difference for selections.

© 2013 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.

Eyewitness confidence is significantly related to the accuracy of lineup selections (Leippe & Eisenstadt, 2007; Sporer, Penrod, Read, & Cutler, 1995), and is an important determinant of belief in eyewitness testimony (Boyce, Beaudry, & Lindsay, 2007) which partly explains why misidentifications are associated with wrongful convictions (The Innocence Project, n.d.). Selections produce low-to-moderate positive correlations between confidence and accuracy (Leippe & Eisenstadt, 2007). Such confidence-accuracy (CA) relations for choosers are found with both simultaneous and sequential lineups, with little or no indication the relations differ based on lineup type (Brewer, 2006; Lindsay & Wells, 1985; Sporer, 1993). Significant CA relations are rarely found for lineup rejections; that is, when witnesses reject the entire lineup (Sauer, Brewer, Zweck, & Weber, 2010; Sauerland & Sporer, 2007, 2009; Smith, Lindsay, & Pryke, 2000; Sporer, 1993). This pattern of results for selections and rejections also occurs with mean comparisons (Sauerland & Sporer, 2009; Sporer, 1992; for exceptions see Sauerland, Sagana, & Sporer, 2012) and with calibration analyses (Brewer & Wells, 2006; Sauer et al., 2010; Sauerland & Sporer,

2009), even though calibration can detect CA relations when correlations are weak (Brewer, Keast, & Rishworth, 2002).

Although a disparity in the relation between accuracy and confidence exists between selections and rejections, there is a dearth of research investigating why the disparity exists, which is problematic as lineup rejections make up 27–40% of real-world witness decisions (e.g., Behrman & Richards, 2005). Legal practitioners often disregard lineup rejections when deciding whether a suspect is culpable (Clark & Wells, 2008; McAllister & Bregman, 1986, 1989; Wells & Lindsay, 1980). However, if correct, a lineup rejection indicates that the investigators must pursue alternative lines of enquiry in order to apprehend the culprit. Thus, identifying reliable indices of rejection accuracy is a matter of practical importance.

With selections (i.e., identification attempts), the witness compares the selected face to the face in memory (a one-to-one comparison; Weber & Brewer, 2006). In this situation, accuracy and confidence likely reflect how closely the lineup member resembles the image in memory (Stretch & Wixted, 1998; Wixted & Gaitan, 2002; Wixted & Mickes, 2010), though there may be limits on this process for particularly strong memories (Mickes, Hwe, Wais, & Wixted, 2011). However, when lineups are rejected, witnesses must compare each of the faces in the lineup to the one held in memory (a many-to-one comparison). In this situation, witnesses' estimates of confidence may reflect the confidence with which the best match to the image in memory is rejected, perhaps generating

\* Corresponding author at: Department of Psychology, Queen's University, Kingston, ON, Canada K7L 3N6. Tel.: +1 613 533 2880; fax: +1 613 533 2499.

E-mail address: [rod.lindsay@queensu.ca](mailto:rod.lindsay@queensu.ca) (R.C.L. Lindsay).

a low level of confidence. Alternatively, confidence may reflect how easily the majority of lineup members were rejected, such that easy rejection of mediocre fillers may lead to a relatively high level of confidence. Finally, the confidence rating may reflect the average confidence with which fillers were rejected, producing intermediate levels of confidence (Weber & Brewer, 2006). Of course, there is no reason that the same strategy for estimating confidence in lineup rejections will be employed across witnesses, lineups, or occasions. As a result, confidence ratings of lineup rejections should have an inconsistent relation with accuracy (Weber & Brewer, 2006).

If this logic is correct, making decisions in a one-to-one context leads to stronger CA relations than making decisions in a many-to-one context because the evidential bases for response and response confidence are directly linked. Poor CA relations for rejections result from making a single confidence judgment to reflect confidence in multiple rejection decisions. We tested this hypothesis by manipulating whether selections and rejections occurred in either a one-to-one or a many-to-one context. To create one-to-one and many-to-one situations for all decisions, a face recognition paradigm was used to present lineups either sequentially or simultaneously (Lindsay & Wells, 1985). Participants were required to remember sets of four targets per trial. At recognition, participants were presented with individual faces, with confidence assessed after each selection/rejection decision resulting in one-to-one decisions. To create a many-to-one context for all decisions, participants were shown large arrays of faces and required to state how many targets were present in the array, and provide a single confidence estimate for each of those decisions (Experiment 1). Finally, to determine if single presentation versus one-to-one decisions was the critical factor, participants were shown simultaneous arrays but required to make separate responses for each face (Experiment 2).

