+Model
ARTICLE IN PRESS

Journal of Applied Research in Memory and Cognition xxx (2020) xxx–xxx

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Empirical article

Live Presentation for Eyewitness Identification is Not Superior to Photo or Video Presentation☆

Eva Rubínová*
University of Portsmouth, UK

Ryan J. Fitzgerald
University of Portsmouth, UK

Stefana Juncu
University of Portsmouth, UK

Eva Ribbers
Institute of Criminal Justice Studies, UK

Lorraine Hope
University of Portsmouth, UK

James D. Sauer
University of Tasmania, Australia

Eyewitnesses are widely believed to have a better chance of identifying a perpetrator from a live identification procedure than from photo or video alternatives. To test this live superiority hypothesis, prospective students and their parents (N = 1048) became unsuspecting witnesses to staged events and were randomly assigned to live, photo, or video identification procedures. In Experiment 1, participants viewed a single person at the identification procedure. In Experiment 2, participants viewed a lineup of six people. Across experiments, live identification procedures did not improve eyewitness identification performance. The results show that even under experimental settings designed to eliminate the disadvantages of conducting live lineups in practice, live presentation confers no benefit to eyewitnesses.

Keywords: Eyewitness Identification, Face, Body, Lineup, Showup

Police lineups are common in criminal investigations and also in film. In *The Usual Suspects,* an eyewitness views criminals through a one-way mirror as they step forward and read an expletive-laden phrase. Number 1 reads his line in a cold, calm manner. Number 2 gestures his hand like a gun toward his observers and mockingly repeats the phrase with overbearing intensity and bravado. By Number 3's turn, the lineup members have burst into laughter. The *Usual Suspects* lineup was intentionally farcical, a caricature of a police lineup gone awry. Nevertheless, watching the chaotic lineup unfold, a kernel of truth is evident: live lineups are hard to control. Much like a viewer cannot unhear a swearword blurted out on live television, an eyewitness cannot unsee an irregularity at a live lineup. Live lineups have other drawbacks, too: they are expensive (BBC

☆ This research was supported by a Future Research Leaders grant from the Economic and Social Research Council to Ryan J. Fitzgerald (ES/N016602/1). The research materials, data, and list of acknowledgements are available in the Online Supplemental Materials (osf.io/6r5b2).

Ryan J. Fitzgerald is now at Department of Psychology, Simon Fraser University.

* Correspondence concerning this article should be addressed to: Eva Rubínová, Department of Psychology, University of Portsmouth, King Henry Building, King Henry I Street, PO1 2DY, Portsmouth, Hampshire, UK.
E-mail address: eva.rubinova@port.ac.uk (E. Rubínová).

Please cite this article in press as: Rubínová, E., et al. Live Presentation for Eyewitness Identification is Not Superior to Photo or Video Presentation. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2020.08.009

News, 2003), logistically difficult to organize (Pike, Brace, & Kynan, 2002), and stressful for witnesses (Brace, Pike, Kemp, & Turner, 2009). So why would law enforcement ever use live lineups?

The Live Superiority Hypothesis

Each year hundreds of thousands of eyewitnesses are asked to identify a stranger they observed committing a crime (Sauer, Palmer., & Brewer, 2019). Does it matter whether the identifications are conducted live or using photographs or videos?

People tend to think it does matter. When asked which lineup medium they would prefer, 82% of jury-eligible participants chose a live lineup over a photo or video lineup (Price, Harvey, Anderson, Chadwick, & Fitzgerald, 2019). The belief that live presentation yields the best identification outcome—the live superiority hypothesis—has also permeated identification policies (Brewer & Palmer, 2010; Fitzgerald, Price, & Valentine, 2018). A review of 54 countries showed that policies in 25 countries (46% of those sampled) implicitly or explicitly indicated a preference for live lineups, compared with two that preferred video lineups (4%) and none that preferred photo lineups (Fitzgerald, Rubínová, & Juncu, in press).

There are ample grounds to predict an advantage of live identification procedures. In line with the encoding specificity hypothesis, live identification could maximize available cues encoded from a live crime (Tulving & Thomson, 1973). The deficit in cues present in photos or videos may be especially pronounced if, as is common practice, only the head and shoulders are depicted. Although faces provide the best cues (Burton, Wilson, Cowan, & Bruce, 1999), bodies can also provide diagnostic identity cues (e.g., Hahn, O'Toole, & Phillips, 2016; O'Toole et al., 2011; Robbins & Coltheart, 2012) and the emerging literature on whole person identification indicates it is most advantageous to see the face and body of a moving person. Movement is theorized to bind the face and body into a coherent whole and distribute attention to all available cues (Yovel & O'Toole, 2016). For these reasons, it seems plausible that a live viewing should improve identification decisions.

However, the live superiority hypothesis has never been subjected to rigorous experimentation. Brewer and Palmer (2010) reviewed the literature and deemed the handful of previous experiments insufficient for policy direction due to small samples, methodological confounds, and neglecting to test performance in target-absent lineups. We similarly reviewed the live identification literature and found three more limitations (see Online Supplemental Materials, osf.io/6r5b2). First, researchers usually tested identification of only a single target person: in 86% of all previous studies, the target's identity did not vary. This neglect of stimulus sampling poses a risk to external validity and increases the likelihood of stimulus-specific effects (Wells & Windschitl, 1999). For example, imagine that live presentation is only advantageous for targets with distinctive bodies. The advantage would be exaggerated if the single target in a study has a distinctive body, and it would go undetected if the single target has a nondistinctive body. Second, only a minority of the authors (36%) explicitly mentioned randomly assigning participants to live and non-live conditions. Although reporting omissions could be to blame, random assignment should not be taken for granted in such experiments, given that it makes live testing substantially more difficult. Finally, none of the authors reported recording the stimuli for the photo or video conditions on the same day as the live identification procedure. Again, same-day recording makes the experiment more difficult, but recording on a different day could produce inconsistencies across conditions and compromise the manipulation (e.g., if a target's hairstyle at the live identification matched the hairstyle at the witnessed event more closely than the hairstyle at the photographic identification). We developed a new methodology to address these limitations.

