

This is the peer reviewed version of the following article: *Davis J. P., Gibson S. and Solomon C. (2014), The Positive Influence of Creating a Holistic Facial Composite on Video Line-up Identification, Applied Cognitive Psychology, 28, pages 634–639, doi: [10.1002/acp.3045](https://doi.org/10.1002/acp.3045)*, which has been published in final form at <http://dx.doi.org/10.1002/acp.3045>

The positive influence of creating a holistic facial composite on video lineup identification

Dr Josh P Davis

Department of Psychology and Counselling, University of Greenwich
Stuart Gibson and Chris Solomon
School of Physical Sciences, University of Kent

Corresponding author: Dr. Josh P Davis, Department of Psychology and Counselling,
University of Greenwich, Eltham, London, SE9 2UG, UK, +44(0) 208 8331 8859;
j.p.davis@gre.ac.uk

Dr Stuart J Gibson, School of Physical Sciences, University of Kent, Canterbury, Kent, CT2
7NZ, + (0) 1227 82 3271; s.j.Gibson@kent.ac.uk

Dr Chris J Solomon, School of Physical Sciences, University of Kent, Canterbury, Kent, CT2
7NZ, +44(0) 1227 823270; c.j.solomon@kent.ac.uk

Keywords: Eyewitness identification, EFIT-V, E-FIT, repeated identification, video lineup

Acknowledgements: The researchers would like to thank Kelty Battenti, Josi Cakebread, Ima Fagerbakke, Rebecca Fell, Beckie Hogan, Andreea Maigut, Amy Skelton, and Felicia Trembl for their assistance in collecting data for this project.

Part of this research was presented at the American Psychology-Law Society Annual Meeting, 4th International Congress of Psychology and Law, *Miami, Florida, USA*, March, 2011, and at the Identifying the Suspect: Improving Facial Composites Workshop, *Institute of Psychological Sciences, University of Leeds*, January 2013.

Chris Solomon and Stuart Gibson are faculty members of the University of Kent and directors of VisionMetric Ltd. VisionMetric Ltd markets the EFIT-V and E-FIT facial composite systems. Solomon and Gibson's contribution to this work was to facilitate the proportion of the data collection that took place at the University of Kent, development of the software, and provision of software support. Data collection, analysis and interpretation were performed by Davis.

The research was partly funded by an internal University of Greenwich grant to the first author.

Abstract

Witnesses to a crime may be asked to create a facial composite of the offender from memory. They may then view a suspect in a police lineup. Previous research on this topic has found both recognition impairment and enhancement following composite construction. In Experiment 1, creator-participants employed the holistic system EFIT-V or the feature-based E-FIT system to create a single composite, and in Experiment 2 creators constructed up to three EFIT-Vs. In both experiments, facial composite creators were one-and-a-half times more likely than non-composite creating controls to make correct target identifications from a video lineup. No between condition effects were found in target-absent trials in Experiment 1. The development of holistic facial composite systems has enhanced suspect identification rates in police investigations, and these results suggest that the use of such a system can also have a positive influence on a composite-creating witness' later recognition of the suspect.

Introduction

In a police investigation, eyewitnesses may, from memory, create a facial composite of the offender, and then subsequently view an identity lineup. Eyewitnesses may make mistaken identifications (Garrett, 2011) and accuracy can be enhanced (e.g., Meissner & Brigham, 2001), or impaired (e.g., Wells, Charman, & Olson, 2005) by composite creation. Indeed, eyewitnesses created composites in 46 of 250 US DNA-exoneration cases, with many identifying the innocent suspect from a lineup (Garrett, 2011). It is not credible to posit a direct link from composite to miscarriage of justice, as alternative case evidence may have been available. Nevertheless, this topic has implications for the fair administration of justice.

