Disguising Superman: How Glasses Affect Unfamiliar Face Matching

ROBIN S. S. KRAMER* and KAY L. RITCHIE

Department of Psychology, University of York, York, UK

Summary: Could a simple pair of glasses really fool us into thinking Superman and Clark Kent are two different people? Here, we investigated the perception of identity from face images with a task that relies on visual comparison rather than memory. Participants were presented with two images simultaneously and were asked whether the images depicted the same person or two different people. The image pairs showed neither image with glasses, both images with glasses, and 'mixed' pairs of one image with and one without glasses. Participants' accuracies, measured by both percentage correct and d' sensitivity, were significantly lower for 'mixed' trials. Analysis of response bias showed that when only one face wore glasses, people tended to respond 'different'. We demonstrate that glasses affect face matching ability using unconstrained images, and this has implications for both dis- guise research and authenticating identity in the real world. Copyright © 2016 John Wiley & Sons,Ltd.

INTRODUCTION

In the fictional city of Metropolis, the character of Clark Kent lives the life of an unassuming journalist with (for the most part) little worry that those around him will make the association with his alter ego, Superman. Although Kent makes several alterations to his appearance in order to hide his true identity, perhaps his most famous prop is a simple pair of glasses. While we are generally required to suspend our disbelief with such stories, it may actually be of practical interest to consider just how successful donning glasses can be when trying to disguise oneself.

To date, research in this area has tended to investigate how glasses affect memory for faces (for a meta-analysis, see Shapiro & Penrod, 1986). Typically, participants are shown a set of face images and are then presented with test images, to which they are asked to respond 'old' (i.e. the face appeared in the first set) or 'new' (i.e. it had not previously been seen). Recognition accuracy was impaired when a face initially seen with glasses is shown without glasses in the recognition phase, and vice versa (Leder et al., 2011; Patterson & Baddeley, 1977; Righi et al., 2012; Terry, 1993, 1994). As one might expect, this detriment was also apparent when sunglasses were used (Hockley et al., 1999; Vokey & Hockley, 2012).

To simulate a courtroom scenario sometimes faced by juries, Davis and Valentine (2009) asked participants to watch staged CCTV footage of an actor while a 'defendant' was present in the room. Surprisingly, higher accuracy in matching the person in the room to the person in the video was found when the actor wore dark glasses in the footage in comparison with no disguise. Perhaps crucially, participants were able to compare the full bodies of the actors in the videos with the bodies of the live actors. As such, comparisons with prior work using facial photographs may be problematic. Using a similar design, other researchers showed video footage of actors (from the shoulders up) with or without sunglasses, and participants were asked to identify the actor in a set of undisguised photographs (Mansour et al., 2012). As expected, sunglasses significantly reduced identification accuracy.

Researchers typically suggest the eyes to be the most important internal feature for face recognition (Davies et al., 1977; Fraser et al., 1990; Haig, 1986). Therefore, any trans- formations to this region between learning and test are likely to reduce recognition performance. If glasses are present during the initial encoding of the face, then less information may be available for use during later recognition of the un- disguised face. However, participants may adapt to the task by emphasising facial features other than the eyes during learning (Hockley et al., 1999). Interestingly,

there are substantial individual differences in face recognition abilities (Bobak et al., 2016a; Robertson et al., 2016; White et al., 2015a; White et al., 2015b), and it may be that those who perform best on such tasks focus more on the central area of the face rather than relying as heavily on the eyes (Bobak et al., 2016b).

Research in this field has primarily focussed on how glasses affect learning and subsequent memory for faces. Surprisingly, there has been little consideration of glasses in a face matching context, where stimuli are presented simultaneously and participants decide whether images depict the same person or different people. Despite decreased memory requirements, people still perform poorly on this task (Bruce et al., 2001; Burton et al., 2010; Megreya & Burton, 2006, 2008).

The study by Davis and Valentine (2009), mentioned earlier, allowed participants to match a live person to a video they were watching simultaneously. Diamond and Carey (1977) asked children to match images with varying types of disguise, finding that the younger children placed too much weight on matching these artificial clothing additions, to their detriment. Finally, Leder and colleagues (Experiment 2; Leder et al., 2011) considered simultaneous matching, where at least one of the two faces was presented without glasses. Importantly, in their experiment, the 'glasses' and 'no glasses' photographs were virtually identical, passport- style images. Unsurprisingly, the task was easy for participants, and so the focus was on reaction times, with participants responding slower when one of the faces wore glasses. To our knowledge, these represent the only investigations of glasses and their effect on face matching.

