Applied Cognitive Psychology, *Appl. Cognit. Psychol.* **26**: 308–315 (2012)
Published online 22 September 2011 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/acp.1826

The Headscarf Effect: Direct Evidence from the Eyewitness Identification Paradigm

AHMED M. MEGREYA[1,2]*, AMINA MEMON[3] and CATRIONA HAVARD[4]

[1]*Department of Psychology, Menoufia University, Shebin El-Kom, Egypt*
[2]*Umm Al-Qura University, University College at Qunfudah, KSA*
[3]*Department of Psychology, Royal Holloway University of London, London, UK*
[4]*Department of Psychology, University of Aberdeen, Aberdeen, UK*

Summary: Internal and external features dominate familiar and unfamiliar face recognition, respectively. However, this finding is not universal; Egyptians showed a robust internal-feature advantage for processing unfamiliar faces (Megreya & Bindemann, 2009). This bias was speculatively attributed to their long-term experiences for individuating female faces with headscarves, which completely cover the external features. Here, we provided an empirical test for this suggestion. Participants from Egypt and UK were presented with a staged crime, which was committed by an own-race woman with or without a headscarf. All participants were then asked to identify the culprit from a line-up involving 10 faces with or without headscarves. British participants showed an advantage when the culprit left her hair uncovered. In contrast, Egyptian observers showed an advantage when the culprit wore a headscarf. This Egyptian headscarf effect was also replicated using British faces, suggesting that it reflects a specific characteristic of participant nationality rather than face nationality. These results therefore provide evidence for how culture influences cognition. Copyright © 2011 John Wiley & Sons, Ltd.

INTRODUCTION

Face recognition research provides consistent evidence that the central region of a face that contains the brows, eyes, nose and mouth (the internal features) aids perception and recognition more efficiently than its outer parts that include the hair, ears and chin (the external features), but only when faces are *familiar* (Bonner, Burton, & Bruce, 2003; Campbell et al., 1999; Clutterbuck & Johnston, 2002; Ellis, Shepherd, & Davies, 1979; Young, Hay, McWeeny, Flude, & Ellis, 1985). In contrast, the external features become more important when faces are *unfamiliar* (Bonner et al., 2003; Bruce et al., 1999; Frowd, Bruce, McIntyre, & Hancock, 2007; Nachson & Shechory, 2002; Want, Pascalis, Coleman, & Blades, 2003). Accordingly, it has been recently suggested that the well-established dissociation between familiar and unfamiliar face processing (e.g. see Megreya & Burton, 2006, 2007) may be located in the way the internal features are encoded (Bonner et al., 2003; Clutterbuck & Johnston, 2002). Therefore, the internal-feature advantage is viewed as an indirect measure of face familiarity (Clutterbuck & Johnston, 2002, 2004, 2005; Osborne & Stevenage, 2008).

However, this view was recently challenged. Megreya and Bindemann (2009) presented Egyptian participants with an image of a whole male or female unfamiliar face and another concurrent face image in which either the internal or external features had been removed. Notably, all female faces were wearing headscarves, which completely covered their external features (see Figure 1 for example). Participants' task was to decide whether both images showed the same or two different persons. As expected, participants were able to match female faces in the internal condition (where all internal features were clearly visible) much more accurately than in the external condition (where the hair and ears were completely covered by a standard headscarf). Intriguingly, however, Egyptian participants also showed superiority for matching male faces from the internal features. This internal-feature advantage (for male faces) was robust, and it appeared using a variety of matching tasks, with no effect for the sex of observers. Furthermore, Egyptian observers showed an internal-feature advantage for both Egyptian *and* British (Caucasian) faces. On the other hand, British participants showed an external-feature advantage for all of these face stimuli, consistent with previous research with Caucasian participants (e.g. Bonner et al., 2003; Bruce et al., 1999; Frowd et al., 2007; Want et al., 2003). Therefore, Megreya and Bindemann (2009) concluded that these differential effects reflect characteristics of participant nationality rather than face nationality.