Studies finding positive effects from composite creation on recognition tended not to match normal witness experience (e.g., Maudlin & Laughery, 1981; Yu & Geiselman, 1993). Those finding negative effects employed *feature-based* systems which require the assembly of individual facial features to produce a final composite (e.g., Identi-Kit; Cornish, 1987; Photo-Fit, Davies, Ellis & Shepherd, 1978; FACES, Kempen & Tredoux, 2012; Wells *et al.*, 2005). For instance, Wells *et al.* (2005) found that in contrast to non-composite creating controls (84%), very few composite creators (10%) identified the target from a subsequent six-person simultaneous photo lineup - the majority rejected the lineup (58%). The authors suggest composite construction induces an analogue of the *Verbal Overshadowing Effect* (VOE; see Meissner & Brigham, 2001; Schooler & Engster-Schooler, 1990), or the negative influence of describing a face on its subsequent recognition. Feature-based composites such as FACES are often of poor quality (see Davies & Valentine, 2007 for a review), and the VOE is primarily implicated when verbal descriptions contain errors. However, for many creators, the most salient memory of the target may have been the composite, and the high rate of lineup rejections may be a consequence of no lineup member meeting this representation. In addition, some creators selected a foil. Creator's foil selections increase if

manipulated to resemble composites (Cornish, 1987), and the results of the Wells *et al.* study suggests that for the foil-selecting creators the most salient memory of the target may have been the composite, which through happenstance, resembled the selected foil.

In a second experiment, Wells *et al.* (2005) replicated their target-present results and additionally found no between-condition differences in target-absent trials, suggesting that composite construction does not increase innocent suspect identification risk.

Faces are primarily processed holistically (e.g., Tanaka & Farah, 1993), and recently introduced *holistic* composite systems (e.g., EFIT-V: Solomon, Gibson & Maylin, 2012; EvoFIT: Frowd, Bruce, Ness *et al.*, 2007; ID: Tredoux, Nunel, Oxtoby & Prag, 2007) may offer a solution. Instead of feature-based assembly, witnesses primarily make selections from whole face image arrays. Empirical research (Davis, Sulley, Solomon & Gibson, 2010; Frowd *et al.*, 2007; Tredoux *et al.*, 2007), and police investigation surveys (Frowd *et al.*, 2010; Solomon *et al.*, 2012) have demonstrated the superior quality of *holistic-system* composites over *feature-based* composites. For creators, their production might therefore induce less conflict with the original target memory, and indeed, the serial selection process might aid memory. Our rationale for the current research was to examine whether composite creation using one such system - EFIT-V - would *enhance* subsequent video lineup accuracy, by comparing the performance of composite *creators* with non-creating *controls*, as well as the *operators* who based decisions on composite memory only, providing a measure of suspect- and foil-composite similarity.

Experiment 1

In Experiment 1, creators and controls viewed a ‘suspect’ video. The creators and their paired operator then employed EFIT-V or the *feature-based* system E-FIT. Our aim was to isolate the influence of composite type on the subsequent target-present and target-absent video lineup outcomes of creators, operators and controls. There were a number of

hypotheses: 1) Due to EFIT-Vs holistic creation process, correct target-present lineup identification rates were predicted to be highest for EFIT-V creators in comparison to controls and E-FIT creators – with operators the least accurate; 2) Based on Wells *et al.*, no target-absent creator-control lineup differences were expected; 3) Consistent with Wells *et al.* (2005), a positive relationship was predicted between composite-target likeness ratings and correct creator target identifications; 4) Decision confidence was also collected immediately after the lineup decisions, and consistent with previous research (e.g., Sporer, Penrod, Read & Cutler, 1995), confidence was expected to be positively related to accuracy – particularly in lineup *choosers*; 5) Finally, creator-operator foil selections were expected to match at higher than chance rates – suggesting that for these creators the most salient memory of the suspect was that of the composite, rather than the suspect displayed in the initial video.

Design and participants

Experiment 1 employed a 2 (target presence: target-present, target-absent) x 3 (role: creator, operator, control) x 2 (system: E-FIT, EFIT-V) independent measures design. The dependent variable was identification accuracy. Forensic science undergraduate creators ($n = 117$; 66 female; *Mean age* = 21.1) and psychology undergraduate non-creating controls ($n = 151$; 123 female; *Mean age* = 22.6) viewed a video of a female ‘suspect’ unknown to all participants. Creators paired with an operator ($n = 117$; 62 female, *Mean age* = 20.3), produced either an E-FIT or an EFIT-V. All participants viewed a target-present or, containing a replacement foil, a target-absent lineup.