Here, we ask whether glasses can affect accuracy in the matching of facial photographs. We utilise regular glasses rather than dark glasses or sunglasses, and importantly, we use unconstrained images. As mentioned earlier, matching passport-style photographs, where the only difference is the presence or absence of glasses, is trivial for participants (Leder et al., 2011). The use of images that incorporate real-world variability allows us to address how people may be affected by glasses in realistic viewing contexts above and beyond the inherent difficulties that come with matching unconstrained images. We hypothesise that, where glasses are only present in one of the images, performance will be reduced.

METHOD

Participants

Fifty-nine students (52 women; age $M = 20.78$ years, $SD = 3.35$; 93.22% self-reported ethnicity as White) at the University of York took part in exchange for course credits. All participants provided written informed consent and were verbally debriefed at the end of the experiment. Sample size was based on past research using a face matching paradigm (Dowsett & Burton, 2015; Estudillo & Bindemann, 2014), and our stopping point was set for the end of a pre- determined data collection run (based on the term timetable). The University of York's psychology department ethics committee approved this study, which was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Stimuli

For 48 White identities (half women), four different facial photographs were downloaded from Google Images. Two of these depicted the person wearing glasses and two showed them without. Care was taken to make sure that the four images depicted different situations (i.e. they were not taken minutes apart and so did not share clothing, backgrounds, etc.). This allowed us to avoid confounds that may otherwise have aided matching within conditions but decreased performance between conditions (e.g. if both 'glasses' images also shared hair style or head angle, and these differed from the characteristics of the 'no glasses' images). As such, there were no systematic differences (other than the presence or absence of glasses) between image conditions for a given identity.

In addition, two photographs of 'foils' (other people who resembled the identities) were downloaded for each person using descriptive search terms that matched their general appearance (e.g. 'blonde woman' and 'grey haired man'). The foil images for each identity were of two different people, and one of these images depicted a person wearing glasses. All images were high quality, colour, unconstrained, naturalistic photographs, and cropped loosely around the person's head, as in Figure 1. Images were resized to 190×285 pixels and were approximately 9×13.5 cm onscreen. Identities were chosen in order that they would be unfamiliar to our predominantly British students.

Figure 1. Images illustrating the three trial types for an example identity. These were as follows: both images contain glasses (top row); neither image contains glasses (middle row); and one image contains glasses and one does not (bottom row). Here, the correct responses are 'different' (top and middle rows) and 'same' (bottom row). Note that images appeared in colour during the experiment. (Copyright restrictions prevent publication of the original images. Images shown here feature three women who did not appear in the experiment. All have given permission for their images to be reproduced here).

The set included international news anchors, singers, models, athletes, and so on. However, after piloting the study on a separate group of nine participants (whose data were discarded), four identities were replaced because they were frequently recognised.

Procedure

On each of the 48 trials, two images were presented onscreen, one to the left and one to the right of centre. Viewing distance was not fixed. The task was to judge whether the two images were of the same person or two different people.

Participants responded using the keyboard, pressing 'A' for 'same' and 'L' for 'different'. These

labels remained onscreen throughout the experiment. Trials were self-paced. Each identity appeared only once for each participant. Trials were divided equally into three conditions: (i) 'glasses', both images contained glasses; (ii) 'no glasses', neither image contained glasses; and (iii) 'mixed', one image contained glasses and one did not. For each condition, half the trials showed two images of one of our identities ('match'), and half showed one image and one foil ('mismatch'). In these mismatch trials, where only one of the images of a given identity was utilised, the particular image shown was selected at random (e.g. from the two glasses-wearing images for that identity).

Identities were randomly assigned to each trial type for each participant with the proviso that half the identities in each of these six trial types were women and half were men. Trial order was randomised for each participant.

After completing the experiment, participants were asked if they had recognised any of the people in the photographs. If they had, they were given a list of the identities' names and were instructed to indicate those with which they were familiar.

RESULTS

The data from seven participants were excluded from our analyses because they reported recognising more than five identities. For the remaining 52 participants, the average number of identities recognised was 1.67 ($SD = 1.63$).

Overall percentage correct was analysed using a one-way (Condition: glasses, no glasses, mixed) within-subjects analysis of variance (ANOVA). We found a significant main effect, $F(2, 102) = 7.95$, $p < .001$, $\eta^2 = 0.13$. Paired t-tests indicated that percentage correct for the 'mixed' condition was significantly lower than for both the 'glasses', $t(51) = 3.08$, $p = .003$, $d = 0.43$, and 'no glasses' conditions, $t(51) = 3.72$, $p < .001$, $d = 0.52$ (Table 1). Both differences remain significant after applying Bonferroni correction ($\alpha = 0.05/3$) for multiple comparisons. The 'glasses' and 'no glasses' conditions did not significantly differ from each other, $t(51) = 0.72$, $p = .473$, $d = 0.10$.