**Hypothesis 1.** When participants reject single faces, confidence will be greater when the decisions are correct than when the decisions are incorrect (consistent with the often reported positive correlation between confidence and accuracy for choosers).

**Hypothesis 2.** When participants make a single decision to reject multiple faces from a many-to-one presentation, confidence will be unrelated to decision accuracy (consistent with the frequently reported lack of CA relation for nonchoosers).

**Hypothesis 3.** Correct selections will be made with greater mean confidence than incorrect selections in the one-to-one situation.

**Hypothesis 4.** The difference in mean confidence of correct and incorrect selections will be smaller in the many-to-one than the one-to-one situation.

## 1. Method: overview

### 1.1. General methodology

We reasoned that the positive CA relationship for selections is due to a single confidence judgment being attached to a single item-to-memory (one-to-one) comparison, while the lack of a positive CA relationship for rejections is due to a single confidence rating being attached to multiple item-to-memory (many-to-one) comparisons. To manipulate this difference in comparisons, participants were asked to make decisions and confidence judgments about multiple faces (mimicking the multiple-face decision currently inherent in rejections of simultaneous lineups) or single faces (mimicking the single-face decision currently inherent in selections).

### 1.2. Participants

Across the two experiments, 101 undergraduate students participated (83 females, 18 males; ages ranged from 17 to 43 years,  $M = 19.58$ ,  $SD = 3.83$ ). They were of European (81.2%), Asian (13.9%), Middle Eastern (3.0%), and African (2.0%) ancestry. Participants were compensated with a chocolate bar and either course credit (Introductory Psychology students) or cash (\$10 per hour).

### 1.3. Materials

#### 1.3.1. Targets

Forty targets were presented in 10 sets of four faces. The targets were 20 males and 20 females of European (22), East Asian (10), South Asian (3), and African (5) ancestry. Sets of targets were created to minimize "confusability" within a set (e.g., European male, European female, African female, Asian male). At exposure, targets were shown in a horizontal row in front of a background scene with no other visible people. Targets were either smiling in a frontal pose, or with a neutral expression in a 3/4 pose, with three variations across sets (all smiling, all 3/4 pose, or two of each). Each set of target faces was visible via computer monitor for 12 s. Exposure was constant across conditions.

#### 1.3.2. Lineups

Each target was associated with a 6-person target-present (TP) and a 6-person target-absent (TA) lineup. Lineup members faced front, with neutral expressions, on a uniform-colored background. Faces seen during recognition differed from those seen during encoding in the following ways: background, size (slightly smaller at recognition), and either expression or pose. Differences between exposure and recognition increase the probability that selection is based on facial recognition rather than image-matching. Lineup members were visible only from the neck up and matched the targets on important physical dimensions (e.g., sex, ethnicity, hair color, etc.). Position of targets in lineups was balanced across targets but was always the same for specific targets. Variations existed such that 0, 1, 2, 3, or all 4 targets were present. Across participants, the presence versus absence of all targets was balanced.

For many-to-one presentation, the four 6-person lineups for each target set were presented at the same time on a large (140 cm) screen, resulting in a total of ten 24-person arrays. Each face in the array was labeled from 1 to 24 to facilitate identification decisions. Each of the four 6-person lineups filled one of the four quadrants of the screen (upper-lower, left-right). Position of the lineups on screen and lineup members within the lineups were constant within a set but randomly assigned across sets.

For one-to-one presentation, photos were shown individually. For this condition, the order of presentation corresponded to positions 1–24 in the many-to-one condition. The top-left lineup in the many-to-one array was presented first (lineup members 1–6), followed by the top-right (lineup members 7–12), bottom-left (lineup members 13–18), and bottom-right (lineup members 19–24). This order was constant across all 10 sets. The size of the individual face images was identical in both presentation conditions and the large area surrounding the individual faces in the one-to-one condition was blank (except for the response options) and a pastel color to prevent excessive brightness and eye fatigue.

