The ‘double face’ illusion

Peter J.B. Hancock and Catherine Foster
Psychology, School of Natural Sciences
University of Stirling
FK9 4LA

Abstract

We report three experiments intended to characterise aspects of the ‘double’ face illusion, formed by replicating the eyes and mouth below the originals. Such doubled faces are disturbing to look at. We find there are wide individual differences in ability to detect that a face has been doubled when presented briefly and masked. These differences appear to relate to perceptual speed, since they correlate with the ability to identify a briefly presented famous face. Doubling has a significant effect on identification, though much less than inversion. In a reaction time study, participants are faster to decide that a face has been doubled as it is rotated away from upright. The final study shows that normal and doubled faces do not pop out from each other, but reveals a processing overhead of 40-60ms per doubled face. We offer some speculations as to the cause of the perceptual effects.

1 Introduction

What we shall call the ‘double face’ illusion has been in evidence on the World Wide Web for a few years, but to date has received little formal attention (Kitaoka 2008). It is interesting, not least because it is so disturbing to look at. As is apparent from Figure 1, inverting the face does not make the damage inconspicuous, unlike the Thatcher illusion (Thompson 1980). This paper reports three studies investigating the effects it has on perception.

Figure 1. The 'double face' illusion

The obvious analogy, especially to readers of Perception, is indeed the Thatcher illusion. This involves cutting out and inverting the eye and mouth regions, producing an image that is grotesque when viewed upright, but inoffensive when inverted. The cause of this is generally held to be that inverting the eyes and mouth affects the configuration of the face and that inversion impairs our ability to detect this configuration. What effect does doubling have? In some respects, like the Thatcher, it does not affect the main face features; it simply replicates two of them. It clearly affects the configuration, in the sense of the relationship between the features, because it is no longer clear which set of features to relate. However, inverting a doubled face does not prevent perception of the alteration, where it does obscure changes in configuration caused by simply moving, say, the mouth up or down (Searcy and Bartlett 1996).

Here, we wish to begin the process of formally testing the effects this manipulation has on our perception of faces, for example what effect it has in a search task. Our first experiment considers the ability to detect that a face has been doubled, and the effect of doubling on recognition of familiar faces. Informal observation indicated that doubling will not prevent recognition, though it might well delay it. We reasoned that a reaction time to recognition study would not tell us very much, however, and that it would be more interesting to test the effects on recognition, by reducing presentation time to the point where errors are made (Dakin and Watt 2009). This provides a less ambiguous response: either the participant recognises the face or they do not, whereas reaction times carry with them the worry of a speed/error trade-off. In order to investigate the nature of the disruptive effect of doubling, we also tested recognition when inverted. If the disruption to recognition is mostly configural, then inversion might be expected to reduce the effects, resulting in an interaction between the two factors (Leder and Bruce 2000; Leder and Carbon 2006).

Earlier work indicated that participants often failed to notice that faces had been doubled in the brief presentation times used to test recognition. We therefore added a separate test of ability to detect whether faces had been doubled, using unfamiliar faces, presented both upright and inverted and at the same presentation times used in the recognition study. Since our earlier work also indicated large individual differences in detection ability, we further tested this using an adaptive algorithm that sought each participant's threshold display time.

2 Experiment 1.

2.1 Participants

Twenty-four students from the University of Stirling took part in return for course credit, 9 male, average age 20 years, with normal or corrected to normal vision.

2.2 Materials

For the detection task, 10 male and 10 female faces were taken from a set of images, gathered several years previously by the UK Home Office, and therefore unfamiliar to participants. The people depicted are young adults, mostly police trainees; a 'doubled' version was created using Adobe Photoshop. Copies were saved upright and inverted. Head width on screen was about 320 pixels, 11cm or 9° at the viewing distance of about 75cm.

For the recognition task, 5 different frontal pictures of 60 celebrities were collected from the internet, selected to be at least 200 pixels across the face. Four images of each were edited using the GIMP image editor (www.gimp.org) to double the eyes and mouth and appeared the same size on screen as those used for the detection task. The fifth image was used only to confirm recognition and was deliberately chosen to be rather different from the others in terms of pose and cropping, to reduce the likelihood of simple image priming of the main experiment. A mask image was created by taking a different face image of the same size and repeatedly sampling randomly sized rectangles and pasting them elsewhere in the frame. The result was a highly jumbled image, with face-like tones but no recognisable features, which was found to be an effective mask in practice.

The monitor used was a 19" Iiyama CRT running at 1024x768 pixels resolution; the detection and recognition tasks were presented at 75Hz; the adaptive detection task at 120Hz, to allow finer display time step sizes.

2.3 Design

Both detection and recognition tasks were within-participants, with independent variables of face type, normal or doubled; orientation, upright or inverted; and presentation time, 53 or 80 ms (4 or 6 frames at 75Hz). Dependent variable was proportion correct for both. The adaptive detection task estimated the presentation time required to get 72% correct.

2.4 Procedure

Participants were tested individually in a cubicle; the whole procedure took about 30 minutes. They first completed the detection experiment. They were seated approximately 75cm from the display, which was controlled by Eprime presentation software. They were explicitly told that accuracy was more important than speed of response and to take their time. After an instruction screen that illustrated the two types of face, with an identity not used in the main experiment, they were shown four practice trials, one for each face type and rotation, displayed for 107ms (7 frames at 75Hz) before being masked for 250ms. Screen background was mid-grey. A response screen appeared, prompting the participant to respond ‘s’ for a single (normal) face, ‘d’ for double. After they did so, there was a central fixation cross for 1 second before the next face. If the participant was clear about the procedure after the practice block, they proceeded to the main sequence of 64 trials, broken into two blocks with the opportunity to rest between them. Faces were displayed for either 53ms (4 frames at 75Hz) or 80ms (6 frames). There were 8 trials for each condition (2 face types x 2 duration x 2 orientation).