Megreya and Bindemann (2009) have attributed these cross-cultural differences to the extensive experience of Egyptians at recognising female faces with headscarves, which are commonly worn in Egypt and the vast majority of Middle Eastern countries. It has been suggested that perceptual expertise critically affects the function and plasticity of the visual system (e.g. see for reviews Bukach, Gauthier, & Tarr, 2006; Gauthier & Nelson, 2001). For example, Gauthier and Tarr (1997) extensively trained participants to be experts at identifying a set of novel objects (called 'Greebles') that made them process these stimuli using the behavioural and neural mechanisms involved in face recognition. For example, this training enabled the participants to differentiate Greebles using configural information, which is considered a key component of face processing (e.g. Bartlett & Searcy, 1993; Freire, Lee, & Symons, 2000; Tanaka & Sengco, 1997). In addition, the training selectively activated the brain regions (e.g. the Fusiform Face Area), which are commonly thought to be involved in face processing (e.g. for a review see Haxby, Hoffman, & Gobbini, 2000). Similar effects were observed using experts of other non-face objects such as dogs (Diamond & Carey, 1986), birds (Gauthier, Skudlarski, Gore, & Anderson, 2000; Rhodes & McLean,

*Correspondence to: Ahmed M. Megreya, Department of Psychology, Faculty of Arts, Menoufia University, Shebin El-Kom, Egypt.
E-mail: amegreya@yahoo.com

Figure 1. The line-up images of the British and Egyptian targets with and without headscarves

1990), and cars (Gauthier et al., 2000). Therefore, it appears that any putatively face-specific cognitive and neural mechanisms are attributed to the high expertise of identifying individual faces, which are one of the most encountered visual stimuli in the real life (e.g. for reviews see Bukach et al., 2006; Gauthier & Nelson, 2001; but see McKone, Kanwisher, & Duchaine, 2006 for a different viewpoint).

The aim of the present paper was to provide a direct test for what is termed the headscarf effect (Megreya & Bindemann, 2009) using the ecologically valid eyewitness identification paradigm. Participants from Egypt and UK were presented with a video of a staged theft, which was committed by an own-race woman with or without a headscarf. After a short delay, all participants attempted to identify the culprit using target-present (TP) or target-absent (TA) photographic line-ups. The effects of the presence of a headscarf at study *and/or* test were examined. In a further experiment, we examined whether this Egyptian headscarf effect could generalise to other-race faces.

EXPERIMENT 1

Method

Participants

A total of 952 Egyptian and British undergraduate students (from the Universities of Menoufia and Aberdeen, respectively) volunteered to participate in this study. The Egyptian sample consisted of 485 students (176 men), with a mean age of 19.2 years ($SD = 0.7$), and the British group included 467 participants (141 men), with a mean age of 19.7 ($SD = 2.41$). All of these participants had been resident in their native country for a majority of their lives.

Stimuli

The stimuli comprised four video clips showing an Egyptian or British actress taking part in a staged theft of an exam. In each clip, a young woman entered an unoccupied university office and searched shelves and a filing cabinet to find an exam sheet. This event lasted approximately 90 seconds, and there were several opportunities to record the target's face closely in frontal and profile viewpoints. The video clips were identical, except that, in each country, the actress wore a headscarf in one version (the headscarf study condition) but left her hair uncovered in another version (the no headscarf study condition). Immediately after acting in the videos, two full-face photographs were taken from each actress while they were wearing the headscarves or left their hair uncovered (see Figure 1). These images were used in the identification line-up tests. Crucially, the headscarves that were worn in these line-up images were very different to those that were worn in the videos in order to avoid pictorial recognition.

For constructing the line-up tests, a total of 40 distractor images were taken from 20 Egyptian and 20 British female undergraduate students, who resembled the actresses in age, hairstyle, and colouring. Furthermore, the targets and distractors had similar ratings of distinctiveness according to a prior rating study. All of these images showed frontal faces with or without a headscarf and were taken on the same day, under the same lighting conditions, and using standard headscarves. For each target nationality, we constructed two versions of target-present and target-absent line-ups, depending on whether the targets were wearing a headscarf or not. Each Egyptian or British line-up consisted of 10 faces including the target (or a target replacement in the absent condition). These line-ups were converted to gray-scale, and each face sized approximately 5×7 cm.

Design

The study employed a 2 (nationalities: British versus Egyptian) $\times$ 4 (conditions: 'headscarf study/headscarf test', 'no headscarf study/no headscarf test', 'headscarf study/no

headscarf test', or 'no headscarf study/headscarf test') between-subjects design. Furthermore, for each condition, TP and TA line-up tests were separately used. TP line-ups involve three independent responses (correct identifications, foil identifications and incorrect rejections); whereas TA line-ups involved two responses (correct rejections and foil identifications).