#### 1.3.3. Intervening task

For each set of targets, participants performed an intervening task between exposure to the targets and the recognition task. The task involved a complex, cartoon beach scene (Where's Waldo<sup>1</sup>) of

<sup>1</sup> TM & © 2008 Entertainment Rights Distribution Limited. All rights reserved.

numerous characters engaged in a wide variety of activities. While the image was on screen (30 s), participants answered questions such as “How many people are wearing red bathing suits?” Participants answered as many questions as they could each time the image was presented. The intervening task ended automatically after the allotted time.

#### 1.4. Procedure

Participants were directed to a computer, read a letter of information describing the study, and indicated their consent to participate. Sets of targets were presented in random order. Each participant completed 10 trials, with each trial consisting of the presentation of a set of to-be-remembered targets (exposure), followed by the 30 s intervening task, then the 24-person lineup, which included 0, 1, 2, 3, or all 4 of the targets seen at exposure. Each combination (0–4 targets present) was seen twice by each participant. Each many-to-one condition trial required a single decision indicating the number of targets present, a rating of the participant's confidence in that decision, and the subsequent identification of an appropriate number of specific faces from the lineup (e.g., if the participant said two targets were present, then the participant identified two specific faces). The one-to-one condition trials started with identical exposure and intervening tasks. Recognition consisted of 24 decisions of whether or not each individual face was one of the targets, each immediately followed by a confidence rating. All confidence ratings were made on an 11-point scale that ranged from 0% (not at all confident) to 100% (completely confident). Finally, participants were debriefed.

#### 1.5. Data analysis

Eyewitness researchers usually refer to “choosers” and “non-choosers” when discussing the CA relation. This terminology suggests that there is a significant relation for people who choose but not for people who reject lineups. We hypothesize that the nature of the decisions, rather than something about the people making the decisions, accounts for the discrepancy in observed CA relations. Thus, we discuss the data in terms of the decisions: selections and rejections. The appropriateness of this approach is further evident when a within-participants design is used because the same participants are both choosers and nonchoosers. Rejections occurred whenever participants responded that 0 targets were present in the many-to-one presentation or said ‘No’ (indicating that they did not see the person during exposure) in response to a photo in the one-to-one presentation. A selection occurred whenever participants indicated that 1, 2, 3, or 4 targets were present during the many-to-one presentation or indicated ‘Yes’ (they did see the person during exposure) to a face during the one-to-one presentation. The confidence ratings for correct and incorrect selections and correct and incorrect rejections were averaged such that each participant (potentially) generated four data points.

## 2. Experiment 1

Participants were randomly assigned to either the many-to-one or one-to-one condition. In the many-to-one condition participants ( $N=42$ ) were required to state how many of the four targets were present in a simultaneous array of 24 faces and state their confidence in these decisions. We expected that confidence in rejections would not be related to accuracy and that the CA relation for selections would be weaker than in the one-to-one condition. In the one-to-one condition, participants ( $N=43$ ) saw and responded to the lineups one face at a time and provided confidence after each decision. We predicted a significant difference in mean confidence

between accurate and inaccurate decisions, for both selections and rejections.

## 3. Results

### 3.1. Accuracy

#### 3.1.1. Many-to-one decisions

For many-to-one decisions, where at least one target was present, correct selection decisions were made on .34 ( $SD=.19$ ) of the trials, incorrect selections (indicating an incorrect number of targets) on .57 ( $SD=.18$ ) of the trials, and incorrect rejections on .09 ( $SD=.12$ ) of the trials. When asked to identify specific targets, the average proportion of correct identifications was .60 ( $SD=.11$ ). When participants correctly stated the number of targets that were present in the many-to-one condition, half (.50) of these decisions were only serendipitously correct – the number of targets present was correct but one or more of the subsequent individual identifications were of lineup fillers. Of the two many-to-one trials where all targets were absent, participants correctly rejected the array .37 of the time ( $SD=.38$ ).

#### 3.1.2. One-to-one decisions

Participants correctly identified targets .60 of the time ( $SD=.16$ ) and correctly rejected fillers .91 of the time ( $SD=.05$ ).