They then completed the adaptive detection task, which used the same image set. The program was written in Matlab, using the Psychophysics toolbox (Brainard 1997; Pelli 1997; Kleiner et al. 2007) with the Quest adaptive search procedure (Farell and Pelli 1999), using default parameters. The program selected normal or double, upright or inverted at random, for a presentation period selected by the adaptive algorithm, followed by the same mask, for 250ms. As before, response was by using the ‘s’ or ‘d’ keys; participants were reminded that accuracy was more important than speed. An initial practice run showed 10 trials starting at 167 ms; most got these all correct, such that the presentation time stepped down after each trial. The main run showed 40 trials, starting at an initial estimate of 83ms (10 frames at 120Hz). Quest works by establishing an initial Gaussian probability density function for the threshold, which was given a standard deviation of 5 frames (41.6ms). After each trial, the density function is updated by Bayes rule. The new maximum probability, rounded to the nearest whole number of frames, is used for the next trial.

They then completed the recognition task and were first shown the fifth picture of each celebrity in turn and asked to identify them verbally, with no time limit. An experimenter sat with them and accepted any unambiguous answer, e.g. Indiana Jones for Harrison Ford. The experimenter operated the computer keyboard, entering ‘y’ or ‘n’ for correct recognition or not. There then followed four blocks of recognition trials, using a maximum of 40 of those celebrities recognised in the first stage, selected at random; the participant was aware of this. Five participants recognised fewer than 40, the minimum being 34. After a fixation cross for 500ms, the celebrity image was presented for either 53 or 80 ms, followed by the mask image for 250ms, followed by the question, ‘who was that?’ Again the participant attempted to name the celebrity. A small number on the screen allowed the experimenter to confirm the correct identity from a crib sheet. The experimenter entered ‘y’ or ‘n’, and this initiated the next fixation cross. Across the four blocks, each participant saw different images of a given celebrity in four of the eight possible conditions; which image of which celebrities appeared in each condition were counterbalanced across participants. There was a break between each block. This and all other experiments were carried out in accordance with BPS ethical guidelines.

2.5 Results

2.5.1 Detection

The dependent variable is proportion correct: Figure 2 shows the average for each condition.

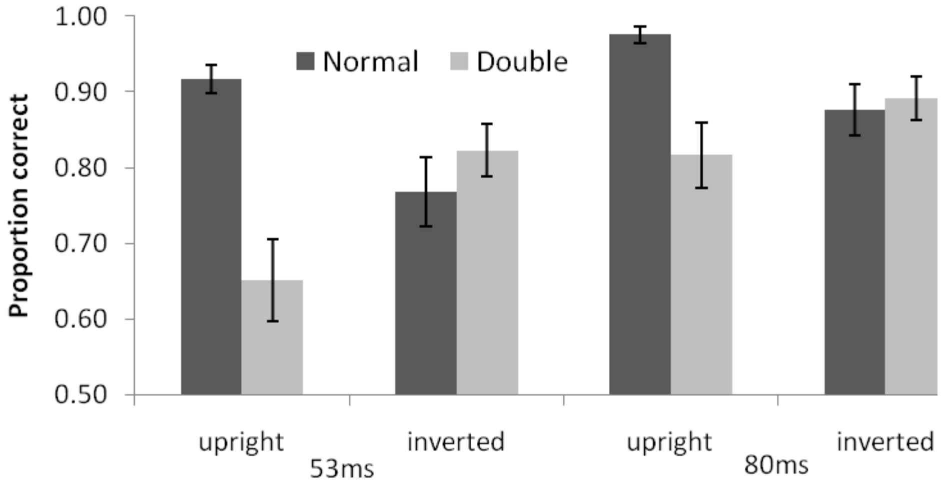

Figure 2 Detection proportion correct in each condition at 53ms (left) and 80ms (right)

The data were analysed using a 2 (rotation) x 2 (display time) x 2 (face type) repeated measures ANOVA. This showed a significant effect of display time ($F_{1,23}$=23.6, p<.001, η^2_p=.51). The overall proportion correct was higher at 80ms (M=.89) than at 53ms (M=.79). There was an effect of image type ($F_{1,23}$=8.94, p<.007, η^2_p=.28); normal faces (M=.89) were correctly identified more often than doubled (M=.79). There was also a significant interaction between orientation and face type ($\mathrm{F}_{1,23}$=10.4, p=.004, η^2_p=.31). As is evident from Figure 2, the proportion correct for the doubled faces rises as the face is inverted, whereas for normal faces it falls. The pattern is similar at 53 and 80 ms, though less pronounced at 80. A signal detection analysis, regarding double faces as hits and misidentified normal ones as false positives, confirms a significant effect of orientation on criterion, $F_{1,23}$ = 10.2, p=.004, η^2_p=.31 (with no significant effect of duration and no interaction). When the faces are upright, the average criterion is 0.35, falling to -0.08 when inverted. Participants are disinclined to respond ‘double’ when the faces are upright, but are more likely to do so when they are inverted.

The adaptive detection task returned estimates of display time threshold for 72% correct ranging from 35ms to 103ms, with a mean of 60ms. As expected, the threshold estimates showed a strong negative correlation with the overall proportion correct in the main detection task: r = -.65 at both 53ms and at 80ms, p<.001.

2.5.2 Recognition

Of the 60 celebrities shown, participants recognised on average 46.8 (78%) in the initial identification check. Correct identification rates for the main experiment, as a proportion of those identified initially, are shown in Figure 3.