Procedure

Identical testing procedures were followed in each country. Participants were tested in small groups (consisting of 10–15) in a university teaching room equipped with data projection facilities. Seating position of participants was designed so that each had a good view of the screen (on which the video was projected), and all participants were seated a little distance apart. They were instructed to watch the video for a later recognition test. Following a 10-minute filler task (completing a non-relevant questionnaire), each participant was handed a target-present or target-absent line-up, which was congruent or incongruent with the headscarf study condition. Participants who watched the targets with headscarves were congruently presented with the line-ups of faces with headscarves, and participants who watched the targets with their hair visible were congruently presented with line-ups of faces with their hair visible. In the incongruent conditions, watching the targets with headscarves was followed by line-ups of faces with their hair visible or vice versa. Participants had to make their responses using an answer sheet containing one of the line-ups. They were told that the young woman from the video might or might not be present among these faces. Their task was to decide whether or not she was present, and if present, to identify her by recording the appropriate image number in their response sheets. No time pressure was applied for their responses.

Results

In the first step, hierarchical log linear analyses (HILOGs) showed no effect for participants' sex on TP and TA responses in all conditions (all $\chi^2 < 1$). Therefore, data from female and male participants were collapsed into new HILOGs, which were conducted separately on TP and TA line-ups.

Figure 2 shows the percentages of responses for TP line-ups in all experimental conditions (frequencies in parentheses). These data were subjected to a 2 (nationalities) $\times$ 4 (conditions) $\times$ 3 (responses) HILOG analysis, which yielded a three-way interaction [$\chi^2(6) = 30.453$, $p < .001$]. The British and Egyptian data were then subjected separately to a series of chi-square tests in order to follow up this interaction. For the British data, there was an effect for conditions and responses [$\chi^2(6) = 26.39$, $p < .001$]. British participants made more correct IDs in the 'no headscarf study/no headscarf test' as compared with the three remaining conditions: 'headscarf study/headscarf test' [$\chi^2(1) = 9.76$, $p = .002$], 'no headscarf study/headscarf test' [$\chi^2(1) = 10.42$, $p = .001$] and 'headscarf study/no headscarf test' [$\chi^2(1) = 9.8$, $p = .002$]. The Egyptian data also showed an effect for conditions and responses [$\chi^2(6) = 13.92$, $p = .03$]. There were more correct responses for the 'headscarf study/headscarf test', as compared with the 'no headscarf study/no headscarf test' [$\chi^2(1) = 6.56$, $p = .01$] and 'no headscarf study/headscarf test' [$\chi^2(1) = 4.84$, $p = .03$]. However, there was no difference between the 'headscarf study/headscarf test' and 'headscarf study/no headscarf test' [$\chi^2(1) = .87$, $p = .35$]. No other effect was found.

Figure 3 shows the percentages of responses on TA line-ups in all experimental conditions (frequencies in parentheses). $2 \times 4 \times 2$ HILOG analysis yielded a three-way interaction [$\chi^2(3) = 10.47$, $p = .01$]. Once again, British and Egyptian data were analysed separately. There was an effect for conditions and responses for the Egyptian data ($\chi^2(3) = 12.45$, $p = .006$) but not for the British [$\chi^2(3) = 3.02$, $p = .3$]. Subsequent χ^2 tests revealed that the headscarf study/headscarf test received more correct rejections than the 'headscarf study/no headscarf test' [$\chi^2(1) = 5.71$, $p = .02$], but it was similar to the 'no headscarf study/no headscarf test' [$\chi^2(1) = 3.02$, $p = .08$] and the 'no headscarf study/headscarf test' [$\chi^2(1) = .52$, $p = .47$]. This latter condition also received more correct rejections than the 'no headscarf study/no headscarf test' [$\chi^2(1) = 5.94$, $p = .015$] and the 'headscarf study/no headscarf test' [$\chi^2(1) = 5.94$, $p = .002$].