### 3.2. Confidence-accuracy

#### 3.2.1. Many-to-one decisions

When participants made a single decision to reject multiple faces from a many-to-one presentation, confidence of correct decisions ( $M=48.18\%$ ,  $SD=25.03$ ) was not significantly different from the confidence of incorrect decisions ( $M=46.36\%$ ,  $SD=24.60$ ),  $t(10)=0.21$ ,  $p=.84$ ,  $d=.06$ . This finding is consistent with the frequently reported lack of CA relation for nonchoosers. The number of observations in this analysis is small because few participants made exactly one correct and one incorrect rejection from the two TA trials. When participants selected multiple faces from a many-to-one presentation by indicating 1–4 targets were present, confidence was significantly higher when the decisions were correct ( $M=65.59\%$ ,  $SD=19.04$ ) rather than incorrect ( $M=54.96\%$ ,  $SD=16.17$ ),  $t(40)=4.88$ ,  $p<.001$ ,  $d=0.76$ . Even under this more complex situation, choosers were more confident when correct than incorrect.

#### 3.2.2. One-to-one decisions

As predicted, when participants rejected single faces from a one-to-one presentation, confidence was greater when the decisions were correct ( $M=84.11\%$ ,  $SD=11.71$ ) than when the decisions were incorrect ( $M=71.24\%$ ,  $SD=20.21$ ),  $t(42)=5.91$ ,  $p<.001$ ,  $d=0.90$ . It is possible that this result is an artifact. Many single decisions to reject individual faces may have been obvious and associated with high levels of confidence thus biasing the mean. To address this issue, we designated an “innocent suspect” based on the “worst case scenario” as the individual most often identified from each of the six-person TA lineups (Pryke, Lindsay, Dysart, & Dupuis, 2004). Still, confidence was significantly higher when rejecting innocent suspects ( $M=79.37\%$ ,  $SD=12.87$ ) than when rejecting targets ( $M=71.24\%$ ,  $SD=20.21$ ),  $t(42)=-4.19$ ,  $p<.001$ ,  $d=0.64$ .

Participants who selected a single face from a one-to-one presentation overall displayed significantly higher confidence when the decision was correct ( $M=81.81\%$ ,  $SD=11.49$ ) than when the decision was incorrect ( $M=59.52\%$ ,  $SD=12.13$ ),  $t(42)=16.15$ ,  $p<.001$ ,  $d=2.46$ . Confidence was significantly higher for target selections ( $M=81.81\%$ ,  $SD=11.49$ ) than for innocent suspect selections ( $M=62.05\%$ ,  $SD=16.43$ ),  $t(42)=9.61$ ,  $p<.001$ ,  $d=1.46$ .

This finding is consistent with the often reported positive CA correlation for choosers.

### 3.2.3. Interaction

The interaction data may be of interest to the reader. However they are not critical to the claims made by the authors. Our main intent was to test whether a positive CA relationship between correct and incorrect rejections could be facilitated by tying confidence ratings to single rejection decisions and whether the positive CA relation between correct and incorrect selections could be minimized by tying single confidence ratings to multiple selection decisions. Thus, our central concern was to investigate whether the CA relation is due to the nature of the task (e.g., one-to-one decision) rather than the type of decision (e.g., selection, rejection). This is demonstrated above. Note that interpretation of the following results must be made with caution due to confounds (e.g., the number of confidence decisions was 10 in the many-to-one condition and 240 in the one-to-one condition).

One might ask if there is a significant interaction between one-to-one versus many-to-one decisions and correct versus incorrect decisions in regard to mean confidence. To address this question, we combined the data in a 2 (presentation: many-to-one versus one-to-one)  $\times$  2 (accuracy: correct versus incorrect) ANOVA. The results indicated that correct decisions were made with greater confidence ( $M = 72.72\%$ ) than incorrect decisions ( $M = 58.40\%$ ),  $F(1, 83) = 123.26, p < .001, \eta_p^2 = .60$ . One-to-one presentation of photos was associated with greater confidence ( $M = 74.17\%$ ) than many-to-one presentation ( $M = 56.95\%$ ),  $F(1, 83) = 32.63, p < .001, \eta_p^2 = .28$ . The interaction between presentation and accuracy was significant,  $F(1, 83) = 6.38, p = .01, \eta_p^2 = .07$ . The interaction reflects the fact that the difference between mean confidence for correct and incorrect decisions was greater in the one-to-one condition ( $M_D = 17.58\%$ ),  $t(42) = 12.88, p < .001, d = 1.96$ , than in the many-to-one condition ( $M_D = 11.06\%$ ),  $t(41) = 5.02, p < .001, d = 0.77$ . Confidence in the one-to-one decision situation was significantly higher than confidence in the many-to-one decision situation for both correct,  $t(64.39) = 6.04, p < .001, d = 1.32$  (equal variances not assumed), and incorrect decisions,  $t(83) = 4.38, p < .001, d = 0.95$ .