Procedure and materials

Operators and creators were firstly allocated to role and system. Creators and controls viewed the 1min 18sec ‘suspect’ video, prior to providing prospective target recognition confidence ratings to ensure that these were equivalent between conditions. Creators, assisted by an operator, then produced an E-FIT (production time = 35-60 min) or EFIT-V (20-50

min). All participants, reminded that the target may or may not be present, viewed a ‘surprise’ nine-person PROMAT video lineup constructed by an experienced police officer (see Valentine, Hughes & Munro, 2009 for a description of this procedure, proscribed in England and Wales by Code D of the Police and Criminal Evidence Act, 1984) and provided lineup decision and confidence data. All participation took approximately two hours.

The mock witness paradigm (Tredoux, 1998) demonstrated the lineups were unlikely to be biased, as target selections from the target-present lineup were at close to chance rates ($n = 46$; 10.9%), and most foils were plausible (Tredoux’s $E = 7.05$). For the target-absent lineup ($n = 42$), Tredoux’s $E = 4.85$.

Composite quality assessment: Additional target-unfamiliar ($n = 83$) and target-acquaintances ($n = 29$) provided target-composite likeness ratings (1-10) by viewing a series of random and counterbalanced slides, on which two video stills of the suspect were presented alongside a different composite¹.

Results

Table 1 depicts lineup outcomes and confidence as a function of target presence, role and system.

Table 1 about here

Prospective recognition: A non-significant t-test (role: creator, control) demonstrated that prospective recognition confidence was closely matched, $t(266) = .57$, $p > .2$.

Target-present trials: A 3 (creator, control, operator) x 2 (chooser, non-chooser) chi-squared test examining lineup selections by role regardless of accuracy was not significant, $\chi^2(2, n = 204) = 5.55$, $p = .062$, $\Phi = .165$. In contrast, a similar 3 (role) x 2 (correct, incorrect) chi-squared test examining the hypothesis that creator accuracy would be highest was significant, $\chi^2(2, n = 204) = 11.37$, $p = .003$, $\Phi = .236$. As predicted, creators made more

¹ Only 76 out of 117 composites were rated by target-acquaintances.

correct selections than controls (OR = 1.49; $p < .05$) and operators (odds ratio (OR) = 1.75; $p < .01$), who did not differ (OR = 1.18; $p > .2$).

With operators excluded, a 3 (E-FIT-creator, EFIT-V-creator, control) x 2 (chooser, non-chooser) chi-squared test examining lineup selections was not significant, $\chi^2(2, n = 141) = 1.24, p > .2, \Phi = .094$. A similar 3 (E-FIT-creator, EFIT-V-creator, control) x 2 (correct, incorrect) test was significant, $\chi^2(2, n = 141) = 6.94, p = .031, \Phi = .222$. EFIT-V creator selection rates did not significantly differ from E-FIT creators (OR = 1.10, $p > .2$). However, as predicted, EFIT-V creators were more accurate than controls (OR = 1.49, $p < .05$).

Target-absent trials: To test our predictions for target-absent trials, a 3 (creator, control, operator) x 2 (lineup rejection, foil identification) chi-squared test was conducted. This was significant, $\chi^2(2, n = 181) = 8.19, p = .017, \Phi = .213$. However, on the critical comparison there were as expected no differences between creators and controls (OR = 1.11, $p > .2$). Instead, creators (OR = 2.30, $p < .05$) and controls (OR = 2.07, $p < .05$) made more correct rejections than operators. With operators excluded, a 3 (E-FIT creator, EFIT-V creator, control) x 2 (lineup rejection, foil identification) chi-squared test was not significant, $\chi^2(2, n = 127) = .31, p > .2, \Phi = .049$.

Composite quality ratings: A Pearson's correlation test examining the relationship between the ratings given by the target-familiar and target-acquaintance raters was significant, $r(76) = .42, p < .001$. A 2 (E-FIT; EFIT-V) x 2 (rater: familiar, unfamiliar) mixed ANOVA on these ratings revealed a significant main effect of composite type, $F(1, 74) = 5.46, p = .022, \eta^2 = .069$, E-FITs received higher ratings ($M = 2.38$) than EFIT-Vs ($M = 2.12$). The interaction was also significant, $F(1, 74) = 18.10, p < .001, \eta^2 = .196$. Target-unfamiliar ratings were higher to E-FITs ($M = 2.53$) than EFIT-Vs ($M = 1.98, p < .05$). E-FIT ($M = 2.24$) and EFIT-V ($M = 2.26$) target-acquaintance ratings were virtually identical ($p > .2$).