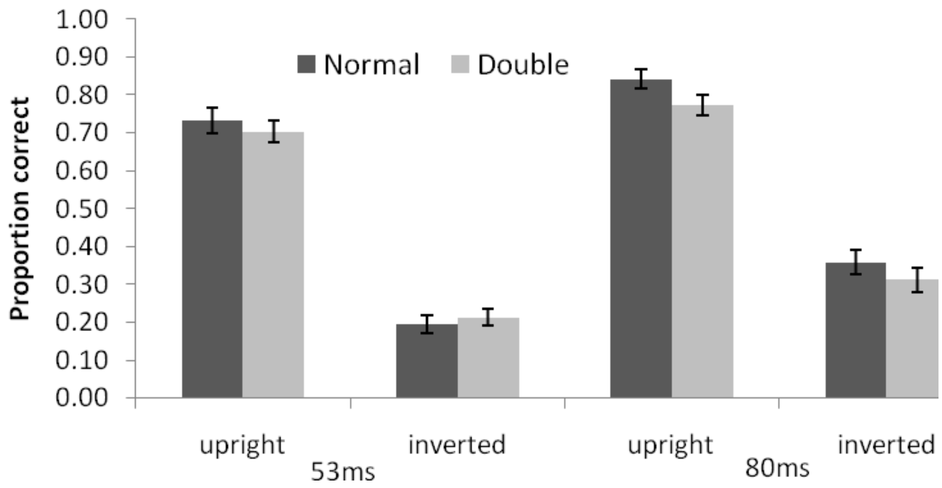

Figure 3 Proportion of correct identifications in each display condition, 53ms on the left, 80ms on the right. Error bars are standard errors.

A repeated measures ANOVA confirmed the expected large effect of inversion, $F_{1,23}$ = 925, p<.001, partial η^2=.98. There is an effect of image type, $F_{1,23}$=7.47, p=.012, η^2_p=.25 and of duration, $F_{1,23}$=112, p<.001, η^2_p=.83. Note that while the statistical effect size for image type is classed as large, Figure 3 shows the actual impact on recognition to be surprisingly small, certainly much less than inversion. There is also a significant interaction between image type and orientation, $F_{1,23}$=4.46, p=.046, partial η^2=.16. No other interactions are significant, all p>.15. The source of the image type x orientation interaction is that doubling the face has an even smaller effect on inverted images than it does on upright ones (or, equally, that inversion has a smaller effect on doubled images than it does on normal ones). Averaging across the durations, paired t-tests show an effect of doubling on upright faces, t_{23}=3.89, p=.001, effect size d = 0.39, but not inverted, t_{23}=0.22, p=.83, d = 0.03. Although there is no three way interaction, Figure 3 suggests that there might be an effect of doubling on inverted faces at 80ms so this was explicitly checked: it is not significant (t_{23}=1.63, p=.12). Overall, therefore, doubling does not have a significant effect on the recognition of inverted faces. It seems unlikely that this is a result of floor effects: the average identification rate for doubled faces, inverted, at 53ms is 0.21, which is 1.9 standard deviations above zero. It is therefore hard to argue that the doubled face score was artificially raised by having the bottom of its distribution cut off.

Previous data had suggested a few individuals were much more affected by doubling than the majority and it was thought possible that this was related to individual differences in ability to detect that faces were doubled. However; these data show no such variation in the effect of doubling on identification, with a very strong correlation between overall scores on normal and doubled faces, r = 0.85, p<.001. There is a mild correlation between the threshold for detection and the overall average identification performance r_{24} = -.41, p=.048; those who are better at detecting whether a face has been doubled at short presentations are also better at identification of rapidly presented faces. There is no hint of a correlation between detection threshold and the overall difference between recognition of normal and double faces, r=.03. The detection threshold did not correlate with this difference in any individual condition either, the largest being upright at 53ms, r = -0.27, ns. It therefore does not appear to be the case that those who are able to see that a face has been doubled are more affected during recognition.

2.6 Discussion

As expected, there is a very strong effect of inversion on recognition rates. Doubling has a relatively small effect. The analysis shows an interaction between the two factors; doubling does not further reduce identification when inverted. A possible explanation is that participants simply couldn't tell that a face was doubled in this condition, however, the evidence from the detection experiment is otherwise. On average, detection performance is clearly above chance (Figure 2). While there are big individual differences in detection ability, these do not relate to the *differences* in identification performance between normal and doubled faces, only with overall identification performance. If difficulty in seeing that a face was doubled was the explanation for the lack of difference in identification when inverted, then those participants who could see the doubling ought still to show a difference.

One aim of the identification study was precisely to see whether there would be an interaction with inversion. The causes of the effect of inversion are debated (Rossion 2008; Rossion 2009; Riesenhuber and Wolff 2009; Yovel 2009) but it seems agreed that our ability to process a face holistically is at least much reduced by inversion (e.g. Hole 1994). Rossion (2009) argues that this loss of holistic processing is the underlying cause of other apparent disruption, such as the ability to do configural processing, which is generally held to be the reason for the Thatcher illusion disappearing on inversion. Adding an extra mouth and pair of eyes ought to have a severe effect on the apparent configuration of the face, depending on exactly how they are added, and how they are processed. If the original eyes and mouth are moved up on the face, to accommodate a copy below, then the average position of the doubled features might remain relatively unaffected. Whether our perceptual system would perform such averaging for face configurations is questionable. In any case, most of the doubled images here left the original eyes and mouth in place and simply added a copy underneath, with exceptions mostly for mouths where the chin was too small to accommodate an extra one. Therefore the apparent configuration of the face should alter, even if some averaging of the location does go on during perception. The individual features, however, remain relatively unaffected; there are simply more of them. Inverted faces are recognised more by features; on both counts, therefore, it might be expected that doubling would have less effect on the recognition of inverted faces. Our data support this expectation; doubling does not significantly affect the routes to recognition afforded by inverted faces.

