Target article

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

R.C.L. Lindsay [a,*], Natalie Kalmet [a], Jaime Leung [a], Michelle I. Bertrand [b], James D. Sauer [c], Melanie Sauerland [d]

[a] Queen's University, Canada
[b] University of Winnipeg, Canada
[c] University of Portsmouth, United Kingdom
[d] Maastricht University, Netherlands

ARTICLE INFO

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 (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[1]) of

---