In an unsupported test of our third hypothesis, two point-biserial correlations unexpectedly found no relationship between creator's target-present lineup accuracy and either target-acquaintance, $r(45) = -.05, p > .2$ or target-unfamiliar composite quality ratings, $r(63) = .04, p > .2$.

Confidence: A point-biserial test (correct = 1, incorrect = 0) revealed that there was an overall significant accuracy-confidence relationship across all participants, $r(268) = .34, p < .001$. To test for confidence differences by role, data were collapsed across target-present and target-absent trials in a 3 (E-FIT creator, EFIT-V creator, control) x 2 (accuracy: correct, incorrect) x 2 (lineup choosing: chooser, non-chooser) ANOVA. The main effect of role was significant, $F(2, 256) = 6.97, p = .001, \eta^2 = .052$. Tukey's test found no differences between E-FIT and EFIT-V creators ($p > .2$), although creators were more confident than controls ($p < .05$). The accuracy main effect was significant, $F(1, 256) = 5.95, p = .015, \eta^2 = .023$, confidence was higher for correct than incorrect decisions. The final significant effect was the accuracy x choosing interaction, $F(1, 256) = 7.65, p = .006, \eta^2 = .029$. Simple main effects revealed that as predicted confidence was higher for accurate than inaccurate choosers ($p < .05$). There were no differences in non-choosers ($p > .05$).

Operator-creator convergence: To test our final hypothesis, the lineup selections of creators and operators were examined for evidence of creator's memory being modified from suspect to composite. Unfortunately relevant cases numbers were infrequent, although operator and creator foil choices were consistent in 25% of the critical target-present trials ($n = 8$) - twice chance expectations (12.5%). Similarly, in target-absent trials ($n = 54$), no convergence could be measured when creators or operators rejected the lineup ($n = 27$). However, in 25.9% of the remaining 17 trials, creators and operators selected the same foil, more than twice chance expectations (11.1%).

Discussion

Experiment 1 demonstrated that as hypothesised, composite creation enhanced video lineup accuracy in target-present conditions, without influencing target-absent accuracy. The correct suspect identification rates of EFIT-V creators (70%) did not significantly differ from E-FIT creators (63%). However, EFIT-V creators were as expected more accurate than controls (45%), although the latter did not reliably differ from the operators (38%). Operators never viewed the suspect, and as the lineup was fair, this suggests their composites possessed a close likeness to the suspect. For lineup choosers, there was also as expected a positive relationship between accuracy and confidence.

Our expectations that EFIT-V creators would be more accurate than E-FIT creators were predicated on EFIT-Vs being superior to E-FITs. However, target-acquaintances' E-FIT and EFIT-V ratings were virtually identical, and in contrast to the findings of Wells *et al.* (2005), there was no relationship between creator lineup accuracy and composite quality. The current results can partly be explained as a consequence of the high composite-target similarity of the E-FITs – nearly half (42.4%) of the E-FIT operators made a correct identification of the target. For their creators, the composite creation process is unlikely to have conflicted with target memory. The highly similar composite-suspect ratings to E-FITs and EFIT-Vs provided by the target-acquaintances were virtually equal, contrast with the target-unfamiliar raters who provided higher E-FIT than EFIT-V ratings. Those tasked with making such decisions in real investigations will mainly be unfamiliar with suspects (e.g., police), and these results suggest that care should be taken when decisions as to composite utility are taken.

Operator and creator lineup data were compared to examine whether composite construction can increase foil identifications, if, through happenstance, they meet the composite's appearance. Although relevant case numbers were too low for any substantive conclusions, approximately one-in-four creator foil selections matched with operators

(25.8%) and the implications in terms of foil selection methods are covered in the General Discussion.