Present Research

The live superiority hypothesis was tested at a series of mass data collection events. To recruit large samples, we staged interruptions at lectures attended by prospective students and their parents who visited a university to learn about its psychology program. This strategy yielded a total $N > 1000$. To increase generalizability, we recruited over 30 targets to stage the interruptions. This is more targets than the combined total of all previously published experiments with live identification procedures. After witnessing the interruption as a group, participants were randomly assigned to identify the interrupter from a live, video, or photo identification procedure. Videos and photos were recorded on the day of testing, and the procedures were administered individually to each participant by trained research assistants who were blind to the target's identity. This methodology provided the strongest test of the live superiority hypothesis to date.

Experiment 1

The choice of lineup medium affects the availability of fillers, which creates a tradeoff between internal and ecological validity in lineup medium experiments. Unlike live lineups, which involve finding people available to physically appear with the suspect, fillers in photo and video lineups are found by searching through massive databases of recorded images (Bergold & Heaton, 2018). The increased availability of suitable fillers in photo and video lineups causes the tradeoff: ecological validity is best achieved by constructing non-live lineups with better fillers than live lineups, but internal validity is best achieved by constructing live and non-live lineups with the same fillers. We sidestepped this tradeoff in Experiment 1 by presenting a suspect for identification with no fillers. Police use this procedure, conventionally known as a showup, if a suspect matching the perpetrator's description is located in the vicinity shortly after the crime.

Consistent with the live superiority hypothesis, we predicted that live identification would outperform photo and video identification. In Experiment 1a, to equate the conditions as much as possible, a full-body view of the suspect was presented in live, video, and photo showups. A live superiority effect under these conditions would indicate that live presentation increases the availability of identity cues for targets witnessed at live events.

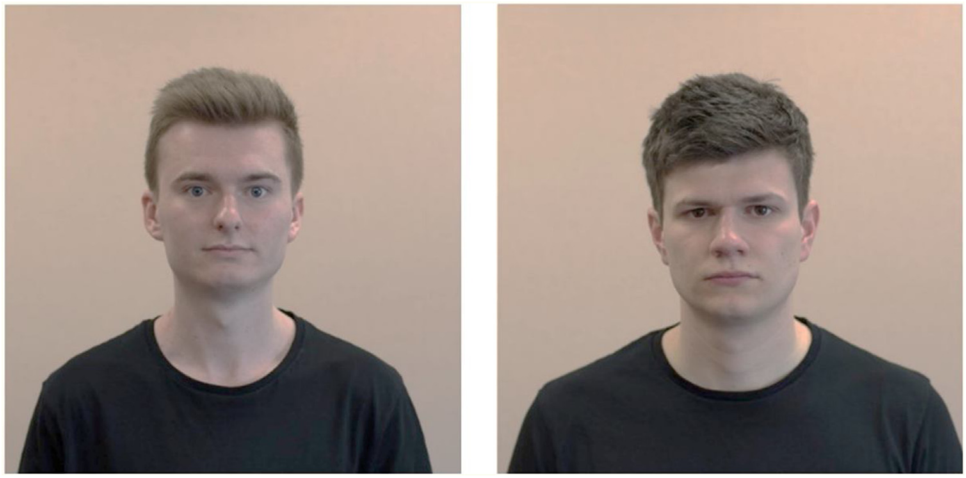

Figure 1. Example of a suspect pair.

In Experiment 1b, a full-body view of the suspect was only available at the live showups. At the video and photo showups, only the head and shoulders were visible (mugshot view). This comparison was hypothesized to result in an even larger advantage for live presentation and would have implications for the typical convention in applied settings of showing mugshot views at photo and video identification procedures.

Method

Participants. Most participants were prospective students and their parents, recruited during university visits to learn about a psychology program. The remaining participants (∼15%) were undergraduate psychology students recruited from lectures. We planned to terminate data collection when each condition reached 50 participants (a sample size with 70% power to detect a medium sized effect, Cohen's h = 0.50, and 90% power to detect a medium-to-large sized effect, Cohen's h = 0.65). In Experiment 1a the recruitment target was achieved at the 9th event, resulting in a final sample of 304 participants, and in Experiment 1b the recruitment target was achieved at the 9th event with a final sample of 306 participants. Exclusions and demographics are reported in Online Supplemental Materials.

Suspects. At the identification test for each event, participants viewed a guilty or innocent suspect (Figure 1). The guilty suspect was the target who previously interrupted a lecture. The innocent suspect was a different person who never appeared at the lecture. To be eligible for the role of innocent suspect, an actor was required (a) to be available for the testing session and (b) to match to a description of the target's age, race, and general appearance enough that the actor could be plausibly suspected of a crime committed by the target. Plausibility was judged by the first and second authors (for photos of all suspect pairs, see Experiment 1 log file in Online Supplemental Materials). During the interruption of the lecture, the target wore street clothes. At the identification test, all suspects wore matching black t-shirts, same color jeans or sweatpants, and black socks (they did not wear shoes, watches, or jewelry). Across events, Experiment 1a used 8 targets and Experiment 1b used 8 targets.

At test, participants saw the suspect live, in a video, or in a photo. For consistency of appearance across conditions, photos and videos were recorded on the day of data collection whenever possible. For two events, suspects were recorded one or two days earlier and were instructed not to change their appearance or hairstyle. To standardize lighting and background across conditions, photos and videos were recorded in the same room used for the live procedure.

The difference between experiments was whether the suspect's body was visible in the recorded images (Figure 2). In Experiment 1a, the suspect's body was visible regardless of the identification medium. The live suspect walked up to a one-way mirror, faced forward, and turned for left and right profile views. The video suspect did the same and was recorded from behind the one-way mirror. The photo suspect was depicted in distant frontal view, mugshot frontal view, and mugshot profile views. In Experiment 1b, the live condition was the same as in Experiment 1a; however, videos and photos depicted only mugshot frontal and profile views. All videos and photos are available in Online Supplemental Materials.