Finally, the data for TP and TA line-ups were collapsed to investigate the overall choosing behaviour. The choosing decision was calculated by combining all choice frequencies, regardless of whether they were correct or incorrect. On the other hand, the non-choosing decision involved all non-choice frequencies, regardless of whether they were correct or incorrect rejections. A 2 (nationalities) $\times$ 4 (conditions) $\times$ 2 (choosing) HILOG analysis yielded a three-way interaction [$\chi^2(3) = 20.13$, $p < .001$]. Similar to TP and TA analyses, follow-up chi-square tests were conducted separately for British and Egyptian participants. There was an effect for conditions and choosing for the Egyptian data ($\chi^2(3) = 16.26$, $p = .001$) but not for the British data ($\chi^2(3) = 7.14$, $p = .07$). Subsequent χ^2 tests showed that Egyptian participants were more likely to choose on 'the headscarf study/headscarf test' (56%) than on the 'no headscarf study/no headscarf test' (41%) [$\chi^2(1) = 5.64$, $p = .02$] and the 'no headscarf study/headscarf test' (34%) [$\chi^2(1) = 11.78$, $p = .001$]. Furthermore, they were more likely to choose on 'the headscarf study/no headscarf test' (54%) than the 'no headscarf study/headscarf test' [$\chi^2(1) = 9.73$, $p = .002$] and 'no headscarf study/no headscarf test' [$\chi^2(1) = 4.28$, $p = .04$]. No other effects were found.

Discussion

This experiment showed an interesting contrast between the British and Egyptian data. British participants were more accurate when their own-race target was not wearing a headscarf, whereas Egyptian observers were more accurate when their own-race target was wearing a headscarf. This contrast might provide support for what so called the headscarf effect (Megreya & Bindemann, 2009). Importantly, however, this contrast could be artificial for some characteristics of faces that were used in this experiment. Indeed, if the British faces have more distinctive hair styles and colours than the Egyptian faces, then a '*hair variance effect*' rather than a '*headscarf effect*' might be underlying these data. Therefore, if this hypothesis turns to be correct,

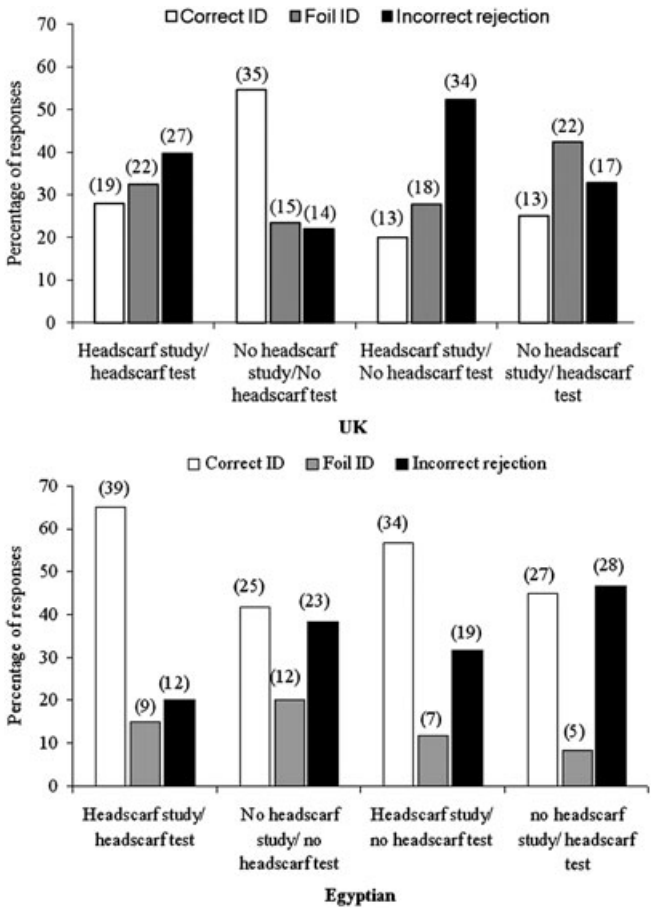

Figure 2. The percentage of responses for target-present line-ups for the UK and Egyptian participants (frequencies are in parentheses)

then we expect that Egyptian participants would be more accurate in identifying British faces when they appear without headscarves. Experiment 2 aimed to examine this hypothesis using the same British target from Experiment 1. British participants were more accurate at recognising the target when her hair was visible, than when covered with a headscarf.

EXPERIMENT 2

Method

Participants

Two-hundred undergraduate students (37 men) from Menoufia University participated in this experiment, with an age mean of 21.2 ($SD = 0.6$). No one had participated in Experiment 1, and none had any real-life interactions with Caucasians.