Experiment 2

A video of a single 'suspect' was employed in Experiment 1 and this raises the question of generalisability (see Wells & Windschitl, 1999, for a discussion of actor-specific effects). Furthermore, the lineups were conducted following a forensically unrealistic brief delay. These issues were addressed in Experiment 2 in which multiple actor videos were employed and lineup viewings were delayed by up to three days. Operator participation was a final component of EFIT-V and Cognitive Interviewing (CI) residential training courses, fulfilling accreditation requirements of the Association of Chief Police Officers (ACPO, 2009). Many operators had previously employed E-FIT (> 25%), and this provided a rare opportunity for research with professionals, who also administered the CI prior to each composite construction (see Memon, Meissner, & Fraser, 2010, for a review of this technique's positive effects).

Real eyewitness may create more than one composite if they believe the next could be better, and in Experiment 2, creators constructed between one and three composites, each with a different operator. Creating more than one composite has other benefits. *Morphing* up to four composites made by the same witness can improve recognition likelihood (Davis *et al.*, 2010; Valentine, Davis, Thorner, Solomon & Gibson, 2010), although effects are stronger when constructed by different witnesses (Valentine *et al.*, 2010). As a consequence, ACPO (2009) recommend *morphing* for real investigations. Nevertheless, multiple composites created by the same creator can vary in target likeness (Valentine *et al.*, 2010), and each construction might conceivably induce multiple, conflicting memory traces. The design therefore allowed a test of whether the positive effects on subsequent recognition found in Experiment 1 from creating a single composite would be replicated with multiple creations.

No target-absent trials were conducted, as no effects had been found in Experiment 1 or previous research (e.g., Wells *et al.*, 2005).

As creator target-present video lineup performance was compared to controls and operators, the primary hypotheses were consistent with Experiment 1. However, there is normally a negative relationship between delay and face recognition (Deffenbacher, Bornstein, McGorty, & Penrod, 2008), and similar effects were predicted in Experiment 2.

Method

Design and Participants

An independent measures design compared the video lineup outcomes of volunteer student composite creators ($n = 41$), controls ($n = 157$) and trained operators ($n = 39^2$). Creators and controls were randomly assigned to view one of six suspect videos depicting three male and three female actors, including the Experiment 1 actress. Creators were then paired with different operators to create one, two or three EFIT-V facial composites of the same suspect. All participants were invited to view a target-present PROMAT video lineup at the end of the training session, or online within 72 hours.

Procedure

The Experiment 1 procedure was replicated except composite production commenced after approximately 30 minutes of viewing the suspect video; operators employed the Cognitive Interview prior to EFIT-V construction with up to three creators who had each viewed different suspect videos. EFIT-V constructions took between 30 and 140 min. Controls viewed an initial suspect video in lectures. Video lineups were mainly viewed online within 72 hours.

² To protect anonymity no operator demographic data were collected, and data of E-FIT experienced operators were not separately analysed. Some operators failed to provide complete lineup data – 21 out of 39 (53.8%) provided data, to 45 out of 85 lineups (52.9%) within 72 hours. Data were excluded for missing lineups only.

Mock witnesses ($n = 46$) tested the fairness of the six lineups. Suspects were selected by between 16.7% and 9.5% ($M = 13.1\%$), suggesting none stood out. Tredoux's E (Tredoux, 1998) ranged from 6.73 – 5.69 indicating that most foils were plausible.

Composite quality assessment: Additional target-unfamiliar raters ($n = 41$) provided target-composite likeness ratings.

Results

Video lineup outcomes: The data associated with different actors were combined as initial analyses found no actor effects on lineup outcomes. As a function of role, Table 2 displays the mean delay prior to lineup viewings, lineup outcomes and confidence.

Table 2 about here

A chi-squared test examining lineup selections by role (creator, control, operator) was not significant, $\chi^2(2, n = 243) = 2.39, p > .2, \Phi = .099$. However, a similar test on accuracy was significant, $\chi^2(2, n = 243) = 6.41, p = .041, \Phi = .162$. Creators made more correct target selections than controls ($OR = 1.53, p < .05$) and operators ($OR = 1.86, p < .05$). The latter did not differ ($OR = 1.21, p > .2$).