The detection data show wide individual variations in ability. A couple of participants were responding reliably at durations of around 30ms, at which speed the first author is barely able to see a face. Others clearly struggled with the task, with performance at chance at 53ms, and the adaptive procedure yielding thresholds of around 100ms. Detection ability showed a mild overall correlation with identification rate, such that those who detect doubling faster were also better at identification. This is consistent with general differences in speed of processing. However, there was no evidence that detection ability had any bearing on the relative effect of doubling on identification.

An unexpected feature of the results is the shift towards reporting double with inversion. The pattern, in Figure 2, is similar at both presentation durations, but less extreme at 80ms. One possibility is that participants are using a 'looks odd' decision, and since inverted faces do 'look odd' there may be a tendency to confuse the two. This would be in contrast to the Thatcher illusion, where inversion makes the illusion less conspicuous. Lewis (2001) found that participants were faster to classify a face as Thatcherised than as normal when upright, but became slower as the faces were rotated. The next experiment sought to test doubled faces in the same way.

3 Experiment 2

The aim of this experiment was to test how quickly a face could be classified as normal or doubled at various orientations. Lewis (2001) tested the Thatcher illusion every 10 degrees from upright to inverted. Such fine step sizes seemed unnecessary here, since, unlike the Thatcher, the double illusion does not fade with inversion, so there is less interest in determining exactly when a change occurs. Accordingly, we used steps of 30 degrees.

3.1 Participants
Twenty members of staff and students from the University of Stirling participated voluntarily, 12 male, mean age 28.1 years, normal or corrected to normal vision.

3.2 Stimuli
The same ten male and female unfamiliar faces as in Experiment 1 were used, but with versions of each pair of images at 0, 30, 60, 90, 120, 150 and 180 degree rotations, with 30-150 both clockwise and anti-clockwise.

3.3 Design
Within participants: independent variables were face type, normal or doubled; and orientation, 7 angles. Dependent variable was reaction time to decide on the face type.

3.4 Procedure
Participants were tested in a variety of quiet locations. Images were displayed using Eprime on a laptop computer, with a 14" matte screen, and a resolution of 1024x768. Participants were shown an instruction screen, explaining their task, with examples of the two types of images. They then had a practice block of 24 trials: one of each type of face at each of the 12 orientations (every 30 degrees around the circle). Responses were made via a two button box, with the button allocated to normal and double faces counterbalanced across participants. For each trial, the software chose one of the 20 stimulus identities at random. At the end of the practice block there was a break to establish that the participant was clear about the procedure. When ready, they started on the main sequence of 96 trials (4 in each condition), with a break programmed half way. The whole study lasted about 6 minutes.

3.5 Results
Any responses taking more than 3sd (214ms) above the mean (502ms) were eliminated from the data, this removed a total of 17 responses out of 1920. There were no improbably fast reaction times. Since with reaction time experiments, there is a danger of a speed-error trade off, the accuracy of the decisions was first checked. Average accuracy was 96.8%, varying between 94.2% and 100% in individual conditions. Pearson correlation between average RT and average accuracy in each condition was not significant, r_{14}=.07, p=.81. There is no evidence for a speed-error trade off.

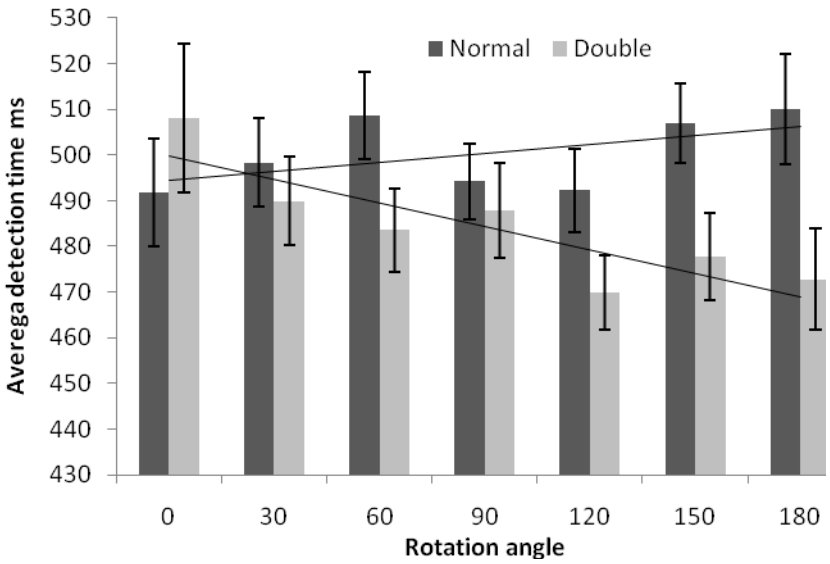

Figure 4 Mean reaction time to decide whether face is normal (dark bars) or double (shaded bars), for correct responses. Regression lines are best linear fit, error bars are standard errors.

Figure 4 shows the average reaction time to decide whether the face is normal or doubled at each of the 7 orientations, with incorrect responses omitted. For normal faces, there is a weak upward trend as the rotation increases; for doubled faces, there is a distinct downward trend. A 2(image type) x 7(orientation) repeated measures ANOVA shows no significant main effects or interaction (all p>.1), but does show a significant linear trend on the interaction between image type and orientation, $F_{1,19}$=4.56, p=.046, η^2_p=.19. Pearson correlation between RT and orientation is not significant for normal faces, r_7 = .53, but is for doubled faces, r_7 = -.86, p=.013. In summary, participants get faster to identify that a face has been doubled as it is rotated away from upright.

