EvoFIT composite face construction via practitioner interviewing and a witness-administered protocol

Abstract—Police require reliable facial-composite systems to help identify, arrest and convict criminals. Recent developments, however, have allowed newer versions of the EvoFIT composite system to be made available for policing. The outcome is an online (cloud-based) version and a new system called Witness At Home, both using a simpler interface. Here, we formally compare these two versions to establish potential benefits to policing. Two experiments were conducted. In Experiment 1, participants observed a target identity for 1 minute and returned 4 hours (Witness At Home) or 24 hours (EvoFIT Online) to construct a composite from memory. No significant difference in composite accuracy was found. In Experiment 2, participants constructed a composite, 24-hours after seeing a target identity, using either EvoFIT Online or Witness At Home. A significant increase in accurate identification was found for EvoFIT Online, with some utility for the self-administered procedure, together indicating benefit for these newer systems plus some areas for development.

Keywords: EvoFIT; Witness At Home; Internet procedures; Internet connection speed; Composite naming

I. INTRODUCTION

Where police find no physical evidence (e.g., fingerprints, CCTV images) at a crime scene, they may ask any witnesses present at the time to produce a facial composite (or likeness) of the suspect [1]. Developed over the past 50 years, most early systems (e.g. PhotoFit) required witnesses to select from sets of individual facial features (e.g. a pair of eyes, a nose, a mouth) [2]. Police would initially interview a witness using a cognitive interview (see [3] for details), designed to aid witness recall and obtain a detailed description of the offender’s face. Witnesses and victims would then select from individual facial features to construct a likeness of the offender’s face. Under certain circumstances, in particular where composite construction is carried out immediately or within a few hours of viewing a target, “feature” systems result in composites with mean correct naming of up to 20% [4]. However, inclusion of a forensically-relevant delay of upwards of 24-48 hours results in composites with mean naming of only a few percent correct (e.g., [5, 6]).

The problems involved are at least two-fold. Firstly, previous research (e.g. [7, 8]) has found that recall of faces decays rapidly, which may account for inferior composites constructed 24-48 hours after a target has been seen. Further, if the witness has not had a good look at the offender’s face or does not have good recall of the facial features, this should have a detrimental effect on subsequent composite naming accuracy. The latter effect is likely to occur irrespective of when composite construction is attempted. Secondly, there appears to be difficulty with the processing shift from unfamiliar to familiar face recognition [9]. Unfamiliar faces are generally much more difficult to recognise and witnesses typically rely on external features (e.g. hair and face width) to guide identification [8, 10]. In contrast, familiar face recognition relies more on internal features (e.g. eyes, nose and brows) and is much easier to achieve accurately [11]. Feature systems are thus fundamentally flawed as they do not reflect the mechanism by which faces are naturally processed.

Consequently, newer approaches such as EFIT-V [12], ID [2] and EvoFIT [13] were developed to better reflect recognition-based (or holistic) face processing mechanisms. The process of evolving a face using a holistic system (e.g., see [2] for detailed procedures using EvoFIT) reduces the impact of recall and utilises more robust recognition (e.g., [14]). Despite this, early trials of the EvoFIT implementation produced composites with low naming levels, of around 10% [15]. However, naming levels have increased following different iterations of the system (e.g. blurring of and eventual removal of external features, [16]) and the introduction of enhanced
interviewing techniques (character-based or “holistic” mnemonics of the cognitive interview [17]). Consequently, EvoFIT has been found to have high identification levels, at least 60% in both the lab setting and in police field trials [16].

More recently, EvoFIT developers have created an online alternative, one utilising a potentially much simpler user interface than the “standalone” version currently used by forensic practitioners to interview witnesses. Also, these developers have created a new system that could improve composite effectiveness while reducing impact on police resources. Called Witness At Home (WAH), witnesses can use this system by themselves in their own home or any location with internet access. Police can issue an email link which allows a witness access to the online face construction site. Witnesses are directed to make a witness statement as well as constructing a likeness of the offender. Finally, the finished composite face and witness statement are forwarded to police for use in the relevant criminal investigation. This entire procedure is time efficient (about an hour) and, crucially, could be deployed to the witness within a few hours of a crime being witnessed. EvoFIT has been shown to support superior recognition processes [18], therefore, this shorter delay should improve recall, resulting in improved composite recognition and ultimately further facilitate conviction rates.