Procedure. Participants witnessed a target interrupt a lecture. At each event the target announced their arrival by claiming that they were teaching in the room earlier and forgot something. The lecturer permitted the target to collect a laser pointer from a desk at the front of the classroom. The target then thanked the lecturer, apologized for the interruption, and left. At the end of the lecture, an experimenter invited the audience to participate in a live, photo, or video identification procedure.

Approximately one hour after the interruption, participants were randomly allocated to a 2 (target: present, absent) × 3 (medium: live, video, photo) between-subjects design. Trained administrators, blind to the target's identity, escorted each participant separately to individual testing rooms. Utilizing multiple administrators and rooms enabled simultaneous testing of each condition. All rooms were equipped with a computer, which was programmed with standard instructions for the administrators to read. Administrators instructed all participants that the suspect may or may not be the person who interrupted the lecture and that they could take as long as needed to make their decision.

Please cite this article in press as: Rubínová, E., et al. Live Presentation for Eyewitness Identification is Not Superior to Photo or Video Presentation. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2020.08.009

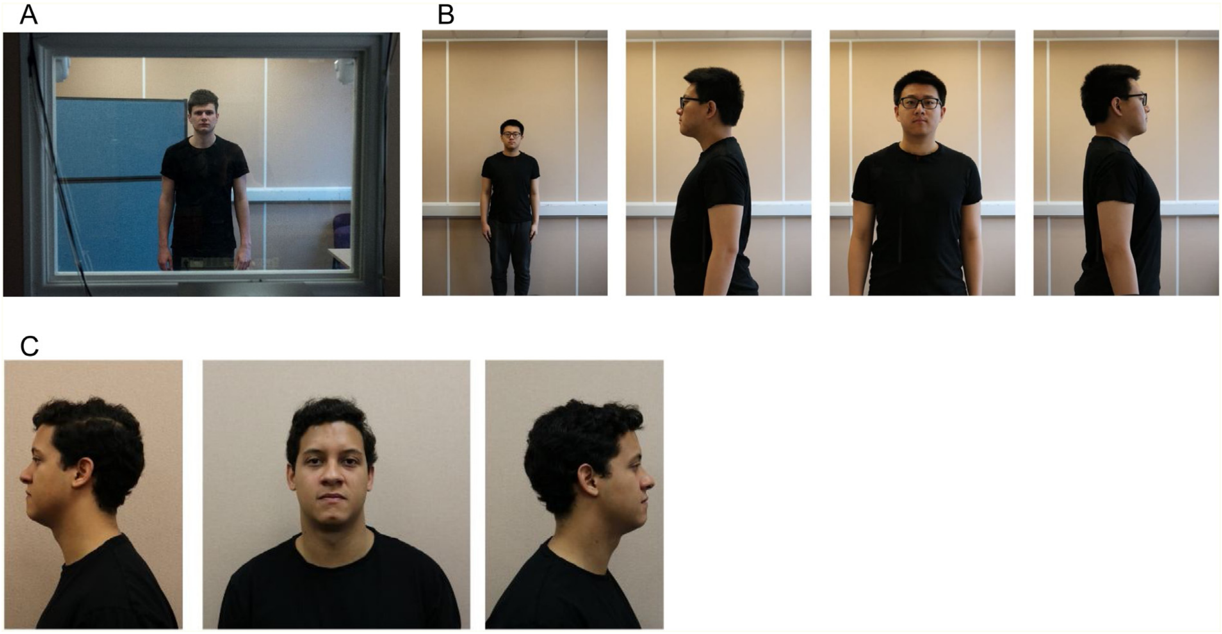

Figure 2. Example of (A) the participants' view of the suspect in the live condition, (B) the photo images in Experiment 1a, and (C) the photo images in Experiment 1b. Videos from Experiment 1a and 1b are available in Online Supplemental Materials.

At test, each participant viewed a single suspect (i.e., a showup). In the live condition, the suspect waited to be called from behind an opaque screen. The participant and administrator were in another room connected via one-way mirror and intercom. When the administrator called, the suspect walked toward the mirror, faced forward for ∼5 s, turned right for ∼5 s, and then left for ∼5 s. These motions were repeated until the participant was ready to decide. In the video condition, the suspect faced forward, then turned right and left for the same durations as in the live condition. Videos looped until the participant was ready to decide. In the photo condition, frontal and profile images were presented simultaneously.

Participants were instructed to press a key when they were ready to make a decision. Once pressed, the suspect exited from view and a new display appeared on the computer with the following options: (a) "YES, I believe this is the person" (b) "NO, I believe this is not the person" or (c) "I DON'T KNOW." After the decision, participants rated their confidence on a scale from 1 (*not at all confident*) to 7 (*very confident*). To conclude, administrators thanked participants and asked them not to discuss the experiment before they left the university.

Statistical analyses. Eyewitness identification data can be analyzed in multiple ways. A recent innovation has been to plot identification responses as Receiver Operating Characteristic curves (Mickes, Flowe, & Wixted, 2012; Wixted & Mickes, 2015), though this approach has not been universally endorsed (Lampinen, 2016; Wells, Smalarz, & Smith, 2015) and alternative measures have been proposed (Smith, Lampinen, Wells, Smalarz, & Mackovichova, 2019). We compared the various approaches to see whether different analytic decisions would affect our conclusions (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016). These additional tests were consistent with those reported here and can be found in Online Supplemental Materials. Confidence Accuracy Characteristic curves can also be found there (Mickes, 2015).

We conducted null-hypothesis significance tests (NHST) and computed Bayes factors (BF). Significance tests were run in *R*, with $\alpha = .05$. Bayesian analyses were run in *JASP*, with strength of BF evidence interpreted as follows: 1–3 = weak, 3–20 = positive, 20–150 = strong, >150 = very strong (Raftery, 1995). In the following analyses, "don't know" responses were treated as non-identification responses. Data and code are available as Online Supplemental Materials.

