Being observed detrimentally affects face perception

Abstract

Face recognition research can be criticised for ignoring the interactive and social nature of real-world face recognition. We present the results of four experiments investigating the effects of social pressure due to being observed on face processing. In Experiment 1, simulated social pressure was manipulated through two factors: whether participants believed they were interacting with other people or not via a webcam and whether they believed they were being recorded or not. Participants who believed they were being recorded, were significantly less accurate at recognising faces than those who did not believe they were being recorded. For Experiment 2, we found that the recognition of own-ethnicity faces was negatively affected by observation but not the recognition of other-ethnicity faces, and then only when observed during learning. In Experiment 3 we demonstrated that observation affected the recognition of upright faces more so than that of objects and inverted faces. Experiment 4 showed that observation does not affect the amount of holistic processing engaged in, but does affect how people view faces. Such results indicate that expert face recognition is susceptible to increased error if participants are being observed whilst encoding faces as a result of changing the way faces are viewed.

Key words: Face recognition; social pressure; ethnicity; stress; face-inversion effect; ecological validity.

Being observed detrimentally affects face perception

The ability to discriminate between so many homogenous stimuli makes face recognition unsurpassed by other human visual skills (Ellis, 1981); this is particually true for familiar faces. This expertise is likely based on some form of expert processing; Faces, in contrast with objects (e.g., Biederman, Mezzanotte, & Rabinowitz, 1982), are processed as a gestalt whole, according to the inter-relations between features, or as a template (see e.g., Maurer, Le Grand & Mondloch, 2002). It is clear from many lines of evidence, including fMRI (Kanwisher, McDermott, & Chun, 1997), EEG (Bentin, Allison, Puce, Perez, & McCarthy, 1996), and behavioural effects such as the face-inversion effect (Yin, 1969) that faces are processed differently to objects, potentially due to the massive expertise that we have processing faces (Gauthier, McGugin, Richler, Herzmann, Speegle, & Van Gulick, 2014). However, much of what we know about face recognition stems from neurological (e.g., Gobbini & Haxby, 2007) or cognitive (e.g., Hole, George, Eaves, & Razek, 2002) research methods employing stringent, laboratory-based experimental paradigms. Given that face-to-face communication (Williams, 1977) and the mere presence of others (Markus, 1978) affects behaviour, these face recognition studies neglect the impact of social cues and have limited ecological validity (Logie, Baddeley, & Woodhead, 1987). This paper reports a series of experiments that explore laboratory-based face recognition while under social observation to provide empirical insights into how social factors can impact on perceptual abilities.

In real life, face processing occurs in social situations. Specifically, people engage in processing and recognition of those they are communicating with. Similarly, there are many situations when face recognition is completed under scrutiny (during eye-witness testing, the computer-based line-up, looking through mug-shots, and at identity checkpoints with pressure to be accurate as part of the job). In social situations, participants need to be accurate to avoid embarrassment. While all these tasks require multiple different processes, most laboratory-studies have neglected the role of social pressure resulting from observation.

There are many reasons to believe that the mere presence of other people will affect face perception. Social facilitation (Allport, 1920; Zajonc & Sales, 1966) is the improvement on task performance due to the mere presence of others. The natural environment provides a wider range of cues (including social ones) than a laboratory setting and is likely to attract a deeper level of processing (Swann, 1984). Social facilitation can be contrasted with social inhibition, where social presence is detrimental to task performance (Pessin, 1933). This inhibition could be caused by a fear of negative evaluation. This has been found to lead eye-witnesses to conform more with co-witnesses (Wright, London, & Waechter, 2010). Zajonc (1965) explained such conflicting findings with drive theory: the presence of others leads to improved performance for well-learnt or easy tasks exploiting the individual's existing behavioural repertoire but inhibits performance for novel or complex tasks. Similarly, evaluation apprehension (the stress elicited by being observed) may elicit anxiety, resulting in performance deterioration or improvement contingent upon task familiarity (Cottrell, Wack, Sekerak, & Rittle, 1968). This would suggest that face recognition performance under observation will depend on how difficult or familiar participants find the task: the more difficult or novel the task the more likely that observation will cause a detriment to performance. This argument would suggest that the effect of observation on face perception is no different to its effect on other forms of object perception.