Additionally, while it would be interesting to see whether and how decision type (selection versus rejection) interacts with the factors examined above (presentation and accuracy), this omnibus analysis must be interpreted with caution due to the data loss in many-to-one decisions noted earlier, an increased likelihood of Type I error, as well as violations of the assumptions of heterogeneity of variance and covariation. With this in mind, a 2 (presentation)  $\times$  2 (decision type)  $\times$  2 (accuracy) ANOVA produced no main effect of decision type ( $F = 2.26, p = .14$ ) and no three-way interaction between presentation, decision type, and accuracy ( $F = .29, p = .59$ ). Decision type did interact with presentation,  $F(1, 51) = 20.41, p < .001, \eta_p^2 = .29$  wherein the difference between selections (60.50%) and rejections (46.50%) in the many-to-one presentation  $t(9) = 2.69, p = .02, d = 0.86$ , was greater than – and in the opposite direction to – the difference between selections (70.67%) and rejections (77.68%) in the one-to-one presentation,  $t(42) = 3.69, p = .001, d = 0.59$ .

### 3.3. Calibration

To investigate the extent to which confidence could be used to assess decision accuracy we calculated within-subjects calibration (C), over/underconfidence (OU), and Adjusted Normalized Resolution Index (ANRI; Yaniv, Yates, & Smith, 1991) statistics for each participant. The C statistic indexes the degree of correspondence between the subjective assessment (confidence) and the objective probability of accuracy, and varies from 0 (perfect calibration) to 1. The OU statistic indicates the extent to which participants are

more or less confident than they are accurate. The ANRI assesses the extent to which confidence discriminates correct from incorrect decisions from 0 (no discrimination) to 1 (perfect discrimination).

We argued that confidence would be a better predictor of accuracy when decisions were based on one-to-one comparisons, rather than many-to-one comparisons. For selections, mean within-subjects ANRI statistics were significantly greater than zero in both the many-to-one ( $M = .14, SD = .27$ ) and one-to-one ( $M = .30, SD = .19$ ) conditions,  $t(40) = 3.42, p = .001, d = 0.54$  and  $t(42) = 10.45, p < .001, d = 1.65$ , respectively. Thus, in both conditions, confidence discriminated correct from incorrect decisions. Consistent with our predictions, discrimination was greater in the one-to-one, than the many-to-one condition,  $t(82) = 3.11, p = .003, d = 0.69$ . No differences in the C and OU statistics were found between conditions,  $ts(82) \leq 1.56, ps \geq .123, ds \leq 0.34$ .

ANRI are meaningful only in conditions where accuracy varies. Ten participants made no rejections and 21 showed no variation in rejection accuracy. Thus, within-subjects ANRI statistics could be calculated for only 11 participants providing too few data points in the many-to-one condition to support a between-conditions comparison of discrimination. The reported analyses for the one-to-one condition used only 'suspect' rejections (reducing chance accuracy). The mean within-subjects ANRI ( $M = .15, SD = .21$ ) for rejections<sup>2</sup> in the one-to-one condition was significantly greater than zero,  $t(42) = 4.46, p < .001, d = 0.71$  (i.e., confidence discriminated between correct and incorrect rejections). Further, consistent with predictions, a comparison of mean C statistics demonstrated superior calibration in the one-to-one condition ( $M = .06, SD = .05$ ), than in the many-to-one condition ( $M = .24, SD = .20$ ),  $t(73) = 5.58, p < .001, d = 1.31$ . Over/underconfidence did not differ significantly between conditions ( $M = -.09, SD = .40$  and  $M = -.13, SD = .11$ , for the many-to-one and one-to-one conditions, respectively),  $t(73) = 0.55, p = .58, d = 0.14$ .

To summarize, our analyses provide three lines of support for our hypothesis that many-to-one comparisons undermine the CA relation. First, for selections, discrimination was reduced in the many-to-one compared to the one-to-one condition. Second, contrary to previous research, confidence discriminated between correct and incorrect rejections when these decisions were based on one-to-one, rather than many-to-one, comparisons. Finally, CA calibration for rejections was better when based on one-to-one, compared to many-to-one, comparisons.