Results

Table 1 displays identification response rates and measures of discriminability and response bias. Discriminability (d') is the ability of an identification procedure to distinguish between guilty and innocent suspects. Positive values of d' signify that a procedure was more likely to result in an identification of a guilty suspect than an innocent suspect, negative values signify a greater likelihood of an innocent versus guilty suspect identification, and values of zero signify that a procedure was just as likely to result in identification of an innocent suspect as a guilty one. Response bias (c) is the tendency towards identification or nonidentification. Positive values of c signify that the group was biased toward nonidentification (conservative response bias), negative values signify that the group was biased toward identification (liberal response bias), and zero values indicate no response bias.

Experiment 1a. NHST for comparing group-level discriminability and response bias (Gourevitch & Galanter, 1967; Mickes, Moreland, Clark, & Wixted, 2014) revealed no significant differences in d' (live vs. photo, $G = 0.01$, $p = .996$; live vs. video, $G = 0.22$, $p = .823$; photo vs. video, $G = 0.24$, $p =$

Please cite this article in press as: Rubínová, E., et al. Live Presentation for Eyewitness Identification is Not Superior to Photo or Video Presentation. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2020.08.009

Table 1
Signal Detection Measures and Identification Response Rates in Experiments 1a and 1b

				Target Present				Target Absent			
Exp	Medium	*d'*	*c*	Hit	Miss	Don't Know	*n*	False Alarm	Correct Reject	Don't Know	*n*
1a	Live	1.11	0.74	.43	.51	.07	61	.10	.82	.08	51
	Photo – Full Body	1.11	0.21	.63	.31	.06	52	.22	.73	.04	45
	Video – Full Body	1.02	0.12	.65	.28	.07	46	.27	.57	.16	49
1b	Live	1.14	0.86	.38	.44	.17	52	.08	.89	.04	53
	Photo – Mugshot	0.86	0.56	.45	.43	.13	56	.16	.74	.10	50
	Video – Mugshot	1.34	0.57	.54	.38	.08	39	.11	.82	.07	56

.814), and significantly more conservative responding for live showups compared with photo showups $G = 2.65$, $p = .008$, and video showups, $G = 3.14$, $p = .002$. Photo and video response bias did not significantly differ, $G = 0.48$, $p = .631$. The medium was significantly associated with identification of guilty suspects, $\chi^2(2, N = 159) = 7.19$, $p = .027$, $BF_{10} = 1.67$. Live identification reduced correct identification rates compared with both photo identification, $\chi^2(1, N = 113) = 4.89$, $p = .027$, $BF_{10} = 2.60$, and video identification, $\chi^2(1, N = 107) = 5.37$, $p = .021$, $BF_{10} = 3.41$. Note that the Bayes Factors indicate that evidence for these associations was either weak or just over the threshold for positive evidence. Hits for photo and video showups did not significantly differ, $\chi^2(1, N = 98) = 0.03$, $p = .856$, with positive evidence for the null hypothesis, $BF_{01} = 4.14$. For innocent suspects, false alarms were not significantly associated with the identification medium, $\chi^2(2, N = 145) = 4.84$, $p = .089$, with weak evidence for the null hypothesis, $BF_{01} = 2.42$.

Experiment 1b. No significant differences were detected in *d'* (live vs. photo, $G = 0.69$, $p = .492$; live vs. video, $G = 0.45$, $p = .651$; photo vs. video, $G = 1.18$, $p = .238$). Similarly, no significant differences were detected in *c* (live vs. photo, $G = 1.46$, $p = .145$; live vs. video, $G = 1.35$, $p = .176$; photo vs .video, $G = 0.04$, $p = .968$). For guilty suspects, the medium was not significantly associated with hits, $\chi^2(2, N = 147) = 2.13$, $p = .344$, with positive evidence for the null hypothesis, $BF_{01} = 6.53$. Similarly, for innocent suspects, the medium was not significantly associated with false alarms, $\chi^2(2, N = 159) = 1.86$, $p = .394$, this time with strong evidence for the null hypothesis, $BF_{01} = 20.65$.

Discussion

Experiments 1a and 1b provided no support for the live superiority hypothesis. Participants were not significantly better at discriminating whether a live suspect was guilty or innocent than they were for a photo or video of the suspect. Even compared with videos and photos that omitted most of the suspect's body, live presentation of a suspect in full view did not significantly improve performance. These findings counter the idea that identification is improved by increasing the quantity and type of available cues (Clark, Moreland, & Rush, 2015; Cutler & Penrod, 1988; Cutler, Berman, Penrod, & Fisher, 1994). The only effect of live presentation found in Experiment 1a was more conservative responding, which reduced identification of guilty and innocent suspects. In Experiment 1b, a higher value of response bias in the live showup indicated a consistent although non-significant pattern.

The discrepant criteria model of identification decisions provides a potential explanation for the more liberal responding for the photo and video showups, compared with the live showup (Smith, Wells, Lindsay, & Myerson, 2018; Smith, Wilford, Quigley-McBride, & Wells, 2019). The model, which has roots in the basic recognition memory literature, states that witnesses respond to difficult testing conditions by adopting more lenient decision criteria. In support of the model, Smith et al. (2018) found that mistaken identification of an innocent suspect increased when the quality of the suspect's photo was degraded. Following this logic, if participants perceive video or photo showups as more difficult than live showups, the discrepant criteria model predicts that participants would lower their criterion at video and photo showups to reduce the chances of missing the target. An analogous situation was reported by Hockley, Hemsworth, and Consoli (1999), who found that after studying undisguised faces, participants were more inclined to identify disguised faces than undisguised faces. Similarly, in our experiments participants first encountered a live target in full view and then at test they saw either a live suspect in full view or a reproduced image of the suspect. Although we did not measure task difficulty and can only speculate that participants might have perceived greater difficulty in the non-live showups, this would correspond with anecdotal reports from eyewitnesses to real crimes who claim they could only make a correct identification if they saw the culprit in person (People v. Hoiland, 1971; State v. Buchanan, 2010; State v. Scott, 2019).