Further, a visit to the police station could potentially elevate stress levels or the witness may simply be unsure of procedures and thus be overly anxious, see [19], either of which could adversely affect witness performance [20]. In contrast, working alone, perhaps without perceived raised expectations, may diminish performance stress [21] and increase the ability to recall the face [22]. Moreover, as no forensic practitioners are required to be in attendance, this protocol may help reduce demands on police resources. Indeed, modern policing needs to be efficient, and as such a self-administered procedure might be welcome for less serious crimes (e.g. minor theft or vandalism), in particular when police resources are strained.

In the current work, we look to evaluate the efficacy of the Witness At Home procedure in comparison to the EvoFIT Online system. In Experiment 1, we investigated whether the Witness At Home procedure might be able to take advantage of superior recall offered by same day construction. We anticipated that composites from Witness At Home would be more effective than EvoFIT Online, due to potentially less witness memory decay. In Experiment 2, we conducted a more direct comparison between the two types of system, one using standard interviewing techniques and face constructions conducted after a 24-hour delay.

II. EXPERIMENT 1: SYSTEM USE UNDER PRACTICAL SITUATIONS

A. METHODS

a) Participants - Stage 1: Face Construction: Participants were staff and students from the University of Stirling and members of the community in Stirling. All had normal or corrected-to-normal vision. Target images were photographs of lecturers from this university, selected to be unfamiliar to face constructors. Twenty participants, age range 18 to 55 years (\(\text{Mage} = 40\) years, \(\text{SD} = 15\) were randomly allocated to either EvoFIT Online (Condition 1) or EvoFIT Witness At Home (Condition 2).

b) Stage 2: Composite Naming: Twenty-four participants (14 female), recruited through opportunity sampling from university staff and students, age range 22 to 65 years (\(\text{Mage} = 39.9\) years, \(\text{SD} = 11.6\) evaluated composites (12 per condition) by attempting to name them. All volunteers had normal or corrected-to-normal vision.

c) Materials: Stimuli comprised digital images of 10 members of Psychology department staff (3 female). Images were colour, full faced frontal photographs, clear and of neutral facial expression. Two sets of 10 photographs, size 18cm high by 13cm wide, were in colour on A4 paper for presentation to Stage 1 participants. One set of images was placed in a random order in white envelopes marked 1-10 (Condition 1). Similarly, the other set was placed in envelopes marked 11-20 (Condition 2). A stopwatch was used to ensure accurate target encoding in Stage 1, and answers from the interview in Stage 2 were recorded manually on a paper sheet.

Manual instructions, including experimental procedure, were used for Condition 1 participants. This comprised introduction, cognitive interview and holistic-cognitive interview instructions, currently used with EvoFIT Online. Different manual instructions, for Witness At Home, were given to participants in Condition 2, including a self-administered interview form.

Composites were constructed on a desktop computer using appropriate EvoFIT software as per condition. Two booklets were designed comprising a total of 14 composites from each condition; 10 target images and four foils (unfamiliar composites not from the target images, used to minimise guessing or process of elimination type strategies) were placed in each booklet. A third booklet contained the 10 original digital images of staff.

B. PROCEDURE

The ‘gold’ standard procedure for facial composite construction (see [13]) was utilised. This required laboratory conditions mirror real life circumstances as closely as possible. Firstly, target images should be unfamiliar to constructors but familiar to evaluators. Secondly, a forensically relevant delay should be implemented between participants viewing the target image and attempting composite construction. This delay was dependent on which condition participants had been randomly allocated (approximately 24 hours for EvoFIT Online and 4 hours for Witness At Home). Finally, interviewing techniques varied by condition. Condition 1 used elements of cognitive interviewing (free recall) and holistic-cognitive interviewing, including a new enhanced holistic interview which required focus on eye region only. For Condition 2, a self-administered interview form was completed.

a) Face-to-Face Construction Stage: Participants followed procedures typically recommended to forensic practitioners to achieve best results [2]. Here, participants engaged in a cognitive interview (e.g. recall everything about the target face) and a holistic interview (e.g. to judge perceived personality traits about the target, such as pleasantness, health and honesty). The latter interview was the recently enhanced version of this interview mnemonic that invited witnesses to
focus on the eye region while again making character attribution.