As with other research in this field (e.g. Terry, 1994), we investigated signal detection measures in order to take into account potentially separate effects caused by the three conditions on 'same' and 'different' trial types. We calculated sensitivity indices (d') using the following: *Hit*, both images are of the same identity and participants responded 'same'; and *False alarm*, the two images are of different people and participants responded 'same'. Trials from the three glasses conditions were considered separately.

These d' values were analysed using a one-way within- subjects ANOVA producing a significant main effect, $F(2,102) = 7.66$, $p < .001$, $\eta^2 = 0.13$. Paired t-tests indicated that d' for the 'mixed' condition was significantly lower than for both the 'glasses', $t(51) = 3.06$, $p = .004$, $d = 0.42$, and 'no glasses' conditions, $t(51) = 3.68$, $p < .001$, $d = 0.51$ (Table 1). Both differences remain significant after applying Bonferroni correction for multiple comparisons. The 'glasses' and 'no glasses' conditions did not significantly differ from each other, $t(51) = 0.70$, $p = .489$, $d = 0.10$. These results show the same patterns as the analysis of accuracy (percentage correct) presented earlier.

Finally, similar analyses were carried out for participants' criterion values for each condition. A one-way within- subjects ANOVA produced a significant main effect, $F(2, 102) = 39.24$, $p < .001$, $\eta^2 = 0.43$. Paired t-tests indicated that criterion for the 'glasses', 'no glasses', and 'mixed' conditions all significantly differed from each other (all ps < .001). Comparing criterion values with zero ('unbiased') showed that in the 'glasses' condition, participants were biased towards responding 'same', $t(51) = 2.12$, $p = .039$, $d = 0.29$, although this was no longer significant after Bonferroni correction for multiple comparisons. In the 'no glasses' condition, participants produced an unbiased response, $t(51) = 1.48$, $p = .144$, $d = 0.21$. In the 'mixed' condition, participants showed a bias towards responding 'different', $t(51) = 6.80$, $p < .001$, $d = 0.94$.

Table 1. Summary of the results for the three conditions

Condition	Overall correct (%)	Sensitivity (d′)	Criterion
Glasses	79.6 [76.7, 82.4]	1.77 [1.59, 1.94]	−0.13 [−0.26, −0.01]
No glasses	80.9 [77.7, 84.1]	1.85 [1.65, 2.05]	0.09 [−0.03, 0.21]
Mixed	74.0 [70.9, 77.2]	1.43 [1.23, 1.62]	0.38 [0.27, 0.49]

Note: Glasses, both images contained glasses; no glasses, neither image contained glasses; mixed, one image contained glasses and one did not; 95% confidence intervals are given in square brackets.

DISCUSSION

We investigated how well people were able to determine whether two images, presented simultaneously and with no time restrictions, depicted the same person or not. Participants performed equally accurately when both images did or did not include glasses. However, when only one of the two faces wore glasses, performance was significantly reduced.

Our 'no glasses' condition is equivalent to previous re- search on face matching. Here, overall accuracy in this condition was 80.9%, which is lower than performance levels found with the benchmark Glasgow Face Matching Test (89.9%; Burton et al., 2010) but comparable with the shortened version of that test (81.3%), comprising only the most difficult items. That the current task was difficult for participants is perhaps unsurprising, given that our images were un- constrained in terms of expression, pose, and so on, which was not the case with the Glasgow stimuli.

Intuitively, any change to the face across two images should decrease matching performance. The Glasgow Face Matching Test (Burton et al., 2010) focussed specifically on alterations to the camera parameters, utilising two different static cameras and a video camera in order to capture different images of the same identity. In addition, a change in viewpoint across the two images to be matched produces an additional (although small) decrease in accuracy (Estudillo & Bindemann, 2014). Here, the two images of each face were unconstrained and so could differ dramatically (Jenkins et al., 2011). It is a testament to the effect of glasses that we find a significant decrease due to this specific alteration above and beyond the numerous differences in the pairs of images already present (Figure 1).

Accuracy in the 'glasses' and 'no glasses' conditions did not differ. We might predict that participants may perform worse on the 'glasses' trials because less information about each face is available for matching (owing to occlusion by glasses frames). However, perhaps because we used regular glasses rather than sunglasses or dark glasses, little information was lost and performance was unaffected. Conversely, one could predict higher matching accuracy in the 'glasses' condition because extra information about the glasses them- selves was also available for use in matching. For our particular identities, only 25 of 48 wore the same pair of glasses in their two 'glasses' images. If we consider only match trials in the 'glasses' condition, where the two pairs of glasses were identical for some trials but not others, we do indeed find that participants were more accurate when identical (90.2%) versus different glasses (79.2%) appeared, $t(51) = 3.40$, $p = .001$, $d = 0.47$. This result suggests that participants may have compared the glasses themselves to improve accuracy on these types of trials. Of course, this also means that accuracy might have decreased when the glasses were different (but the identities were the same).