Experiment 2

In Experiment 2 we extended the test of the live superiority hypothesis to lineups. Live lineups were compared with two forms of video lineups: full-body and mugshot. If live lineups are superior to full-body and mugshot video lineups, it would suggest live presentation of the lineup members provides additional behavioral cues that are diagnostic for identification. If live lineups and full-body videos are superior to mugshot videos, it would suggest the body provides additional diagnostic cues.

Method

The methodology was generally consistent with Experiment 1, with a few noteworthy exceptions: (a) Following Cutler and Fisher (1990) participants witnessed two targets interrupt each lecture, rather than just one; (b) participants completed pre-lineup and post-lineup questionnaires; (c) identification was tested via 6-member, sequential lineups; (d) the live lineup was compared with full-body and mugshot video lineups (photo lineups were not tested); and (e) the mugshot videos were recorded using a booth on loan from the Video Identification Parade Electronic Recording (VIPER) Bureau. The VIPER booth is equipped with software that records 15 s clips of a person's head and shoulders under standardized lighting and background conditions. Approximately half of UK police forces use VIPER booths to record videos for eyewitness identification.[1]

Participants. Before data collection, we estimated that a sample of 129 per condition (N = 387 total) was required to achieve 80% power for Cohen's h = 0.35 (a small-to-medium sized effect). Accordingly, we planned to end data collection on the date in which we reached 129 per group. Termination criteria were met after 22 data collection events. The final sample consisted of 856 lineup decisions from 438 participants. Exclusions and demographics are reported in Online Supplemental Materials.

Suspects and fillers. Across events, we used eight male suspect pairs and eight female suspect pairs. To recruit lineup members, we advertised around campus and on the university website for actors who matched a description (e.g., male, 20–35 years, average/normal build, around 6″, short dark hair, and no facial hair). We also contacted actors from our previous experiments and asked recruits if they knew anyone who fit the description.

Procedure. The interruption began with a man and a woman entering the lecture theatre mid-way through a presentation. The man immediately announced that he had left books near the front desk, and the two walked to the front of the room to collect the books. At the desk the woman tried to hand the man two books, but the man said he had left a different one, so the woman returned a book and handed him another. The man then put the books into his backpack and the woman faced the audience and apologized for the interruption. The mean length of the interruptions was 43 s. The invitation to participate was the same as in Experiment 1, with mention that participants would be asked to identify both targets from lineups.

After consenting to take part approximately one hour after the interruption, participants completed a pre-lineup questionnaire (see Online Supplemental Materials). The pre-lineup questionnaire specified the lineup condition at the top of the page (live, full-body video, or head-and-shoulder video). Participants used the questionnaire to report demographic information, their location in the lecture theatre (front, middle, or back), and prospective confidence ratings for the upcoming identifications (separately for the man and the woman). The questionnaire also contained written instructions for the identification task, which were read to the participant by the lineup administrator.

Administrators instructed all participants on how to complete the sequential lineup and emphasized they were not aware of the identity of the targets (see Online Supplemental Materials). The lineups were computer-administered, except that in the live condition the administrator called each lineup member by number. Following the procedures used in Experiment 1, lineup members remained in view until participants indicated that they were ready to make a decision. Participants were asked "Is this the person you saw?" and could respond "Yes", "No", or "Not Sure." Affirmative responses were followed by a confidence rating from 0% (low confidence) to 100% (high confidence).

A response was required before proceeding to the next lineup member. After responding for all lineup members, a display appeared asking if an identification was made. For participants who identified one lineup member or zero lineup members, the computer directed participants to the administrator for further instruction. At this point, the administrator advised that the lineup was complete and their responses were recorded. The administrator then asked the participant if they had any questions before they proceeded. This question was included to give participants who did not identify anyone an opportunity to request another viewing of the lineup, without explicitly offering this option. If participants requested an additional viewing, the administrator arranged for the lineup to repeat. Participants who reported that they identified more than one lineup member were presented with the lineup again to clarify their decision.

After completing separate lineups for each target, participants completed a post-lineup questionnaire (see Online Supplemental Materials) that assessed post-identification confidence, comfort during the procedure, endorsement of cues used to make the identification decision (face, body, posture, height, movements, behaviors/mannerisms), and endorsement of confidence-related questions. The post-lineup queries were completed separately for the male and female lineups.

Results

Table 2 displays identification decisions and compound decision signal detection theory measures (using the independent observations model, see Bruer, Fitzgerald, Price, & Sauer, 2017), which take into account the detection and identification components of lineup decisions (Duncan, 2006; Palmer & Brewer, 2012; Palmer, Brewer, & Weber, 2010). No significant differences in d' were detected (as indicated by the overlapping 95% CIs in square brackets; Tryon, 2001): live = 1.28 [1.09, 1.47], full-body video = 1.41 [1.23, 1.59], mugshot video = 1.17 [1.00, 1.34]. Similarly, no significant differences in c were detected: live = 0.89 [0.80, 0.98], full-body video = 0.80 [0.72, 0.89], mugshot video = 0.90 [0.82, 0.98]. For target-present lineups, the medium was not significantly associated with identification responses, $\chi^2(4, N = 435) = 2.37$, $p = .669$, with very strong evidence for the null hypothesis, $BF_{01} = 424.61$. For target-absent lineups, again the medium was not significantly associated with identification responses, $\chi^2(2, N = 421) = 0.37$, $p = .833$, and evidence for the null hypothesis was strong, $BF_{01} = 46.91$.

We used responses from the target-absent lineups to compute what Quigley-McBride and Wells (in press) refer to as resultant lineup fairness, which indicates the spread of mistaken identi-

[1] These videos were recorded locally and have not been quality assured by the VIPER Bureau. We accept full responsibility for the quality of the videos.