Once the interview had been completed, participants were guided through face construction by the researcher. See [2] for detailed EvoFIT procedures. In brief, this involved participants selecting a face database based on ethnicity, gender and age of their target (e.g. White, European, Female, 20 years). Next, they selected from arrays of internal features—that is, where no external features were present, as this is thought to focus attention on features critical to familiar face recognition. Faces were requested to be selected based on the area around the eyes, but ignoring face width, as this aspect of the face can be adjusted accurately later. The process proceeded through two generations of smooth faces, textured faces and a combination of both smooth and texture. Further adjustments were made available using holistic tools (for altering face characteristics such as width, weight, health, honesty and masculinity) and shape tools (e.g. to adjust thickness of lips, size of eyes), with participants working towards producing the most recognisable face. Finally, external features (hair, ears and neck) were added, and further adjustments made to the face as required. The completed composite was saved to disk. Participants were thanked and debriefed.

b) Witness At Home Construction Stage: Participants worked online to construct the face: they initially viewed an instructional video and then completed face construction on their own. Procedures were equivalent to the EvoFIT Online procedure in terms of selecting the face database, smooth and textured faces, using holistic and shape tools, and adding external features. The researcher remained on hand to answer any questions a witness might have; in all cases questions were limited to procedural (e.g. should I press the ‘next’ button) rather than featural issues (e.g. do you think the nose is big enough). See Fig. 1 for example composites produced in experiment 1.

c) Evaluation Stage Procedure
Participants were tested individually, and the task was self-paced. Participants were presented with two booklets of 14 composites, some of which were university lecturers with which they should be familiar. The 14 composites seen comprised five from Condition 1, five from Condition 2 and four random foils not of university lecturers (‘foil’ composites); there were two booklets of stimuli containing different composite identities from the experiment. Participants viewed individual composites sequentially, in a different random order for each person, and were asked to try to identify the identities depicted therein. A correct response was recorded if a participant gave the correct name or an unambiguous semantic description such as “she taught health psychology lectures in second year”. Booklets were randomised across participants with equal sampling. For example, participant 1 saw Booklet 1, participant 2 saw Booklet 2, participant 3 saw Booklet 2, etc. Booklet 3, contained the 10 original target images, was then presented to participants to name, as a check that targets were known to them. Also, as part of an a priori rule, participants were required to be familiar with at least eight out of the ten targets for their data to be included in the following analysis. In cases where participants failed to meet this a priori rule, another participant was recruited as replacement. Testing sessions took approximately 10 minutes to complete.

C. RESULTS

The following analyses were conducted on the effectiveness of the composites from participants, by examining both accurate and inaccurate responses, as well as by assessing average time taken to construct the faces themselves.

a) Accurate Composite Identification: The data were checked for missing data, of which no cases were found. Conditional naming scores were first analysed for correct responses. This method considers the number of composites that a person has correctly named relative to the number of targets known. For example, if a participant correctly named two composites but only correctly named eight targets, his or her score was calculated as 2 / 8 (25%). A paired-samples t-test conducted on these conditional naming scores from participants found no significant difference between composites from EvoFIT Online (M = 25.4%, SD = 4.2%) and Witness At Home (M = 24.8%, SD = 2.7%), t(23) = 0.124, p = .90, two tailed, Cohen’s d = 0.17.

b) Incorrect Composite Identification: The data were re-scored based on the number of incorrect (mistaken) names provided by participants. In this case, a lower score indicates a more accurate composite. For this analysis, a paired-samples t-test revealed that composites from EvoFIT Online (M = 12.4%, SD = 1.9%) produced significantly fewer incorrect names than composites from Witness At Home (M = 21.7%, SD = 4.1%), t(23) = 2.20, p = .038, two tailed, d = 2.91.

c) Composite Construction Time: Mean face construction time was calculated. Witness At Home composites (M = 64min) took somewhat longer to construct than EvoFIT Online composites (M = 60min), although this difference was not significant using an independent-samples t-test, t(18) = 0.60, p = .56, two tailed, d = 0.02.

d) Discussion: No significant difference was found in correct (conditional) naming between conditions. Six of the 20 composites received no correct naming (4 from Witness At Home condition). In contrast, incorrect naming levels for Witness At Home were significantly higher, suggesting that these composites were of a worse standard. For Witness At Home, participants constructed composites on their own with no help from a trained practitioner. In several sessions,
procedural questions were raised. For example, “Why does the ‘Next’ button not work when I press it?” Such examples suggested that at least some participants were unable to follow the on-screen instructions correctly. In addition, there were occasions involving appreciable delays (slow internet speed) and these may have left participants frustrated with the process, negatively affecting performance. Notably, these system issues were consistent across both conditions, which may have contributed to overall lower naming levels than typically found in EvoFIT laboratory trials (35-50%, see e.g., [23, 24]). That said, correct naming levels were similar to a previous study [25] conducted under similar internet conditions, supporting the notion that slower internet speeds may adversely affect composite accuracy.