The decrease in performance for 'mixed' trials shows that the appearance of glasses in only one image makes matching harder. Participants responded conservatively in these trials by more often saying that images were of two different people. One explanation might simply be that a change in glasses category (wearing vs not wearing) could be similar to a change in hair colour across images—'the two photos depict different hair colours so it is more likely to be two different people'. If this 'rule of thumb' is something that people subscribe to, then glasses represent a simple disguise that could be effective in real-world scenarios. Conversely, one can look more similar to a photograph in a passport

(which may not be your own) by simply wearing or removing a pair of glasses.

Although our untrained participants performed significantly worse when glasses were added or removed, it may be that more experienced forensic examiners or those with higher natural abilities (e.g. 'super recognisers') remain un- affected by such alterations. Evidence suggests that people who perform better on these tasks may spend more time examining the central region of the face (i.e. the nose rather than the eyes; Bobak et al., 2016b), which could minimise the role that glasses may play.

In conclusion, we show that the addition or removal of glasses can affect face matching performance, where images appear simultaneously and participants are not required to remember faces for later comparison. Specifically, when one face is wearing glasses and one is not, matching ability is significantly worse. While we may still find it hard to believe that the inhabitants of Metropolis are unable to match Clark Kent with the numerous appearances of Superman in newspapers and on television, we can at least understand why Kent has chosen glasses as his aid to anonymity.

REFERENCES

Bobak, A. K., Hancock, P. J. B., & Bate, S. (2016a). Super-recognisers in action: Evidence from face-matching and face memory tasks. *Applied Cognitive Psychology, 30*, 81–91.

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2016b). Eye-movement strategies in developmental prosopagnosia and "super" face recognition. *The Quarterly Journal of Experimental Psychology.* Advance online publication.

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied,* 7, 207–218.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, 42, 286–291.

Davies, G., Ellis, H., & Shepherd, J. (1977). *Cue saliency in faces as assessed by the "Photofit" technique. Perception, 6,* 263–269.

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology*, *23,* 482–505.

Diamond, R., & Carey, S. (1977). Developmental changes in the representation of faces. *Journal of Experimental Child Psychology*, *23*, 1–22.

Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British Journal of Psychology, 106,* 433–445.

Estudillo, A. J., & Bindemann, M. (2014). Generalization across view in face memory and face matching. *i-Perception*, *5,* 589–601.

Fraser, I. H., Craig, G. L., & Parker, D. M. (1990). Reaction time measures of feature saliency in schematic faces. *Perception*, *19,* 661–673.

Haig, N. D. (1986). Exploring recognition with interchanged facial features. *Perception*, 15, 235–247.

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

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same person. *Cognition, 121,* 313–323.

Leder, H., Forster, M., & Gerger, G. (2011). The glasses stereotype revisited: Effects of eyeglasses on perception, recognition, and impression of faces. *Swiss Journal of Psychology, 70,* 211–222.

Mansour, J. K., Beaudry, J. L., Bertrand, M. I., Kalmet, N., Melsom, E. I., & Lindsay, R. C. L. (2012). Impact of disguise on identification decisions and confidence with simultaneous and sequential lineups. *Law and Human Behavior, 36,* 513–526.

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

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. Journal of Experimental Psychology: *Human Learning and Memory, 3,* 406–417.

Righi, G., Peissig, J. J., & Tarr, M. J. (2012). Recognizing disguised faces. *Visual Cognition, 20,* 143–169.

Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., & Burton, A. M. (2016). Face recognition by Metropolitan Police super-recognisers. *PLoS ONE, 10(10),* e0150036.

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*, 139–156.

Terry, R. L. (1993). How wearing eyeglasses affects facial recognition. *Current Psychology, 12,* 151–162.

Terry, R. L. (1994). Effects of facial transformations on accuracy of recognition. *The Journal of Social Psychology, 134,* 483–492.

Vokey, J. R., & Hockley, W. E. (2012). Unmasking a shady mirror effect: Recognition of normal versus obscured faces. *The Quarterly Journal of Experimental Psychology, 65,* 739–759.

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015a). Error rates in users of automatic face recognition software. *PLoS ONE, 10(10),* e0139827.

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015b). Perceptual expertise in forensic facial image comparison. *Proceedings of the Royal Society B: Biological Sciences,* 282, 20151292.