## 4. Experiment 2: one-to-one decisions with simultaneous presentation

Sequential presentation is not the only way that people could make one-to-one decisions; they could examine a simultaneous lineup and make a separate decision for each face. In Experiment 2 we presented the same sets of targets and simultaneous 24-person arrays used in Experiment 1, but for each array participants ( $N = 16$ ) were asked to make separate decisions for each face (e.g., Is number 1 a person you saw?) and express confidence in each decision. Examining confidence for selections and rejections made from this format attenuates the potential effects of sequential versus simultaneous presentation that may have contributed to the results observed in Experiment 1.

<sup>2</sup> Initial analyses included all rejection decisions. This approach inflated rejection accuracy (i.e., rejecting every face in a 6-member TP lineup would achieve a rejection accuracy rate of 83%. If half the viewed lineups are TA, the same approach yields an accuracy rate of 92%), constraining co-variation between confidence and accuracy and undermining discrimination. The reported analyses used only 'suspect' rejections (reducing chance accuracy).

## 5. Results

### 5.1. Confidence-accuracy

As predicted, making one-to-one decisions resulted in a significant difference in the mean confidence of accurate ( $M = 79.45\%$ ,  $SD = 12.33$ ) and inaccurate ( $M = 67.55\%$ ,  $SD = 14.50$ ) rejections,  $t(15) = 5.68$ ,  $p < .001$ ,  $d = 1.42$ . The difference in the mean confidence between accurate ( $M = 82.51\%$ ,  $SD = 11.63$ ) and inaccurate ( $M = 63.37\%$ ,  $SD = 13.08$ ) selections also was significant,  $t(15) = 5.65$ ,  $p < .001$ ,  $d = 1.47$ . The difference between accurate and inaccurate rejections ( $M = 11.90\%$ ,  $SD = 8.37$ ) was slightly smaller than that of accurate and inaccurate selections ( $M = 19.14\%$ ,  $SD = 13.56$ ) and approached significance,  $t(15) = 2.02$ ,  $p = .06$ ,  $d = 0.53$ .

### 5.2. Calibration

Mean within-subjects ANRI statistics were significantly greater than zero for both suspect selections ( $M = .09$ ,  $SD = .12$ ) and suspect rejections<sup>3</sup> ( $M = .62$ ,  $SD = .46$ ),  $t(15) = 2.98$ ,  $p = .009$ ,  $d = 0.75$  and  $t(8) = 4.04$ ,  $p < .005$ ,  $d = 1.43$ , respectively. A paired samples  $t$ -test found a significant difference between ANRI statistics for selections ( $M = .06$ ,  $SD = .06$ )<sup>4</sup> and suspect rejections ( $M = .62$ ,  $SD = .46$ ),  $t(15) = 3.57$ ,  $p = .007$ ,  $d = 1.82$ . However, confidence discriminated accuracy for both selections and rejections. There were no significant differences between decision types for C ( $M = .09$ ,  $SD = .05$  and  $M = .09$ ,  $SD = .06$ , for selections and rejections respectively) or OU statistics ( $M = .02$ ,  $SD = .14$  and  $M = -.04$ ,  $SD = .15$ ),  $ts(15) < 1.17$ ,  $ps > .262$ ,  $ds < 0.34$ .

## 6. Discussion

We set out to determine why prior studies consistently report positive CA relations for selections but little or no CA relation for rejections from simultaneous lineups. We hypothesized that witnesses who select a lineup member base their confidence in that decision on the degree of match between the appearance of the person selected from the lineup and their memory for the criminal: the closer the match, the more confident the witness is in his or her decision (e.g., Stretch & Wixted, 1998). This reasoning explains the positive CA relations typically found for “choosers”. But why should “nonchoosers” fail to show a positive CA relation? We hypothesized that people who reject (particularly simultaneous) lineups (by choosing no one) must somehow combine their confidence in multiple decisions to reject each of the individual lineup members into a single confidence judgment. Such a “many-to-one” comparison may weaken the CA relation because the link between the bases for the identification decision and the confidence rating is less direct (Merkle & Van Zandt, 2006).