Table 2
Signal Detection Measures and Identification Response Rates in Experiment 2

			Target Present				Target Absent		
Medium	*d'*	*c*	Hit	Filler	Miss	*n*	Filler	Correct Reject	*n*
Live	1.28	0.89	.31	.25	.44	136	.34	.66	119
Video – Full-Body	1.41	0.80	.37	.21	.42	136	.37	.63	139
Video – Mugshot	1.17	0.90	.29	.25	.46	163	.37	.63	163

fications across the lineup members. We used the distribution of choices in the target-absent lineups to compute Tredoux's E' (Tredoux, 1998). In a 6-member lineup, E' can range from 1 (*biased*) to 6 (*fair*). In our lineups, $E' = 3.52$, 95% CI [2.79, 4.24]. The distribution of identifications across fillers for each lineup is reported in Table SM3 in Online Supplemental Materials.

Discussion

Again, no support was found for the live superiority hypothesis. There were no significant differences in discriminability,[2] bias, or identification decisions, and Bayesian analyses suggested strong or very strong evidence for the null hypothesis. Despite having 80% power to detect an effect of a small-to-medium size, we do not exclude the possibility that with an even larger sample the advantage for live lineup over mugshot video lineup could have crossed the significance threshold. However, the fact that discriminability was numerically highest in the full-body video lineup suggests that it would not have been because of the live presentation per se.

General Discussion

None of our experiments supported the live superiority hypothesis. The choice of identification medium had no significant effects on discrimination between guilty and innocent suspects. These findings suggest that, all else being equal, photo and video identification procedures are just as effective as live procedures.

We will highlight a couple of comparisons across the experiments, although caution is necessary because of methodological differences (i.e., the use of showup and lineup procedures, and the different calculations of signal detection theory measures). Discriminability scores associated with a given medium were relatively similar across experiments, with lower variability in the live procedures (1.11–1.28) and higher variability in the photo (0.86–1.11) and video procedures (especially when the full-body was in the view; mugshot: 1.12–1.17, full-body: 1.02–1.41). Response bias scores showed a higher variability across media in Experiment 1 (0.12–0.86) but more consistent conservative responding in Experiment 2 (0.80–0.90). We will discuss these differences and other related effects in turn.

[2] The only significant difference was indicated by ROC analyses (see Online Supplemental Materials, osf.io/6r5b2), which showed higher discriminability in the full-body video lineup than in the live lineup. This was, however, the only case where this comparison crossed the significance threshold, so we do not discuss it further.

Response Bias

With the showup procedure in Experiment 1, the medium seemed to influence the decision to make an identification. In Experiment 1a, live presentation of showups led to more conservative responding; specifically, live showups led to a reduction of hits and false alarms when compared to photo and video showups. Experiment 1b revealed a similar pattern of data, but the effect was not significant. Taking a social perspective of the identification task may help us understand why live procedures may lead to more conservative responding. Unlike examining photos or videos, live presentation puts the eyewitness into a social encounter with the suspect, which may lead to stress and anxiety (Brace et al., 2009). Dent and Stephenson's (Experiment 2, 1979) behavioral observations suggest that a face-to-face live procedure may increase participants' discomfort (see also Peters, 1991) and lead to fewer identifications when compared to a photo procedure (there was no video procedure in their study). This type of stress-induced aversion to risky decisions is in line with findings from behavioral economics (Cahlíková & Cingl, 2017; FeldmanHall, Raio, Kubota, Seiler, & Phelps, 2015).

The differences in response bias, however, did not replicate in Experiment 2. The live lineup in Experiment 2 revealed a response bias score (0.89) similar to the live showups in Experiment 1 (0.74–0.86), but the video lineups showed consistently high response bias scores (0.80–0.90). An important difference between our experiment and live lineup procedures happening in the real world, though, is the nature of the event: for victims of violent crimes, participating in a live lineup would likely be more stressful than for participants in our experiments, who witnessed a staged event. Such increased stress might lead to more conservative responding in live lineups than we observed in our experimental setting.

Why Live Procedures May Not Be Superior

It is easy to generate theories that predict live superiority, but our results demand explanation for why seeing a suspect live does not improve identification performance. Live identification should maximize the availability of identity cues encoded from a live perpetrator (Tulving & Thomson, 1973). Furthermore, given the positive contributions of motion and body cues to identification, this effect should be especially pronounced in comparison with images that depict only a portion of a static person (Hahn et al., 2016; O'Toole et al., 2011; Robbins & Coltheart, 2012). Yet in our experiments, showing a whole person live was no better than showing a static mugshot.

Please cite this article in press as: Rubínová, E., et al. Live Presentation for Eyewitness Identification is Not Superior to Photo or Video Presentation. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2020.08.009

We offer three possible reasons why live identification was not superior. First, prolonged exposure and elaborative encoding may be required to notice the types of cues afforded by live presentation (Brewer, 2011). People can readily recognize the shape and movements of familiar others, but it may require some time to pick up on diagnostic body features and learn the idiosyncratic movements of strangers (Butcher & Lander, 2017; O'Toole, Roark, & Abdi, 2002). Second, the additional cues at a live identification procedure may not all be diagnostic. In debriefing, some participants in the live conditions referred to cues of questionable reliability. For example, "I knew it wasn't him. This guy was too confident. The one who interrupted the lecture was shy." Behavioral cues such as these, which may be unique to live lineups, vary by context. For example, the same person might be shy when interrupting a crowded lecture hall but comfortable when standing behind a one-way mirror for the 20th time. Thus, it is possible that nondiagnostic behavioral cues are canceling out the benefits of any diagnostic cues that become available at live identification. Finally, relative to photo or video identification, live identification may lead to greater expectations of a strong recognition experience, that is, high ecphoric similarity (Tulving, 1981). If an eyewitness endorses the live superiority hypothesis, as many likely do (Price et al., 2019), they may have unrealistic expectations of their ability to identify a previously-seen person if said person is standing in front of them. When eyewitnesses are then confronted with this person and the match is lower than expected, it could lead them to withhold identification erroneously.