3.6 Discussion

There are a couple of striking differences between these results and those of Lewis (2001) for the Thatcher illusion. First is the difference in average reaction times – around 500ms here, compared to 700-1000ms for the Thatcher. It is evidently much easier to detect that a face has been doubled than Thatcherised. Second, Lewis found an increase in decision reaction time for both normal and Thatcherised faces, where we show a clear decrease for doubled faces and no change for normals. The change for doubled faces is not great, around 35ms, far less than the ~300ms change found by Lewis for Thatcherised faces, but is significant and appears linear with rotation. One explanation is that doubled faces, while still obvious, are less unsettling when inverted. Participants are therefore able to respond without any delay caused by the perceptual unpleasantness of an upright doubled face. This explanation may seem unsatisfactory, because it might suggest that doubled faces should be slower than normal ones when upright, rather than faster when inverted. However, it is possible that there are two effects at work: Lewis (2001) found that Thatcherised faces were identified considerably faster than normal ones when upright. It is plausibly easier to make an ‘odd’ decision than a ‘normal’ one. This would act to speed up the double decision when upright, mitigating any delay caused by the perceptual disquiet. On inversion, the disquiet is removed, leaving an advantage for the ‘odd’ decision.

An alternative is the same bias explanation advanced for the tendency to respond double when inverted in Experiment 1. If there is a tendency to see inverted faces as odd it may be easier, and therefore faster, to respond double when the face is inverted. However, this explanation would suggest that there should be more errors for inverted normal faces, with a tendency to say they are double. Performance is close to ceiling, rendering differences unreliable, but there is no hint of this in the data; accuracy for inverted normal faces was 99%, above the overall average.

4 Experiment 3

As is evident from Figure 1, upright doubled faces are quite arresting. There is mixed evidence about the extent to which faces pop out from other items, but the consensus seems to be that they do not (Nothdurft 1993; VanRullen 2006). Murray (2004) tested search for one face with an inverted mouth in a field of normal faces, amongst other conditions, and found no sign of pop out, with a search time of 73ms per face in the array. Our final experiment was designed to test whether normal and double faces might pop out from each other, with the expectation that they would not. However, it would also show whether double faces are indeed slower to be processed, as indicated by a greater search time by item.

4.1 Participants

Twenty students, staff and visitors to the university took part voluntarily, 10 male, average age 23.7 years.

4.2 Materials

A different set of 5 male and 5 female faces from the Home Office collection used in experiment 2 were used. Double versions were created in Photoshop and both normal and double images saved at 135x180 pixels.

4.3 Design

Design was 2(face type) x 2(target present/absent) x 4 (array size), within participants.

4.4 Procedure

Participants were tested individually in a quiet location, using the same laptop as before. The task was explained and they then had 16 practice trials; two in each condition for both 4 and 6 items. Items were displayed by randomly picking the requisite number out of a total of 20 possible locations, arranged in a 5x4 irregularly spaced grid. The irregularity consisted of image frames being displaced by 10-20 pixels in the horizontal axis, which contributed to a sense that faces were appearing in random locations on the screen. The maximum extent of the array items was about 25 x 22.5 cm, about 28° x 25° at the viewing distance of about 50cm. Each face was approximately 3.5cm, 4° wide. In target absent trials, the faces displayed were either all normal or all double, all the same identity for a given trial. In target present trials, one of the faces was replaced by the other type of the same identity. The task was to decide whether there was an odd one out, and respond via a two-button box, with the button used for each response counterbalanced across participants. At the end of the practice block there was a pause to check the participant was clear about the procedure, whereupon they completed the main sequence of 128 trials, with 8 in each condition for 4, 6, 8 or 10 faces. A break was scheduled after each 32 trials. The whole study lasted about 8 minutes, after which any further questions were answered.

4.5 Results

There were 28 responses out of 2560 taking more than 3sd (1356ms) above the mean (2511ms), but 13 of these were from 10 item displays and only 1 from a 4 item display. This suggests that not all were true outliers, such as participants being distracted. We therefore chose to omit only those responses below 400ms, as being improbably fast, and use median reaction times in analyses.

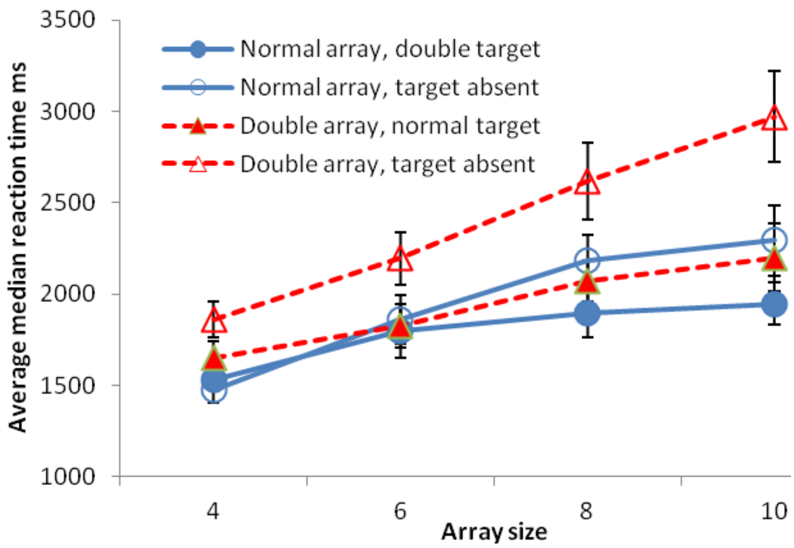

Figure 5 Average median reaction times by array size for the four conditions. Error bars are standard errors, filled symbols are target present; dotted lines are double face arrays.