Consistent with our hypothesis, confidence was significantly related to decision accuracy when the confidence rating referred to a decision about a single face. This pattern held regardless of whether that decision was to select or reject a lineup member. Indeed, decisions to reject the most similar member of the TA lineups were made with significantly greater confidence than the rejection of targets. Prior results indicating that eyewitness confidence is weakly related or unrelated to the accuracy of lineup

rejections are based, at least in part, on the built-in artifact of the many-to-one nature of the single confidence rating for the multiple decisions required to reject a lineup. This finding was further supported by the fact that the difference in confidence between correct and incorrect selections was greater in the one-to-one than the many-to-one selection situation.

Legitimate arguments can be made regarding confounds in the research described above. In particular, the nature of the task between the many-to-one and the one-to-one conditions differs. In the many-to-one task, participants are asked to make a judgment about how many target faces are present. In the one-to-one task, participants make a judgment about a single face. This type of confound is indeed problematic. However, we argue that this confound already exists in the literature. In particular, rejections of simultaneous lineups are – by their nature – single decisions about multiple faces. Although the decision here is not directly about how many faces are not the target's face, that is what a rejection of a simultaneous lineup is telling us in essence (e.g., “zero of these 6 faces are the target face”). So, while a confidence rating about a selection reflects confidence about a particular face, a confidence rating about a rejection reflects confidence about multiple faces and could be based on a number of factors (e.g., average confidence across all faces, maximum confidence based on a single face, minimum confidence based on a single face, etc.). Although the specific cognitive (e.g., Takahashi, 1998) or social psychological (Chernev, 2009) bases for confidence ratings in a many-to-one rejection situation are unclear, we argue that the difference in the nature of selections and rejections (for simultaneous lineups) is the primary reason why the CA relation is generally seen in one situation and not the other.

We find the presence of a CA relation for rejections – when based on one-to-one decisions – promising, particularly when taken with other research that has investigated one-to-one decisions for lineup members. For example, Sauer, Brewer, and Weber (2008) found that having participants provide confidence ratings for each lineup member was useful in determining whether or not the suspect in the lineup was guilty. While the data were originally analyzed in a way that optimized classification accuracy, subsequent analyses showed that idiosyncratic classification criteria are not required for this procedure to work (Sauer, Brewer, & Weber, 2012). In all, we are excited by the diagnostic potential of methods where confidence ratings are collected based on one-to-one decisions.

## 7. Practical applications

Sequential lineups already are used in many places because they lower the rate of false positive selections (Bertrand, Lindsay, Beaudry, Mansour, & Whaley, 2009). To the extent that the one-to-one presentation permits the collection of information related to the accuracy of lineup rejections, there is an additional reason to use sequential lineups. Alternatively, police using simultaneous lineups could request selection decisions and confidence ratings for each lineup member.

## Acknowledgements

This research was supported by grants from the Social Sciences and Humanities Research Council of Canada to R.C.L. Lindsay and Natalie Kalmet.

## References

- Behrman, B. W., & Richards, R. E. (2005). Suspect/foil identification in actual crimes and in the laboratory: A reality monitoring analysis. *Law and Human Behavior*, 29(3), 279–301. <http://dx.doi.org/10.1007/s10979-005-3617-y>
- Bertrand, M. I., Lindsay, R. C. L., Beaudry, J. L., Mansour, J. K., & Whaley, E. (2009, August). *A survey of American officers' lineup procedures*. Toronto, ON, Canada: American Psychological Association.

<sup>3</sup> This calibration analysis includes only data from suspect selections and rejections to avoid inflating accuracy. This approach also reduces both ANRI and variation in accuracy.

<sup>4</sup> The variation in ANRI for selections reflects the reduced number of participants entered in the comparison. Some participants showed no variation in accuracy for either suspect selections or rejections. Thus, we could not calculate ANRI statistics for both decision types for all participants, and could not enter all participants in the analysis.