Composite quality ratings: Three Pearson's correlation tests found no relationship between lineup accuracy and the target-unfamiliar ratings given to the creator's last composite, $r(41) = -.22, p = .174$, their best rated composite, $r(41) = -.19, p = .226$ or to the mean of their composites ratings, $r(41) = -.25, p = .116$. Indeed, in contrast to expectations the non-significant coefficients were negative.

Lineup decision confidence: With operators excluded, a point-biserial correlation revealed a significant lineup accuracy-confidence relationship, $r(196^3) = .48, p < .001$. A two-way ANOVA examining confidence by role (creators, controls) and accuracy (correct, incorrect) found only a significant main effect of accuracy, $F(1, 192) = 32.54, p < .001, \eta^2 =$

³ Two participants provided no confidence data.

.145. Confidence for correct responses was higher ($M = 73.5$) than incorrect decisions ($M = 52.6$).

Delay: With operators excluded, a point-biserial correlation conducted to test our predictions that there would be a negative relationship between delay and lineup accuracy was unexpectedly not significant, $r(198) = -.08, p > .2$. Neither was the correlation between delay and confidence, $r(196) = -.11, p = .143$.

Composite production number and lineup outcome: A Pearson's correlation found no relationship between creator composite number (one, two, three) and lineup accuracy, $r(41) = -.04, p > .2$, or confidence, $r(41) = -.08, p > .2$.

Discussion

Consistent with Experiment 1, EFIT-V creators made more correct lineup selections (48.8%) than controls (35.0%), and operators (28.9%), and there was again a positive relationship between lineup accuracy and confidence. However, unexpectedly there were no effects of delay, and the creation of up to three EFIT-Vs had no additional influence on lineup identifications. Morphing composites made by the same witness can enhance subsequent recognition, and these results suggest that multiple creations do not reduce correct lineup identifications.

General Discussion

When the police have no suspect, a witness, confident in their memory of the offender may work with an operator to create a facial composite. If a suspect is located, that witness may be asked to view them in a lineup. In both Experiment 1 and 2, composite creators were approximately one and a half times more likely than controls to make a correct identification from a video lineup, a figure close to that found in a previous meta-analysis (1.58: Meissner & Brigham, 2001). Consistent with previous research there were no differences in target-absent trials (e.g., Wells *et al.*, 2005).

Previous robust on this topic mainly found a negative composite production influence on recognition (e.g., Wells *et al.*, 2005). We believe that our findings are consistent with Wells *et al.*'s explanations in that construction of a poor quality *feature-based* composite is more likely to induce the VOE. The high creator lineup accuracy in the current research – particularly following EFIT-V construction, may be explained as a consequence of this system's holistic design. Matching our normal experience of real faces, the serial selection process encourages comparison of the configurations of facial features within arrays of whole faces and not analyses of isolated facial features. This is more likely to result in a better target likeness that is unlikely to conflict with memory of the target.

A further aim of the experiment was to examine whether composite creation would induce foil identifications if through happenstance they met the appearance of the composite. There was some evidence of this, although unfortunately, due to the low rates of foil identifications, conclusions are limited. Nevertheless, these results have implications in terms of foil selections. Wells, Rydell and Seelau (1993) argue that to reduce the likelihood of a lineup comprising 'look-a-likes,' foil selections should be based on a witness's description of the offender, and not on suspect-foil similarity. However, verbal descriptions often lack detail, and the police may select foils based on facial composite similarity. Our results suggest that for some witnesses composite memory may be the most salient. The faces included in such a lineup might be more likely to possess a close similarity to the composite, making distinguishing between them difficult. This might reduce correct identifications of an offender. It might also increase the likelihood of an innocent suspect identification, if included purely based on the appearance of a composite. Indeed, there has been at least one wrongful conviction following the creation of a lineup based on the appearance of a facial composite (Garrett, 2012). We therefore recommend that a police officer creating a lineup should not see the composite prior to construction.

Some caution must be taken with respect to generalizing these results. In both experiments there was a brief delay between initial suspect video viewing and composite production. However, many composites are constructed in the UK on the day of the crime, and it is not inconceivable that internet-based hand-held systems may one day be developed so that a remote operator can interact rapidly with a witness.