As a result, a second experiment was conducted using a more effective internet connection, to determine if this situation would lead to more effective composite naming.

III. EXPERIMENT 2: SYSTEM COMPARISON TESTED UNDER EQUIVALENT FORENSIC CONDITIONS

A. METHODS

a) Participants - Stage 1: Face Construction: Participants were students at University of The West of England and members of the community in Bristol. All participants had normal or corrected-to-normal vision. Target images were characters from the ITV soap Coronation Street (five male), with participants recruited to be unfamiliar with these target identities. Twenty participants, age range 18 to 63 years (Mage = 30 years, SD = 17.0) were randomly allocated by condition, 10 in each.

b) Stage 2: Composite Naming: Twenty-one participants (14 female) recruited through opportunity sampling from staff and students at University of West of England, age range 18-76 years (Mage = 45.2 years, SD = 18.7) named the composites (11 EvoFIT Online, 10 Witness At Home). All participants had normal or corrected-to-normal vision.

c) Materials: As Experiment 1, with the exception that target identities were of characters from Coronation Street.

B. PROCEDURE

a) Face-to-Face Construction Stage: As in Experiment 1, with the exception that no interviews (cognitive or holistic-cognitive) were conducted to aid recall. Face construction was carried out after a 20 to 28-hour delay.

b) Witness At Home Construction Stage: Procedure was as Experiment 1, with the exception that no self-administered questionnaires were completed, and face construction was carried out after a 20 to 28-hour delay (again following 30 second encoding of the target). See Fig. 2 for example composites produced in Experiment 2.

C. RESULTS

The analysis focussed on conditional and composite inaccurate naming scores, as in Experiment 1.

a) Accurate Composite Identification: A paired-samples t-test indicated a reliable benefit of composites from EvoFIT Online (M = 34.3%, SD = 16.8%) over Witness At Home (M = 21.1%, SD = 8.6%), t(19) = 2.20, p = .043, two tailed, d = 1.03.

b) Incorrect Composite Identification: Only three false mistaken (inaccurate) names were given in total. Thus, due to infrequent cases of this type, no statistical tests were conducted on these data.

c) Composite Construction Time: Composite construction time for both conditions averaged 45 minutes.

d) Discussion: Experiment 2 was conducted as a direct comparison of the two EvoFIT systems, EvoFIT Online (practitioner help with construction) and Witness At Home (no practitioner help, witness administered), in order to assess the effectiveness of EvoFIT face construction procedures. Additionally, we were interested in assessing any effect of a relatively faster internet connection on naming accuracy. Neither of the face construction procedures used in this experiment involved additional aid to face recall.

A significant effect was found in the correct naming task indicating a benefit for EvoFIT Online. Naming levels for EvoFIT Online were at a higher level than Experiment 1 and were more closely aligned to previous EvoFIT laboratory trials (e.g., [23,24]). While internet speeds were faster (screen loading typically took 3 to 7 seconds) than in Experiment 1 (12 to 15 seconds), the somewhat lower correct naming levels for Witness At Home suggest it is more difficult for witnesses to construct a composite without the assistance of a forensic practitioner.

IV. GENERAL DISCUSSION.

Facial composites constructed with traditional ‘feature’ systems (so called as they require the selection of individual features such as, a pair of eyes, a nose, a mouth) shortly (e.g. immediately or within 3-4 hours) after target face encoding are usually named at around 20% correct [4]. However, the introduction of a forensically relevant delay, in the region of
24 to 48 hours, results in worse performance [5, 6]. The development of holistic (or recognition-based) systems (e.g., EvoFIT) with enhanced interviewing techniques (e.g., cognitive and holistic-cognitive interviews, designed to aid recall) produce composites with much higher correct naming (e.g., 60%, see [16]), despite a delay of 24 to 48 hours. These enhanced interviewing techniques have been shown to improve recall beyond the level achieved by same-day construction [17]; however, it could be the case that same day construction with some sort of facilitated recall could result in even better naming levels. Unfortunately, no system appears to be readily available that a witness could sensibly use early on in a police investigation.