Figure 5 shows the average median reaction times for each condition. It is apparent that all four conditions show a rise with number of items in the array, while in all cases, an array of double faces takes longer than the equivalent normal face condition. A linear regression was performed separately for each participant's data in each of the four conditions, to give slope and intercept estimates. A 2x2 repeated measures ANOVA on the slope estimates showed a significant effect of target presence ($F_{1,19}$=9.15, p=.007, η^2_p=.33) and of face type ($F_{1,19}$=16.4, p=.001, η^2_p=.46) but no interaction (p=.35). As expected, it takes about twice as long to search when there is no target present (M=163ms per item) than when there is a target (M=81ms). It also takes longer to search through double faces (M=141ms) than normal ones (M=103ms). Although the interaction is not significant, this difference is mostly driven by the target absent condition. With target absent, double faces (M=188ms) take significantly longer than normal faces (M=139ms, t_{19}=3.3, p=.004), whereas with target present, the two face types do not differ (double, M=95ms, normal, M=67ms, p=.14).

4.6 Discussion

There is no evidence for popout of either normal or double faces. This would have shown up as a very low slope in the target present condition but the data show the slope in target present to be almost exactly half that of the target absent condition. This confirms the subjective impression from doing the experiment: you simply have to search through all the faces present. In piloting we tried other arrangements and larger faces, about twice the current width, but the subjective experience was the same. The average search time per normal face, at 67ms in the target present condition, is similar to the 73ms found by Murray (2004) when searching for a face with an inverted mouth amongst normal faces.

The search time per face is significantly longer for double faces, principally in the target absent condition, where the full array must be searched each time. It simply seems to take a bit longer to process each face, on average 48ms in the target absent condition. We speculate that it may be no coincidence that this is close to the 35ms difference found in Experiment 3 between decision times for upright and inverted double faces. Refer to Figure 1: the inverted face is obviously wrong, but does not have the same perceptual effect. It would appear that this perceptual effect adds around 40ms to the processing of an upright doubled face, which is removed on inversion. If this is the explanation, then inverting the faces should speed up the search through an array of doubled faces. We tested this prediction by rerunning the experiment with an inverted array condition.

4.7 Experiment 3b: participants

Twenty-eight students from the University of Stirling, average age 20.6 years, took part voluntarily. Three were excluded for excessive error rates (25%, 39% and 44% errors against an average of 8%).

4.8 Design

In order to limit the number of trials, the experiment was simplified to two different sizes of array, 6 and 12, resulting in a 2 (array size) x 2 (upright/inverted) x 2 (normal/double array) x 2 (target present/absent) design, with 8 trials per condition. All other details as before.

4.9 Results

As before, we excluded responses below 400ms and then analysed median RTs. We considered only the target absent condition, as we are interested in the effects of inversion on search time, which is maximised with no target to find. Slope estimates for each participant were computed by subtracting the reaction time for an array size of 6 from that for array size of 12 in each condition and dividing by 6. A 2x2 repeated measures ANOVA showed an effect of orientation ($F_{1,24}$=5.61, p=.026, η^2_p=.19) and a marginal effect of image type ($F_{1,24}$=3.80, p=.063, η^2_p=.14), qualified by a significant interaction ($F_{1,24}$=5.11, p=.033, η^2_p=.18). Planned comparisons showed the search time per item was significantly greater for doubled faces when upright (M=216ms) than when inverted (M=134ms, t_{24}=3.5, p=.002), while orientation had no effect on normal faces (inverted, M= 130ms; upright, M=158ms, t_{24}=0.99, p=.33).

As we expected, inversion speeds up a search through an array of double faces, consistent with our hypothesis that it removes the processing time associated with resolving an upright doubled face. While the effect of inversion is not significant for normal faces, the direction of the difference is the same, with search among inverted faces being faster. This is consistent with upright faces also grabbing attention, as found by Langton et al (2008), who report a 20ms effect of inversion, comparable with the 28ms found here. The search times per item for both normal and double faces when inverted are essentially identical. It therefore seems that there may be a small cost per normal face when upright, but a larger one for double faces. The difference here is 58ms, somewhat larger than the 48ms per item in the main Experiment 3.

4.10 Experiment 3c

One of the referees, Robbie Cooper, suggested that an alternative explanation for the extra time taken to search through double faces is that there are two pairs of eyes looking at you, rather than one. Thus Senju, Hasegawa and Tojo (2005) found that it took about 60ms per item longer to look through a field of direct gaze faces than a field of averted gaze faces when the target, a face of the opposite gaze type, was absent. To test this explanation, we created doubled faces with averted eyes and ran a new version of the search task.

4.11 Participants

Thirty-six students from the University of Stirling, 12 male, average age 21 years, took part voluntarily.

4.12 Materials

Six volunteers were photographed with eyes forward, and looking to either side, taking care to minimise associated head movements. Doubled versions of each were created as before. All faces in a given array were the same identity and all the eyes if averted looked the same way, left and right equally often.

4.13 Design

To limit the total number of trials, only three array sizes (4, 8 and 12) were used, so the design was 3 x 2 (eyes, forward/averted) x 2 (target, present/absent) x 2 (face, normal/doubled), while only three trials were run per condition for a total of 72 trials. All other details as before.