While drive theory makes overarching predictions about the effects of observation on face recognition performance overall, it fails to specify the mechanisms. For this, we must look at both social psychological theories and the face processing system. Social pressure is thought to force participants to engage in more heuristic processing (Lambert Payne, Jacoby, Shaffer, Chasteen, & Khan, 2003). Heuristic processing is likely to be beneficial for tasks that the participant is familiar with but detrimental to tasks that the participant is not familiar with. To provide evidence for their claims, Lambert et al., employed a process-dissociation procedure (Payne, Jacoby, & Lambert, 2004) during person perception. Their results demonstrated that participants who anticipated their responses being scrutinised publicly were more likely to exhibit stereotyping and prejudice. The theory is that social pressure caused by public scrutiny led to the cognitive control processes not being engaged fully. This, in turn, causes participants to be less able to inhibit their responses and use heuristics during perceptual processing. Thus, the degree of facilitation depends on differences in task difficulty and an individual's skill (Holroyd, Westbrook, Wolf, & Badhorn, 1978). Further, the appraisal of the observation will be affected by individual differences: some participants may find it more anxiety-invoking than others.

In face recognition experimentation, it has been found that stress can lead to lower face recognition accuracy, depending on the appraisal of the stressor (Deffenbacher, Bornstein, Penrod, & McGorty, 2004) and the extent to which the observer was stressed or scared (Valentine & Mesout, 2009). More generally, it has been shown that witnesses that are more relaxed are more reliable and less likely to conform (Vallano & Schreiber, 2011). Stress can be induced through social factors and this, therefore, suggests that face recognition may be affected while under observation. Specifically, face recognition performed under observation would be less accurate than face recognition not performed under observation even though face recognition may be an automatic process (given the ease with which faces are detected and face pareidolia), it is not considered to be based on heuristic processing, except potentially for the processing of other-ethnicity faces (Levin, 2000).

While drive theory might predict that social pressure would affect face recognition, it is well known that not all faces are processed equally. Heuristic processing might be engaged for certain groups of faces more than others (Young, Hugenberg, Bernstein, & Sacco, 2012), specifically those of other-ethnicities to that of the observer (Levin, 2000). The own-ethnicity bias (OEB) is demonstrated by greater recognition accuracy for own-ethnicity compared to other-ethnicity faces (e.g., Meissner & Brigham, 2001). While there are many theories to explain this effect, one thing that they have in common is that own-ethnicity faces are processed using some form of controlled (Meissner, Brigham, & Butz, 2005), expert (e.g., Hills & Lewis, 2011), holistic (Michel, Corneille, & Rossion, 2007), or individuating (Sporer, 2001) mechanisms, whereas other-ethnicity faces are processed shallowly (Levin, 2000) in an inexpert (Valentine & Endo, 1992) or categorical manner (Hugenberg, Young, Bernstein, & Sacco, 2010). For example, when participants are presented with an own-ethnicity face they instinctively orientate their fixations to the most diagnostic visual features (Hills, Cooper, & Pake, 2013a) and these are not likely to be as diagnostic for discriminating between other-ethnicity faces (Shepherd & Deregowski, 1981). Instructions to participants to individuate other-ethnicity faces can lead them to be recognised as accurately as own-ethnicity faces (Hugenberg, Miller, & Claypool, 2007) implying that participants are able to engage in the most appropriate and effortful processing for other-ethnicity faces possibly by controlling their processing. Controlled processing is typically employed for own-ethnicity faces rather than other-ethnicity faces. Any condition that limits controlled processing should enhance the use of heuristics and categorisation processes (Dasgupta, McGhee, Greenwald, & Banaji, 2000). Indeed, Marcon, Meissner, Frueh, Susa, and MacLin (2010) have shown that making a recognition task harder, by shortening the presentation time, increasing the set size, and lag between learning and test, enlarges the OEB. Using a repetition-lag paradigm, Marcon, Susa, and Meissner (2009) have shown that other-ethnicity faces are more affected by repetition during the test phase of a recognition experiment implying that they were not processed in a controlled manner. Based on the notion that social pressure causes participants to engage in more heuristic processing (which is that typically employed for other-ethnicity faces), we would anticipate that differential effects of social pressure on own-ethnicity face recognition and other-ethnicity face recognition.

In the preceding paragraph, we have indicated that own-ethnicity faces are processed using some form of controlled or expert processing. It is well known that expertise in cognitive domains rely on chunking and automatisation (Case, 1985) and this allows for more efficient processing (McCutchen, 1996, 2000). It could be argued that holistic processing is an example of this chunking process. Automatic processes are not considered controlled (Conway & Engle, 1994) and therefore, one would argue that holistic processing is in fact a type of heuristic processing. However, this argument is one of commensurability - the difficulty in translating scientific terms across research fields (Pearce, 2012). Heuristic processing as caused by observation is a reliance on simple cognitive shortcuts and is not a reliance on using automatic expert mechanisms. In this sense, we consider the social psychological definition of heuristic processing to differ from the cognitive psychological one and that automatisation of expertise is not a heuristic.