- Brewer, N. (2006). *Uses and abuses of eyewitness identification confidence*. *Legal and Criminological Psychology*, 11, 3–23.
- Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied*, 8(1), 44–56. <http://dx.doi.org/10.1037/1076-898X.8.1.44>
- Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, 12(1), 11–30. <http://dx.doi.org/10.1037/1076-898X.12.1.11>
- Boyce, M., Beaudry, J., & Lindsay, R. C. L. (2007). Belief of eyewitness identification evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology: Vol. 2. Memory for people* (pp. 501–525). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Chernev, A. (2009). Choosing versus rejecting: The impact of goal-task compatibility on decision confidence. *Social Cognition*, 27, 249–260. <http://dx.doi.org/10.1521/soco.2009.27.2.249>
- Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior*, 32, 406–422. <http://dx.doi.org/10.1007/s10979-007-9115-7>
- Leippe, M. R., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology: Vol. 2. Memory for people* (pp. 377–426). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, 70, 556–564.
- McAllister, H. A., & Bregman, N. J. (1986). Juror underutilization of eyewitness nonidentifications: Theoretical and practical implications. *Journal of Applied Psychology*, 71(1), 168–170. <http://dx.doi.org/10.1037/0021-9010.71.1.168>
- McAllister, H. A., & Bregman, N. J. (1989). Juror underutilization of eyewitness nonidentifications: A test of the disconfirmed expectancy explanation. *Journal of Applied Social Psychology*, 19(1), 20–29. <http://dx.doi.org/10.1111/j.1559-1816.1989.tb01218.x>
- Merkle, E. C., & Van Zandt, T. (2006). An application of the Poisson race model to confidence calibration. *Journal of Experimental Psychology: General*, 135, 391–408. <http://dx.doi.org/10.1037/0096-3445.135.3.391>
- Mickes, L., Hwe, V., Wais, P. E., & Wixted, J. T. (2011). Strong memories are hard to scale. *Journal of Experimental Psychology: General*, 140, 239–257. <http://dx.doi.org/10.1037/a0023007>
- Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: A method of calibrating eyewitness identifications. *Journal of Applied Psychology*, 89(1), 73–84. <http://dx.doi.org/10.1037/0021-9010.89.1.73>
- Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, 137, 528–547. <http://dx.doi.org/10.1037/a0012712>
- Sauer, J. D., Brewer, N., & Weber, N. (2012). Using confidence ratings to identify a target among foils. *Journal of Applied Research in Memory and Cognition*, 1, 80–88. <http://dx.doi.org/10.1016/j.jarmac.2012.03.003>
- Sauer, J. D., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior*, 34, 337–347. <http://dx.doi.org/10.1007/s10979-009-9192-x>
- Sauerland, M., Sagana, A., & Sporer, S. L. (2012). Postdicting nonchoosers' eyewitness identification accuracy from photographic showups by using confidence and response times. *Law and Human Behavior*, 36, 394–403. <http://dx.doi.org/10.1037/h0093926>
- Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime & Law*, 13(6), 611–625. <http://dx.doi.org/10.1080/10683160701264561>
- Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied*, 15(1), 46–62. <http://dx.doi.org/10.1037/a0014560>
- Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology*, 85(4), 542–550. <http://dx.doi.org/10.1037/0021-9010.85.4.542>
- Sporer, S. L. (1992). Postdicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology*, 22, 157–180. <http://dx.doi.org/10.1002/ejsp.2420220205>
- Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, 78(1), 22–33. <http://dx.doi.org/10.1037/0021-9010.78.1.22>
- Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, 118(3), 315–327. <http://dx.doi.org/10.1037/0033-2909.118.3.315>
- Stretch, V., & Wixted, J. T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1397–1410. <http://dx.doi.org/10.1037/0278-7393.24.6.1397>
- Takahashi, A. (1998). The relationship of the proportion correct in recognition and the confidence rating. *Japanese Journal of Psychology*, 69, 9–14.
- The Innocence Project. (n.d.). The causes of wrongful conviction. *The Innocence Project*. Retrieved October 12, 2010, from <http://www.innocenceproject.org/understand/>
- Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. *Applied Cognitive Psychology*, 20, 17–31. <http://dx.doi.org/10.1002/acp.1166>
- Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, 88(3), 776–784. <http://dx.doi.org/10.1037/0033-2909.88.3.776>
- Wixted, J. T., & Gaitan, S. C. (2002). Cognitive theories as reinforcement history surrogates: The case of likelihood ratio models of human recognition memory. *Animal Learning & Behavior*, 30, 289–305. <http://dx.doi.org/10.3758/BF03195955>
- Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, 117, 1025–1054. <http://dx.doi.org/10.1037/a0020874>
- Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, 110, 611–617.