Stimuli and procedure

The British materials (two videos of staged thefts committed by a woman with or without a headscarf and their corresponding TP and TA line-ups) from experiment 1 were used in this experiment. The procedures were identical to those of Experiment 1, with the sole exception that, here, only congruent line-ups were used. In short, participants (in small groups) were shown the perpetrator when she was wearing a headscarf or left her hair uncovered. Following a 10-minute filler task, they were asked to identify the perpetrator from a congruent TP or TA line-up.

Results and discussion

Figure 4 shows the percentages of responses on TP and TA line-ups in headscarf conditions (frequencies in parentheses). The TP data were subjected to a 2×3 chi-square test that showed an effect [$\chi^2(1) = 7.79$, $p = .005$]. In a next step, the sub-measures of 'Foil ID' and 'Incorrect Rejection' were combined and compared directly with the Correct ID. This new 2×2 chi-square test yielded an effect [$\chi^2(1) = 7.43$, $p = .006$], with advantage for headscarf condition. These results provide good support for a '*headscarf effect*' rather than a '*hair variance effect*'. Consistently, Megreya and Bindemann (2009) found that Egyptian observers matched both Egyptian and British faces more accurately from their internal features, whereas British participants matched both types of faces more accurately from their external features. However, there was no evidence for this headscarf effect using the TA data [$\chi^2(1) = .16$, $p = .69$]. Also, there was no effect for choosing behaviour [$\chi^2(1) = 1.28$, $p = .26$].

GENERAL DISCUSSION

We report two experiments investigating what is termed as 'the headscarf effect' (Megreya & Bindemann, 2009) using the eyewitness identification paradigm. In Experiment 1, we examined the ability of observers from Egypt (where

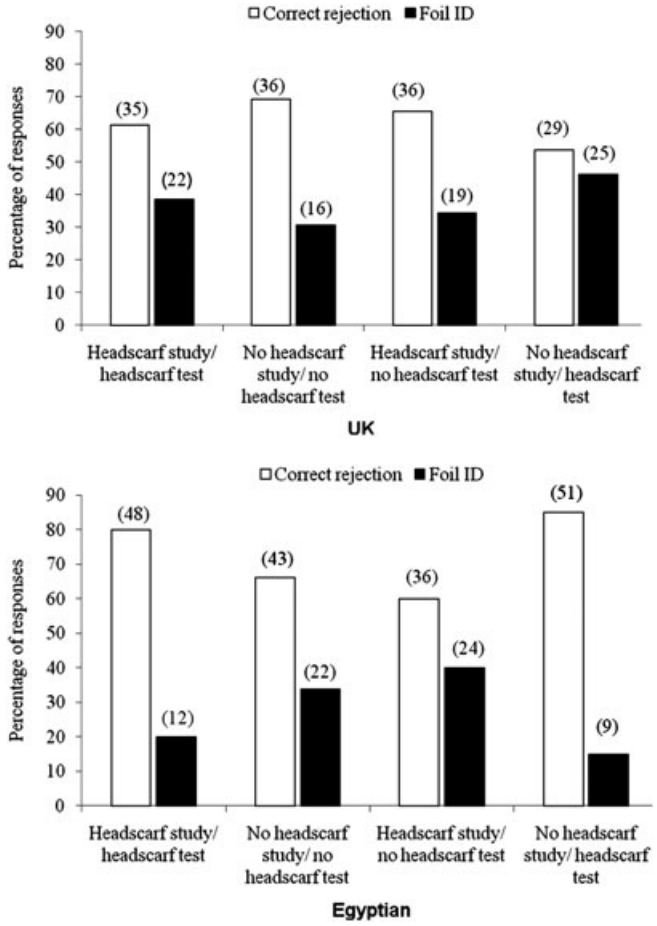

Figure 3. The percentage of responses for target-absent line-ups for the UK and Egyptian participants

the vast majority of women wear headscarves) and the UK (where the wearing of a headscarf is rare, especially among Caucasians) to identify an own-race female target who was previously seen in a crime video tape with or without a headscarf. TP/TA and congruent/incongruent line-ups were used. In congruent conditions, each group of participants initially saw an own-race target with or without a headscarf and was then presented with a line-up of women with or without a headscarf, respectively. In incongruent conditions, half of the participants initially saw a target with a headscarf and were then presented with a line-up of women without headscarves, whereas the other half initially saw the target without a headscarf and were then presented with a line-up of women with headscarves. Although these incongruent conditions have less ecological validity than the congruent conditions, they were critical for determining whether the headscarf effect was occurring at the encoding or retrieval phase and providing important insights for any possible effects for the congruent conditions. Experiment 2 examined how well Egyptian eyewitnesses could identify the British woman, who participated as a target in Experiment 1, when she appeared with or without a headscarf.