In summary, the recent development of holistic composite systems has seen an increase in suspect recognition rates (Frowd *et al.*, 2010; Solomon *et al.*, 2012). The results of the research reported here suggest their use may additionally increase the likelihood of a correct identification from a lineup. It should be stressed that we do not claim that EFIT-V or any similar system is entirely holistic-based, as there are elements of feature-based assembly in all systems. Nevertheless, these positive results should generalise to other similar systems (e.g., ID, EvoFIT), and the research reported here adds to a growing portfolio of evidence supporting the police use of holistic facial composite systems.

References

Association of Chief Police Officers (ACPO, 2009). *Facial Identification Guidance*.

Downloaded 3 January 2013 from

<http://www.acpo.police.uk/documents/crime/2009/200911CRIFIG01.pdf>

Cornish, S.E. (1987). Recognition of facial stimuli following an intervening task involving the Identi-kit. *Journal of Applied Psychology*, 72, 488-491.

Davies, G.M., Ellis, H., & Shepherd, J. (1978). Face identification: the influence of delay on accuracy of Photofit construction. *Journal of Police Science and Administration*, 6, 35-42.

Davies, G.M. & Valentine, T. (2007). Facial composites: forensic utility and psychological research. In: R.C.L. Lindsay, D.F. Ross, J.D. Read & M. P. Toglia, *Handbook of Eyewitness Psychology. Volume 2: Memory for People*. Mahwah: LEA (pp. 59-83).

- Davis, J.P., Sulley, L. Solomon, C., & Gibson, S. (2010). A Comparison of Individual and Morphed Facial Composites Created Using Different Systems. In G. Howells, K. Sirlantzis, A. Stoica, T. Huntsberger and A.T. Arslan (Eds.) *2010 IEEE International Conference on Emerging Security Technologies* (pp. 56 – 60).
- Deffenbacher, K.A., Bornstein, B.H., McGorty, E.K., & Penrod, S.D. (2008). Forgetting the once seen face: Estimating the strength of an eyewitness's memory representation. *Journal of Experimental Psychology: Applied*, 14, 139-150.
- Frowd, C., Bruce, V., Ness, H., Thomson-Bognor, C., Paterson, J., McIntyre, A., & Hancock, P. J. B. (2007). Parallel approaches to composite production, *Ergonomics*, 50, 562-585.
- Frowd, C.D., Hancock, P.J.B., Bruce, V., McIntyre, A., Pitchford, M., Atkins, R., et al. (2010). Giving crime the 'evo': catching criminals using EvoFIT facial composites. In G. Howells, K. Sirlantzis, A. Stoica, T. Huntsberger and A.T. Arslan (Eds.) *2010 IEEE International Conference on Emerging Security Technologies* (pp. 36-43).
- Garrett, B.L. (2011). *Convicting the Innocent: Where Criminal Prosecutions Go Wrong*. Harvard University Press.
- Kempen, K., & Tredoux, C.G. (2012). 'Seeing is believing': the effect of viewing and constructing a composite on identification performance. *South African Journal of Psychology*, 42, 434-444
- Mauldin, M.A., & Laughery, K.R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology*, 66, 351-357.
- Meissner, C.A., & Brigham, J.C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, 15, 603-616.

- Memon, A., Meissner, C.A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, & Law, 16*, 340-372.
- Schooler, J.W., & Engster-Schooler, T.Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 12*, 105-126.
- Solomon, C.J., Gibson, S.J., & Maylin, M. (2012). EFIT-V: Evolutionary algorithms and computer composites. In C. Wilkinson and C. Rynn (Eds.), *Craniofacial Identification*. Cambridge University Press (pp. 24-41).
- Sporer, S.L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence and accuracy: A meta-analysis of the confidence-accuracy relationship in eyewitness identification studies. *Psychological Bulletin, 118*, 315–327.
- Tanaka, J.W., & Farah, M.J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46A*, 225–245.
- Tanaka, J.W., & Sengco, J.A. (1997). Features and their configuration in face recognition. *Memory and Cognition, 25*, 583–592.
- Tredoux, C.G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior, 22*, 217-237.
- Tredoux, C.G., Nunez, D.T., Oxtoby, O. & Prag, B. (2007). An evaluation of ID: an eigenface based construction system. *South African Computer Journal, 37*, 90-96.
- Valentine, T., Davis, J.P., Thorner, K., Solomon, C., & Gibson, S. (2010). Evolving and combining facial composites: Between-witness and within-witness morphs compared. *Journal of Experimental Psychology: Applied, 16*, 72-86.
- Valentine, T., Hughes, C., & Munro, R. (2009). Recent developments in eyewitness identification procedures in the United Kingdom. In, Bull, R., Valentine, T., &