Limitations and Practical Application

In many respects, our experimental conditions were more favorable toward live lineups than would occur in practice. Compared with photo or video lineups, live lineups in criminal cases are associated with increased witness anxiety, longer delays between the witnessed event and identification procedure, and a smaller pool of fillers available to construct a fair lineup (Brace et al., 2009; Pike et al., 2002). On the other hand, the quality of the photos and videos in our experiments may have attenuated the benefits that a live procedure might confer over an identification procedure with lower quality images. In addition to their high resolution, our photos and videos were recorded on the same day as the witnessed event. Therefore, we cannot exclude the possibility of a live superiority effect if compared with more dated images. Further, all the recorded images in our experiments depicted the subject from at least the shoulders upward. This may not always be the case in practice, where only headshots from a driving register database may be available in a photo lineup. These differences may limit the generality of our findings to applied contexts, but they were necessary to ensure a fair test of the medium.

Conclusion

Our experiments were designed to test whether live identification is superior to photo and video identification when practical differences are minimized. Although suitable fillers are more difficult to obtain for live lineups, we omitted or equated fillers across conditions. Live lineups also take longer to organize than photo or video lineups, yet we imposed no difference in retention intervals. In addition, our participants knew it was an experiment by the time of identification and had less reason to be anxious about confronting the suspect than they would in the context of a real crime. If we had detected an advantage for live lineups under these experimental conditions, it would have left us wondering if the performance benefit would remain under more realistic conditions. But given that a live superiority effect was elusive even in our controlled experiments—and live lineups should perform worse without such controls—it would be worth reconsidering the preferential status given to live lineups in so many jurisdictions throughout the world.

Open Practices Statement

Data, analysis codes, and materials used in this experiment have been made publicly available on the Open Science Framework and can be accessed at osf.io/6r5b2.

Conflict of interest

We confirm that the manuscript "Live presentation for eyewitness identification is not superior to photo or video presentation" contains original data, has not been published elsewhere, and is not under consideration by another journal. All authors have approved the manuscript and agree with its submission.

Author Contributions

E. Rubínová and R. J. Fitzgerald developed the study concept, designed the experiments, organized and supervised the data collection events, and drafted the manuscript. E. Rubínová also created experimental stimuli and performed all data analyses. S. Juncu and E. Ribbers created experimental stimuli, collected data, and provided critical revisions on the manuscript. L. Hope and J. Sauer developed the study concept and provided critical revisions on the manuscript.

Acknowledgements

We are grateful to everyone who helped make data collection possible. The full list of acknowledgements is reported in Online Supplemental Materials.

References

BBC News (2003). Police offer virtual ID parades. Retrieved from: http://news.bbc.co.uk/1/hi/technology/2850803.stm

Bergold, A. N., & Heaton, P. (2018). Does filler database size influence identification accuracy? *Law and Human Behavior*, *42*, 227–243. http://dx.doi.org/10.1037/lhb0000289

Brace, N. A., Pike, G. E., Kemp, R. I., & Turner, J. (2009). Eyewitness identification procedures and stress: A comparison of live and video identification parades. *International Journal of Police Science & Management*, *11*, 183–192. http://dx.doi.org/10.1350/ijps.2009.11.2.122

Brewer, N. (2011, May 26). Pictures perfect: why photo lineups can be better at catching crooks. The Conversation. Retrieved from:

http://theconversation.com/pictures-perfect-why-photo-lineups-can-be-better-at-catching-crooks-1217.

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology*, *15*, 77–96. http://dx.doi.org/10.1348/135532509X414765

Bruer, K. C., Fitzgerald, R. J., Price, H. L., & Sauer, J. D. (2017). How sure are you that this is the man you saw? Child witnesses can use confidence judgments to identify a target. *Law and Human Behavior*, *41*, 541–555. http://dx.doi.org/10.1037/lhb0000260

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*, 243–248. http://dx.doi.org/10.1111/1467-9280.00144

Butcher, N., & Lander, K. (2017). Exploring the motion advantage: Evaluating the contribution of familiarity and differences in facial motion. *Quarterly Journal of Experimental Psychology*, *70*, 919–929. http://dx.doi.org/10.1080/17470218.2016.1138974

Cahlíková, J., & Cingl, L. (2017). Risk preferences under acute stress. *Experimental Economics*, *20*, 209–236. http://dx.doi.org/10.1007/s10683-016-9482-3

Clark, S. E., Moreland, M. B., & Rush, R. A. (2015). Lineup Composition and Lineup Fairness. In T. Valentine, & J. P. Davis (Eds.), *Forensic Facial Identification* (pp. 127–157). http://dx.doi.org/10.1002/9781118469538.ch6

Cutler, B. L., Berman, G. L., Penrod, S., & Fisher, R. P. (1994). Conceptual, practical, and empirical issues associated with eyewitness identification test media. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony* (pp. 163–181). http://dx.doi.org/10.1017/CBO9780511759192.009

Cutler, B. L., & Fisher, R. P. (1990). Live lineups, videotaped lineups, and photoarrays. *Forensic Reports*, *3*, 439–448. http://dx.doi.org/10.1037/law0000164

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology*, *73*, 281–290. http://dx.doi.org/10.1037/0021-9010.73.2.281

Dent, H. R., & Stephenson, G. M. (1979). Identification evidence: Experimental investigations of factors affecting the reliability of juvenile and adult witnesses. In D. P. Farrington, K. Hawkins, & S. M. Lloyd-Bostock (Eds.), *Psychology, law and legal processes* (pp. 196–206). London, UK: Palgrave Macmillan.

Duncan, M. J. (2006). *A signal detection model of compound decision tasks. (Technical Report No. TR2006 –256)*. Toronto, ON: Defence Research and Development Canada.

FeldmanHall, O., Raio, C. M., Kubota, J. T., Seiler, M. G., & Phelps, E. A. (2015). The effects of social context and acute stress on decision making under uncertainty. *Psychological Science*, *26*, 1918–1926. http://dx.doi.org/10.1177/0956797615605807

Fitzgerald, R. J., Price, H. L., & Valentine, T. (2018). Eyewitness identification: Live, photo, and video lineups. *Psychology, Public Policy, and Law*, *24*, 307–325. http://dx.doi.org/10.1037/law0000164

Fitzgerald, R. J., Rubínová, E., & Juncu, S. (in press). Eyewitness identification around the world. In Smith, A. M., Toglia, M., & Lampinen, J. M. (Eds.), *Methods, measures, and theories in eyewitness identification tasks*. Taylor and Francis.

Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika*, *32*, 25–33. http://dx.doi.org/10.1007/BF02289402

Hahn, C. A., O'Toole, A. J., & Phillips, P. J. (2016). Dissecting the time course of person recognition in natural viewing environments. *British Journal of Psychology*, *107*, 117–134. http://dx.doi.org/10.1111/bjop.12125

Hockley, W. E., Hemsworth, D. H., & Consoli, A. (1999). Shades of the mirror effect: Recognition of faces with and without sunglasses. *Memory & Cognition*, *27*, 128–138. http://dx.doi.org/10.3758/BF03201219

Lampinen, J. M. (2016). ROC analyses in eyewitness identification research. *Journal of Applied Research in Memory and Cognition*, *5*, 21–33. http://dx.doi.org/10.1016/j.jarmac.2015.08.006

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, *18*, 361–376. http://dx.doi.org/10.1037/a0030609

Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information need to perform ROC analysis? Then compute *d'*, not the diagnosticity ratio. *Journal of Applied Research in Memory and Cognition*, *3*, 58–62. http://dx.doi.org/10.1016/j.jarmac.2014.04.007

Mickes, L. (2015). Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition*, *4*, 93–102. http://dx.doi.org/10.1016/j.jarmac.2015.01.003

O'Toole, A. J., Phillips, P. J., Weimer, S., Roark, D. A., Ayyad, J., Barwick, R., et al. (2011). Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. *Vision Research*, *51*, 74–83. http://dx.doi.org/10.1016/j.visres.2010.09.035

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, *6*, 261–266. http://dx.doi.org/10.1016/S1364-6613(02)01908-3

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior*, *36*, 247. http://dx.doi.org/10.1037/h0093923

Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied*, *16*, 387–398. http://dx.doi.org/10.1037/a0021034

People v. Hoiland. (1971). *22 Cal. App. 3d. 530.*

Peters, D. P. (1991). The influence of stress and arousal on the child witness. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 60–76). Washington, DC: American Psychological Association.

Pike, G., Brace, N., & Kynan, S. (2002). *The visual identification of suspects: Procedures and practice (Briefing Note No. 2/02)*. London: Home Office Research Development and Statistics Directorate.

Price, H. L., Harvey, M. B., Anderson, S. F., Chadwick, L., & Fitzgerald, R. J. (2019). Evidence for the belief in live lineup superiority. *Journal of Police and Criminal Psychology*, *34*, 263–269. http://dx.doi.org/10.1007/s11896-018-9305-x

Quigley-McBride, A., & Wells, G. L. (in press). Methodological considerations in eyewitness identification experiments. In Smith, A. M., Toglia, M., & Lampinen, J. M. (Eds.), *Methods, measures, and theories in eyewitness identification tasks*. Taylor and Francis.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. http://dx.doi.org/10.2307/271063

Robbins, R. A., & Coltheart, M. (2012). The effects of inversion and familiarity on face versus body cues to person recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1098–1104. http://dx.doi.org/10.1037/a0028584

Sauer, J. D., Palmer, M. A., & Brewer, N. (2019). Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychology, Public Policy, and Law*, *25*, 147–165. http://dx.doi.org/10.1037/law0000203

Smith, A. M., Lampinen, J. M., Wells, G. L., Smalarz, L., & Mackovichova, S. (2019). Deviation from Perfect Performance measures the diagnostic utility of eyewitness lineups but partial Area Under the ROC Curve does not. *Journal of Applied Research in Memory and Cognition*, *8*, 50–59. http://dx.doi.org/10.1016/j.jarmac.2018.09.003

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Myerson, T. (2018). Eyewitness identification performance on showups improves with an additional-opportunities instruction: Evidence for present-absent criteria discrepancy. *Law and Human Behavior*, *42*, 215–226. http://dx.doi.org/10.1037/lhb0000284

Smith, A. M., Wilford, M. M., Quigley-McBride, A., & Wells, G. L. (2019). Mistaken eyewitness identification rates increase when either witnessing or testing conditions get worse. *Law and Human Behavior*, *43*, 358–368. http://dx.doi.org/10.1037/lhb0000334

State v. Buchanan. (2010). *9 La. App. 1288.*

State v. Scott. (2019). *191 Conn. App. 315.*

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*, 702–712. http://dx.doi.org/10.1177/1745691616658637

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior*, *22*, 217–237. http://dx.doi.org/10.1023/A:1025746220886

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods*, *6*, 371–386. http://dx.doi.org/10.1037//1082-989X.6.4.371

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior*, *20*, 479–496. http://dx.doi.org/10.1016/S0022-5371(81)90129-8

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373. http://dx.doi.org/10.1037/h0020071

Wells, G. L., Smalarz, L., & Smith, A. M. (2015). ROC analysis of lineups does not measure underlying discriminability and has limited value. *Journal of Applied Research in Memory and Cognition*, *4*, 313–317. http://dx.doi.org/10.1016/j.jarmac.2015.08.008

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*, 1115–1125. http://dx.doi.org/10.1177/01461672992512005

Wixted, J. T., & Mickes, L. (2015). ROC analysis measures objective discriminability for any eyewitness identification procedure. *Journal of Applied Research in Memory and Cognition*, *4*, 329–334. http://dx.doi.org/10.1016/j.jarmac.2015.08.007

Yovel, G., & O'Toole, A. J. (2016). Recognizing people in motion. *Trends in Cognitive Sciences*, *20*, 383–395. http://dx.doi.org/10.1016/j.tics.2016.02.005

Received 12 May 2020;
received in revised form 15 July 2020;
accepted 9 August 2020
Available online xxx