British participants made more correct IDs when their own-race target did not wear a headscarf during the study and test (54.7%), as compared to all other conditions: wearing a headscarf at both the study and test (27.9%) or only at the study (20%) or test (25%). This finding confirms the importance of external facial features for unfamiliar face recognition (Bonner et al., 2003; Bruce et al., 1999; Frowd et al., 2007; Nachson & Shechory, 2002; Want et al., 2003). In stark contrast, Egyptian participants made more correct IDs when their own-race target was wearing a headscarf at study and test (65%) than when she did not wear it at both the study and test (41.7%) or only at the test (45%). However, recognition at this congruent headscarf condition was similar to the other incongruent condition, where the target wore a headscarf only at the study (58.6%). In addition, British participants were more accurate than Egyptian observers in the congruent no headscarf condition (55% vs 42%), whereas Egyptian participants were roughly two times greater than those of British observers in the congruent headscarf condition (65% vs 28%). Importantly, this Egyptian headscarf advantage was noticed again in Experiment 2, where Egyptian participants had to identify the same British target from Experiment 1. This replication suggests that the Egyptian headscarf effect reflects a specific characteristic of participant nationality rather than face nationality.

Megreya and Bindemann (2009) consistently found that Egyptian adults matched both Egyptian *and* British unfamiliar faces more accurately from internal than external features, whereas British counterparts matched the same faces more accurately from external than internal features. Although Megreya and Bindemann provided strong evidence for the existence of this Egyptian internal-feature advantage for processing unfamiliar faces, their data offered little regarding the aetiology of this effect. Rather, they

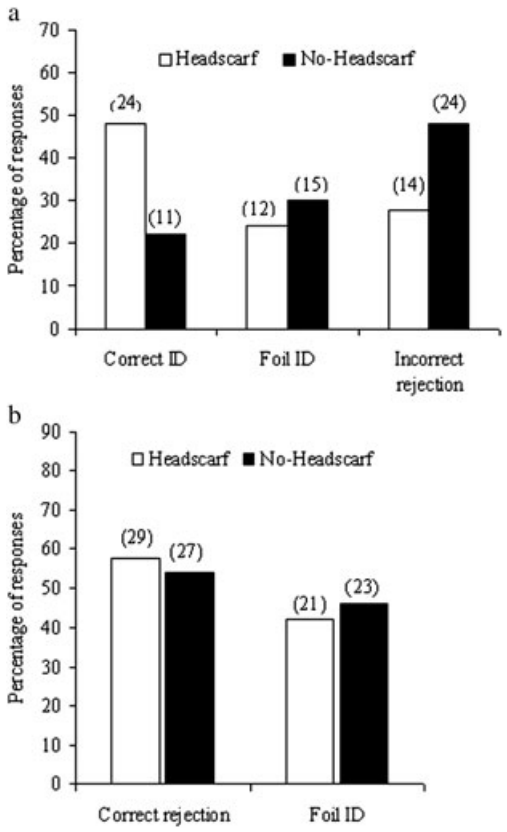

Figure 4. (a) The response percentages for the TP line-ups in Experiment 2 (frequencies in parentheses), (b) the response percentages for the TA line-ups in Experiment 2 (frequencies in parentheses)

speculatively attributed this Egyptian bias the long-term experiences of Egyptian adults at recognising female faces with headscarves that might inherently learn them to pay closer attention to the internal facial features. Our present results provide empirical support for this explanation as Egyptian adults recognise unfamiliar female faces, regardless of their nationality, more accurately when they wear headscarves.

The comparison between congruent and incongruent headscarf conditions further provides important information about the locus of this Egyptian headscarf effect. Interestingly, the correct ID rates for Egyptian participants in the congruent headscarf condition were higher than the incongruent 'no headscarf study/headscarf test' but were similar to the incongruent 'headscarf study/no headscarf test'. In other words, Egyptian participants performed less accurately when the hair was visible during encoding, but viewing the hair during the identification phase, did not seem to reduce correct identifications. Our explanation for this finding is as follows. Studying female faces with their hair visible might increase the perceptual load over the processing of internal features, and therefore it could be difficult to build up an accurate representation of the face. On the other hand, studying female faces with headscarves might offer Egyptian participants a good opportunity for encoding the internal features properly that persists even when the hair is visible during the identification test. However, this explanation would benefit from further investigation using eye-tracking procedures.