4.14 Results

With only three trials per condition, outliers are more of an issue in this experiment and medians are not very stable. We are principally interested in the target absent condition, since this maximises the effect of having to look through the set of faces. We therefore again considered only target absent responses and computed mean and standard deviation of RT for each array size separately, then removed those responses that were more than 2 SDs above each mean. This resulted in three participants for whom there were no correct responses in a condition, so these three were omitted from the analysis. There were no responses below 400ms. Slopes were computed for each participant as before and then analysed with a 2x2 repeated measures ANOVA. This gave a significant effect of face doubling (Double, M=182ms per item, Normal, M=145ms, $F_{1,31}$=11.9, p=.002, η^2_p=.28) but no effect of eye direction (Averted, M=161ms, Forward, M=166ms, $F_{1,31}$=0.1) and no interaction. It therefore seems that the effect is due to the double face, rather than an extra pair of eyes looking at the observer. The absence of any effect of eye direction, contrasting with the 60ms per item reported by Senju et al (2005), may be because their participants were explicitly searching for a face that differed in gaze, whereas gaze was irrelevant for our task. Note that the extra time per item for doubled faces in this experiment is 37ms, rather less than the 48ms from experiment 3and 58ms of Experiment 3b and close to the 35ms of experiment 2.

5 General discussion

Although at first this illusion may seem related to the Thatcher illusion, it actually shares rather little in common with it. Primarily, inversion does not make it inconspicuous, so favourite Thatcher studies such as the rating of grotesqueness with rotation are not relevant.

The detection part of Experiment 1 found an increased tendency to classify a face as doubled when it is inverted. While, as noted above, this might be explained by participants classifying both inverted and doubled faces as ‘odd’, an alternative explanation is offered by the results of the next two experiments. These both indicated that there appears to be a perceptual cost of processing an upright doubled face of around 40ms. In experiment 1, the faces are presented very rapidly. If inversion removes the perceptual overhead, then it should be easier to identify a face as doubled in the time available.

The adaptive procedure revealed wide individual differences in the presentation time required to detect doubled faces; varying by nearly a factor of three, from 35ms to 103ms. These threshold estimates correlated strongly with error rates on the fixed time experiment, which is reassuring. They also correlated, though less strongly, with overall identification performance in the second part of this experiment, suggesting an explanation in terms of speed of perceptual processing. Such variability has been reported by others, for example Codispoti et al (2009) find strong individual differences in the ability to detect the expression of masked faces at 40 and 50ms.

The identification results indicated the expected strong effect of inversion, and a smaller effect of doubling, qualified by an interaction that indicated that there was no additional effect of doubling when faces were inverted. This suggests that the effect that doubling has on identification is likely to be on the apparent configuration, which is less perceptible when inverted.

Experiment 2 was an analogue of the Lewis (2001) study of the Thatcher illusion. It found very different results, with inverted doubled faces being detected more rapidly than upright ones. We have argued above that this is because inversion removes a perceptual overhead associated with upright doubled faces. Doubling is still apparent when inverted, but less disturbing to look at.

Experiment 3 checked whether normal or doubled faces might pop out from each other. They did not, but the experiment showed that it takes longer to search through doubled faces than normal ones. Subsidiary experiments demonstrated that this processing cost disappears when the faces are inverted (3b) and is not due to a second pair of eyes looking at the observer (3c). The three studies suggest that the delay is in the range of 40-60ms per item.

There is a considerable literature on whether some faces either draw attention, or hold it, during a search task. Thus your own face does not draw attention, but holds it when found in an array (Devue et al 2009). Angry faces have been found to draw attention, especially when schematic faces are used, but the evidence with photographic images is more confused, with recent data suggesting that happy faces are more detectable, due to the high saliency of the smiling mouth (Calvo and Nummenmaa 2008). Our search task used neutral faces and it is possible that, had smiling faces been used, the doubled smile would have drawn attention. Angry faces, meanwhile, have been shown to hold attention: Belopolsky et al (2011) used an eye-tracker to show that it takes longer to move away from a threatening face. A similar method could be used to confirm whether doubled faces actively hold attention, or simply take longer to process.

6 Further work

We suspect that the double illusion will not spawn as much research as the Thatcher illusion has; it seems more of a perceptual oddity. Future work might seek a better understanding of what it is about upright double faces that is so unsettling. Our speculation is that it is consistent with some kind of internal face-finding template. New born babies have been shown to orientate towards a very simple face-like stimulus, consisting of two blobs above a single central blob, in an oval (Mondloch et al 1999). Debate continues as to whether this reflects a real face template, or something simpler, but suppose we do have some kind of simple face template, effectively two eyes over a mouth, Figure 6a (note that, for this argument, whether or not any such template has a nose makes little difference). Presented with a doubled face, Figure 6b, it then has four possible ways to fit the features, Figure 6c-f. Thus the template mouth could locate on either the top (Figure 6c and 6e) or bottom (Figure 6d and 6f) mouth of the face, and similarly for the eyes. Our speculation is that the sense of disorientation reflects the inability of this template to find a stable best fit. This may be rather like the effect that is sometimes observed when first viewing a repeated pattern, such as a net. It can be hard to focus on initially, because there are multiple possible solutions for how to match the image coming from the two eyes. Eventually a best fit is found, and the net pops into focus. Observing a doubled face feels similar, but there is no resolution and the unsettled feeling persists. However, while some such internal template is possible, finding a complete fit is clearly not necessary, given that we are able to locate and identify faces despite key features, such as the eyes, being obscured (Roberts and Bruce 1988; Lewis and Edmonds 2003).