- Williamson, T. (Eds.). *The Handbook of Psychology of Investigative Interviewing*.
Chichester: Wiley.
- Wells, G.L., Charman, S.D., & Olson, E.A. (2005). Building face composites can harm
lineup identification performance. *Journal of Experimental Psychology: Applied*, *11*,
147–156.
- Wells, G.L., Rydell, S.M., & Seelau, E.P. (1993). The selection of distractors for eyewitness
lineups. *Journal of Applied Psychology*, *78*, 835-844.
- Wells, G.L., & Windschitl, P.D. (1999). Stimulus Sampling and Social Psychological
Experimentation. *Personality and Social Psychology Bulletin*, *25*, 1115-1125.
- Yu, C.J., & Geiselman, R.E. (1993). Effects of constructing Identi-kit composites on
photospread identification performance. *Criminal Justice and Behavior*, *20*, 280–292.

Table 1: Number of participants and percentage of each lineup outcome and mean confidence as a function of target presence, participant role and composite system in Experiment 1

Compo site	Total	Suspect ID		Conf	Foil ID		Conf	Rejection		Conf	
		n	n	%	n	%	n	%			
Target Present											
Controls	78	35	44.9	63.4	22	28.2	42.3	21	26.9	55.6	
Creators	EFIT	33	21	63.6	74.3	6	18.2	56.7	6	18.2	71.8
	EFITV	30	21	70.0	75.2	3	10.0	60.0	6	20.0	73.3
	Total	63	42	66.7	74.7	9	14.3	57.8	12	19.0	72.6
Operators	EFIT	33	14	42.4	62.9	15	45.5	50.7	4	12.1	58.8
	EFITV	30	10	33.3	46.0	17	56.7	40.6	3	10.0	43.9
	Total	63	24	38.1	55.8	32	50.8	45.3	7	11.1	52.0
Overall	204	101	49.5	66.3	63	30.9	46.0	40	19.6	59.8	
Target Absent											
Controls	73	-	-	-	45	61.6	49.6	28	38.4	71.4	
Creators	EFIT	27	-	-	16	59.3	66.3	11	40.7	68.2	
	EFITV	27	-	-	15	55.6	55.3	12	44.4	65.8	
	Total	54	-	-	31	57.4	61.0	23	42.6	67.0	
Operators	EFIT	27	-	-	20	74.1	58.0	7	25.9	54.3	
	EFITV	27	-	-	24	88.9	52.5	3	11.1	63.3	
	Total	54	-	-	44	81.5	55.0	10	18.5	57.0	
Grand Total	181	-	-	-	120	63.3	54.5	61	33.7	67.4	

Table 2: Mean delay (hours) after viewing the initial target video in Experiment 2 as well as the number and percentage of each target-present lineup outcome, and mean confidence in that outcome

	Mean	Total	Target ID			Foil ID			Rejection		
	Delay	<i>n</i>	<i>n</i>	%	<i>Conf</i>	<i>n</i>	%	<i>Conf</i>	<i>n</i>	%	<i>Conf</i>
Controls	32.1	157	55	35.0	72.5	48	30.6	50.0	54	34.4	53.2
Creators	31.1	41	20	48.8	75.9	7	17.1	71.7	14	34.1	50.8
Operators	26.5	45 ¹	13	28.9	58.5	22	48.9	50.0	10	22.2	57.0
Total	30.9	243	88	36.2	71.3	77	31.7	51.7	78	32.1	53.3

¹ *This value indicates the number of lineups viewed by operators. Some viewed more than one as they created more than one composite and some operators failed to view all lineups within 3-days. These data were not included.*