The TA data also revealed some interesting findings. The addition of a headscarf had no effect on correct rejections for British participants. Egyptian observers also showed similar correct rejection rates but *only* for congruent conditions in Experiments 1 and 2. However, they made more correct rejections in the 'headscarf study/headscarf test' (80%) as compared with the 'headscarf study/no headscarf test' (60%). In addition, correct rejection rates were higher in the 'no headscarf study/headscarf test' (85%) than in the 'no headscarf study/no headscarf test' (66%) and the 'headscarf study/no headscarf test'. Therefore, Egyptian participants were more likely to make a correct rejection when faces were wearing headscarves in the line-up, and this was regardless of whether the target was seen with or without a headscarf during the study phase. This seems to suggest that seeing the hair at retrieval can impair correct rejections for Egyptian participants. Alternatively, perhaps seeing a headscarf in the line-up makes Egyptian participants more cautious in their identification decisions and less likely to choose someone. Indeed, some data from choosing decisions favour this second possibility. Choosing rates on the 'headscarf study/no headscarf test' was higher than those in the 'no headscarf study/headscarf test'. However, other patterns of choosing behaviour did elicit the Egyptian headscarf effect especially in congruent conditions. Egyptian participants were more likely to choose on the 'headscarf study/headscarf test' than on the 'no headscarf study/no headscarf test'.

Comparing the performance on TP and TA line-ups, it seems clear that the headscarf effect can have disparate influences. For British witnesses, the wearing of a headscarf only impaired correct identification of a target from a TP line-up but had no influence over decisions on TA line-ups. On the other hand, Egyptian participants showed the headscarf effect for both types of line-ups, but it manifested during different processes. In TP line-ups, wearing a headscarf appeared to influence the encoding of target face. There was no reduction in identification accuracy when the target had been encoded wearing a headscarf and the line-up test contained faces with their hair visible. In TA line-ups, however, the headscarf effect did elicit during the retrieval phase. Correct rejections were greater when the line-up faces wore headscarves, regardless of whether the target had been encoded wearing a headscarf or not. An interesting avenue for future research would be to examine how Egyptians fare relative to British when the hair changes radically.

The Egyptian headscarf effect supports the power of expertise for face processing (e.g. for reviews see Bukach et al., 2006; Gauthier & Nelson, 2001). At a broader perspective, it also provides strong evidence for the effects of culture on cognition (e.g. for a review see Nisbett & Norenzayan, 2002). For example, it is now well documented that basic cognitive and motivational processes vary dramatically across nations, and therefore the fact that the vast majority of studies, which were conducted using Western participants (mostly Americans), is now believed to present a challenge to the understanding of human psychology and behaviour (Arnett, 2008; Henrich, Heine, & Norenzayan, 2010a, 2010b). The present study adds to this literature the suggestion that face-processing system is adaptive and could be greatly influenced by the cultural contexts.

In addition to this theoretical significance, the present study also provides important information for forensic settings. Eyewitness identification is now well-known to be highly error-prone (for reviews see Cutler & Penrod, 1995; Memon, Vrij, & Bull, 2003; Steblay, Dysart, Fulero, & Lindsay, 2001; Wells, 1993), and may cause unintentionally wrongful imprisonment (e.g. The Innocence Project). Therefore, a great deal of research has examined the sources of this fallibility (e.g. for a review see Wells, Memon, & Penrod, 2006). For example, Megreya and Burton (2008) suggested that a significant part of the difficulty of eyewitness identification involves problems of encoding unfamiliar faces in the first place. Previous research reported that the hair is the most salient area in the human face (e.g. O'Donnell & Bruce, 2001). However, the hair is not a permanent facial feature, and it can be easily changed, or concealed (for example by a hat, which is common in Western nations or by a headscarf which is common in Middle Eastern societies). The perception of unfamiliar faces was found to be greatly impaired by changing hair styles (Kemp, Towell, & Pike, 1997) or wearing a hat (Henderson, Bruce, & Burton, 2001). Therefore, the present study suggests that training Westerners to process faces in a fashion similar to Egyptians, might improve their identification accuracy. In support of this suggestion, research has shown that the recognition of familiar faces is robust (Burton, Jenkins, Hancock, & White, 2005; Burton, Wilson, Cowan, & Bruce, 1999) and it relies on the processing of internal features (Bonner et al., 2003; Campbell et al., 1999; Clutterbuck & Johnston, 2002, 2004, 2005; Ellis et al., 1979; Osborne & Stevenage, 2008; Young et al., 1985). In fact, the relative superiority of Egyptians over British in encoding (Megreya & Bindemann, 2009) *and* remembering (the present study) unfamiliar faces from the internal features hints that this training could be working, as long as there is not a trade off with encoding faces with external features visible.