A. WITNESS AT HOME

Recently developed, the Witness At Home protocol for EvoFIT may be the first implementation that could be deployed on the same day a crime had been witnessed. Witness recall should be better [4] and with the addition of a self-administered interview form (designed to aid recall) could produce more accurate composites. This in turn could improve identification levels and thus provide the police with the tools to increase arrests and conviction rates [16].

a) Composite Naming: Experiment 1 saw no reliable difference between 4 hour and 24-hour delay conditions. Furthermore, naming levels were lower for both conditions (when compared to typical EvoFIT laboratory trials, e.g., [23, 24]) and there was a significantly higher false identification rate for Witness At Home. In Experiment 2, where both systems were deployed without the structured interviews to aid recall, a significantly higher naming level (34.3%) was found for EvoFIT Online. Moreover, there was no corresponding increase in false identification. Over both experiments, a majority of composites were identified by at least 1 evaluator (14 of 20 and 15 of 20 in Experiments 1 and 2 respectively), suggesting that images were recognisable to some extent.

While overall naming levels were higher for EvoFIT Online, the top performing composite in Experiment 1 (11 of 12 accurate identifications, Fig. 1, far right) turned out to be constructed using the Witness At Home system. This may highlight the potential of the system, but also the importance of individual differences in ability of constructors to create an effective image [26]. However, other considerations may explain lower naming levels in Experiment 1 as well as the lower correct naming rate for Witness At Home in Experiment 2.

B SYSTEM ISSUES

Average construction times were around 15 minutes longer for Experiment 1 (cf. Experiment 2). This outcome is likely to be due to worse internet connectivity in the first experiment. Average construction times for EvoFIT Online here (60 minutes) were similar to [25] (64 minutes). Both studies were conducted under virtually identical experimental conditions (same location, equipment and internet service). Delays (e.g., a screen loading slowly) in Experiment 1 may have left participants feeling frustrated, negatively affecting their performance. This may also be an issue when participants were constructing composites on their own using Witness At Home. This system has a 7-minute instructional video that is played at the start along with written on-screen guidelines throughout the process. It is clear, though, that participants were able to construct a face reasonably well. This observation is based on fairly good correct naming of their composites (M = 21.1%). However, there were a few questions of a procedural nature (e.g., “What do I do now?”) raised by participants, suggesting some difficulty with following the on-screen instructions.

C FINAL COMMENTS

EvoFIT composites constructed under laboratory conditions have seen accuracy levels improve from a few percent correct [5, 6] to around 35 to 45% [23, 24], and higher [18]. At first glance, the lower naming levels found here (both Conditions in Experiment 1, and Witness At Home in Experiment 2) may give cause for concern. However, both Experiments were conducted using the internet, and speeds are variable. Experiment 1 had higher construction times of 15 minutes on average than Experiment 2. Frequent delays in screens’ loading do suggest slower connections. Consequently, participants may have become frustrated with the process, promoting less effective composites. Furthermore, there was evidence (from researchers present during construction) that some participants had difficulty with some of the on-screen instructions while using Witness At Home. This may contribute to the lower correct naming levels in Experiment 2, Witness At Home. Conversely, EvoFIT Online, in Experiment 2, did reveal a more positive result, despite no enhanced interviewing techniques being employed. Previous research has shown these enhanced interviewing techniques (e.g., cognitive and holistic-cognitive [17], focused breathing [25]) to be of some use in facilitating recall. In the field, police use these enhanced interviewing techniques with the standalone version of EvoFIT, where screen loading times are fast and results of field trials indicate 60% identification accuracy [16].

Soon, a “standalone” (non-internet) version of EvoFIT, using the simpler interface used here, will be released. This new version should help reduce system delays and when combined with the enhanced interviewing techniques should also improve composite accuracy and subsequent composite identification. Witness At Home shows promise. We think no other system could reasonably be deployed, by police, on the same day a crime was witnessed. The apparent difficulties with participants and instructions will need to be addressed. In the first instance, we intend adding voice prompts where required (e.g., “Choose 2 faces from this screen and press the next button”) to help focus participants’ attention. Our intention is to trial these new procedures, to ideally improve performance, but it is clear that, even as it is, the witness-at-home procedure has fairly good utility with mean naming at around 20-24% correct.

V. REFERENCES