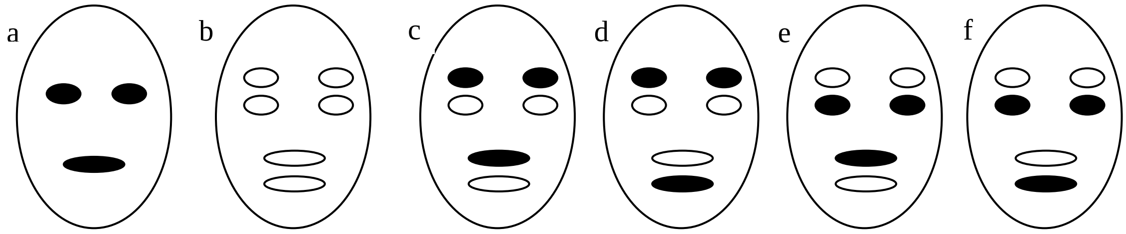

Figure 6 a) simple face template, b) an outline doubled face, c-f) the four possible combinations of eyes and mouth for the template to fit.

Finally, as noted in section 2.6, there are two ways to produce doubled faces, which might possibly have different perceptual effects. The simplest way to produce such an image is to copy the eyes and mouth regions, putting the new copies just below the originals. A slightly more sophisticated alternative is to move the existing eyes, eyebrows and mouth up a little, before placing new eyes and mouth a bit below the originals. The images used here were created by a mixture of the two methods, depending on the configuration of each face. Further experimentation would be needed to establish whether the two have significantly different effects in practice.

Acknowledgements. Thanks to Robbie Cooper and James Bartlett for thoughtful comments on earlier drafts. Amy Smith, Claire McKay, Emma Harrick, Lauren McClung, Stewart Bicker, Catharina Koppitz, Kevin McGurk, Elizabeth Simpson, Tom Atkinson, Kirsten Morrison and Stuart Plowman assisted with data collection.

References

Belopolsky A V, Devue C, Theeuwes J, 2011 “Angry faces hold the eyes” *Visual Cognition* **19** 27-36

Brainard D H, 1997 “The psychophysics toolbox” *Spatial vision* **10** 433–436

Calvo M G, Nummenmaa L, 2008 “Detection of emotional faces: Salient physical features guide effective visual search” *Journal of Experimental Psychology: General* **137** 471-494

Codispoti M, Mazzetti M, Bradley M M, 2009 “Unmasking emotion: Exposure duration and emotional engagement” *Psychophysiology* **46** 731-738

Dakin S C, Watt R J, 2009 “Biological ‘bar codes’ in human faces” *Journal of Vision* **9** 2-10

Devue C, Van der Stigchel S, Brédart S, Theeuwes J, 2009 “You do not find your own face faster; you just look at it longer” *Cognition* **111** 114-122

Farell B, Pelli D G, 1999 “Psychophysical methods, or how to measure a threshold and why” in *Vision research: A practical guide to laboratory methods* Ed R H S Carpenter and J G Robson (New York: Oxford University Press) pp 129–136.

Hole G J, 1994 “Configurational factors in the perception of unfamiliar faces” *Perception* **23** 65-74

Kitaoka A, 2008 “Cognitive Psychology of Visual Illusion” *Japanese Journal of Cognitive Psychology* **5** 177-185

Kleiner M, Brainard D H, Pelli D G, 2007 “What’s new in Psychtoolbox-3” *Perception* **36** ECVP Abstract Supplement

Langton S R H, Law A S, Burton A M, Schweinberger S R, 2008 “Attention capture by faces” *Cognition* **107** 330-342

Leder H, Bruce V, 2000 “When inverted faces are recognised: the role of configural information in face recognition” *Quarterly Journal of Experimental Psychology* **53A** 513-536

Leder H, Carbon C-C, 2006 “Face-specific configural processing of relational information” *British Journal of Psychology* **97** 19-29

Lewis M B, 2001 “The Lady’s not for turning: Rotation of the Thatcher illusion” *Perception* **30** 769-774

Lewis M B, Edmonds A J, 2003 “Face detection: Mapping human performance” *Perception* **32** 903-920

Mondloch C J, Lewis T L, Budreau D R, Maurer D, Dannemiller J L, Stephens B R, Kleiner-Gathercoal K A, 1999 “Face Perception During Early Infancy” *Psychological Science* **10** 419-422

Murray J E, 2004 “The ups and downs of face perception: Evidence for holistic encoding of upright and inverted faces” *Perception* **33** 387-398

Nothdurft H C, 1993 “Faces and facial expressions do not pop out” *Perception* **22** 1287-1298

Pelli D G, 1997 “The VideoToolbox software for visual psychophysics: Transforming numbers into movies” *Spatial vision* **10** 437–442

Riesenhuber M, Wolff B S, 2009 “Task effects, performance levels, features, configurations, and holistic face processing: A reply to Rossion” *Acta Psychologica* **132** 286-292

Roberts T, Bruce V, 1988 “Feature saliency in judging the sex and familiarity of faces” *Perception* **17** 475-481

Rossion B, 2009 “Distinguishing the cause and consequence of face inversion: The perceptual field hypothesis” *Acta Psychologica* **132** 300-312

Rossion B, 2008 “Picture-plane inversion leads to qualitative changes of face perception” *Acta Psychologica* **128** 274-289

Searcy J H, Bartlett J C, 1996 “Inversion and processing of component and spatial-relational information in faces” *Journal of Experimental Psychology: Human Perception and Performance* **22** 904-915

Senju A, Hasegawa T, Tojo Y, 2005 “Does perceived direct gaze boost detection in adults and children with and without autism? The stare-in-the-crowd effect revisited” *Visual Cognition* **12** 1474-1496

Thompson P, 1980 “Margaret Thatcher: A new illusion” *Perception* **9** 483-484

VanRullen R, 2006 “On second glance: Still no high-level pop-out effect for faces” *Vision Research* **46** 3017-3027

Yovel G, 2009 “The shape of facial features and the spacing among them generate similar inversion effects: A reply to Rossion (2008)” *Acta Psychologica* **132** 293-299