ACKNOWLEDGEMENTS

We would like to thank Colin Gray and Mike Burton for their great statistical advice on this study. We thank also Graham Davies (Editor), Peter Hancock and three other anonymous reviewers for great suggestions on early drafts of this paper. Also, we thank Nicola Kent for her help in preparing stimuli and collecting data on the British part of this study. Finally, we thank Hayley Dickson and Nancy George who acted in the British and Egyptian videos, respectively as targets and all of those who acted as foils in the line-ups.

REFERENCES

Arnett, J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, *63*, 602–614.

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, *25*, 281–316.

Bonner, L., Burton, A. M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, *10*, 527–536.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology Applied*, *5*, 339–360.

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trend in Cognitive Sciences*, *10*, 159–166.

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, *51*, 256–284.

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.

Campbell, R., Coleman, M., Walker, J., Benson, P. J., Wallace, S., Michelotti, J., & Baron-Cohen, S. (1999). When does the inner-face advantage in familiar face recognition arise and why? *Visual Cognition*, *6*, 197–216.

Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*, 985–994.

Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition*, *11*, 857–869.

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology*, *17*, 97–116.

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. New York: Cambridge University Press.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology. General*, *115*, 107–117.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, *8*, 431–439.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, *29*, 159–170.

Frowd, C., Bruce, V., McIntyre, A., & Hancock, P. (2007). The relative importance of external and internal features of facial composites. *British Journal of Psychology*, *98*, 61–77.

Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Current Opinion in Neurobiology*, *11*, 219–224.

Gauthier, I., & Tarr, M. J. (1997). Becoming a 'Greeble' expert: Exploring mechanisms for face recognition. *Vision Research*, *37*, 1673–1682.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*, 191–197.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neutral system for face perception. *Trends in Cognitive Sciences*, *4*, 223–233.

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, *15*, 445–464.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not WEIRD. *Nature*, *466*, 29.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*, 61–83.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology*, *11*, 211–222.

McKone, E., Kanwisher, N., Duchaine, B. C. (2006). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, *11*, 8–15.

Megreya, A. M., & Bindemann, M. (2009). Revisiting the processing of internal and external features of unfamiliar faces: The headscarf effect. *Perception*, *38*, 1831–1848.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*, 865 – 876.

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity based dissociation. *Perception & Psychophysics*, *69*, 1175–1184.

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology Applied*, *14*, 364–372.

Memon, A., Vrij, A., & Bull, R. (2003). *Psychology and law: Truthfulness, accuracy and credibility*, 2nd edn. England: John Wiley & Sons Ltd.

Nachson, I., & Shechory, M. (2002). Effect of inversion on the recognition of external and internal facial features. *Acta Psychologica*, *109*, 227–238.

Nisbett, R. E., & Norenzayan, A. (2002). Culture and cognition. In D. Medin, & H. Pashler (Eds.), *Stevens' handbook of experimental psychology*, 3rd edn, Vol. *2*, Memory and Cognitive Processes. New York: John Wiley & Sons.

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception*, *30*, 755–764.

Osborne, C. D., & Stevenage, S. V. (2008). Internal feature saliency as marker of familiarity and configural processing. *Visual Cognition*, *16*, 23–43.

Rhodes, G., & McLean, I. G. (1990). Distinctiveness and expertise effects with homogeneous stimuli: Towards a model of configural coding. *Perception*, *19*, 773–794.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior*, *25*, 459–473.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, *25*, 583–592.

Want, S. C., Pascalis, O., Coleman, M., & Blades, M. (2003). Recognisizing people from the inner or outer parts of their faces: Developmental data concerning 'unfamiliar' faces. *British Journal of Psychology*, *21*, 125.

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553–571.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, *7*, 45–75.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, *14*, 737–746.