While there have not been many studies exploring how social pressure might affect face recognition, we can make several predictions: Firstly, we can anticipate that social pressure will affect face recognition, either by enhancing it (if it is considered an easy task) or by reducing it (if it is a difficult task) according to drive theory (Zajonc, 1965). This empirical question was answered in Experiment 1. Given that some face recognition tasks are more difficult than other ones (e.g., the processing of other-ethnicity faces and inverted faces), we can further assess the relevance of drive theory to face recognition. This was assessed in Experiments 2 and 3. Drive theory was insufficient in explaining the findings of the results in Experiments 2 and 3, potentially because other-ethnicity and inverted faces may not be processed in the same manner as upright faces. This led us to assess whether observation was having selective effects in the processing of expert face recognition, potentially by causing more heuristic processing. This was assessed in Experiments 3 and 4. The present studies address the impact of social pressure on face recognition in a novel manner using the rigor of a laboratory-based study but with realistic social pressure. Since there is lack of empirical studies to base theoretical discussions on, we embed theoretical discussions following each experiment.

Experiment 1

To establish if face recognition is affected by being observed, we ran an old/new recognition task. However, we attempted to manipulate the instructions to establish the effect of social interaction and observation on recognition performance. The present experiment used video stimuli, which in some conditions, is perceived to be a live transmission of a moving, speaking, person, with the intention of developing the methodology necessary for ecologically valid research by inducing social pressures. Foreshadowing our results, the effect of being observed was most critical to recognition performance.

Method

Participants

An opportunity sample of 88 Anglia Ruskin University staff and students aged between 18 and 64 years (56 female, mean age 30.6 years, *SD*=11.23) naïve to the purpose of the study were recruited. Sample characteristics in each group are presented in Table 1 - they did not differ significantly across conditions. Participants self-reported they were ethnically White and had normal or correct-to-normal vision and were compensated £3.00. Sample size was determined based on an estimate of effect size (*r*=.46) for how being observed affects false memories (Reysen, 2007) and a Power of .90 (calculated using GPower 3.1). This was the critical sample size chosen throughout the manuscript.

Materials

Stimuli comprised 38 videos clips (spatial resolution 960x540px, frame rate 25fps) of volunteers (19 female, mean age 23.2 years) seated in front of a white background in frontal view, with a white sheet over their shoulders to cover clothing, leaving the head and neck visible. All volunteers were ethnically White, with no distinctive or extraneous features (such as beards or jewellery). The volunteers were recorded counting from 1 to 5, using a Panasonic HDC-SD5 high definition video camera under the same ambient illumination and setting. The videos were edited to last 5 seconds, achieved by selecting from the larger database, volunteers who spoke at a similar rate incorporating a short period of silence at either end of the counting. Their faces subtended 4.6 degree of visual angle. A still image (165x95mm) of each person in a neutral expression with mouth closed was used for the test stimuli.

The experimental procedure was implemented using E-Prime Professional 2 program (Psychology Software Tools Inc., 2010). Stimuli were displayed on a 15' LCD high resolution colour monitor (1280x1024dpi, refresh rate 50hz). Shape counting and short maths problems were used during the distractor phase. Responses were collected on a standard keyboard. A short pen and paper questionnaire was used to obtain participant demographics. To ensure the manipulations were believable, a webcam and microphone were positioned and a second control computer was positioned in front of the experimenter with Skype Internet communication application running to imply messages with collaborators at other institutions.

Design

A 2x2 (see Table 1) between-subjects design was employed to measure accuracy and response time in an old/new recognition task. The first factor was video status (whether the participants believed they were viewing live or recorded stimuli). The second factor was recording belief (whether the participants believed they were being recorded or not). Participants were randomly allocated to one of the resulting four conditions. Counterbalancing was applied by using each video and image as a target/non-target an equal number of times.

Table 1.
Experimental conditions with a description of what the participant believed and sample demographics (number of female participants, mean age in years and standard deviation in parentheses).

		Recording Belief	
		Recorded	Not Recorded
Video Status	Viewing Live	Interaction with real persons (*Condition 1*)	Watching real people (*Condition 2*)
		16 F, M_{age} = 32.18 (13.27)	15 F, M_{age} = 31.86 (10.28)
	Not viewing live	Stimuli to be used in future experiments (*Condition 3*)	Control condition (*Condition 4*)
		13 F, M_{age} = 29.15 (10.27)	12 F, M_{age} = 28.35 (11.11)

Procedure

The experiment consisted of six consecutive phases: set-up; instructions; learning; distracter; test; and debrief. The only difference between conditions was the instructions given. Participants were positioned against a white background with a white sheet covering their clothes and a microphone clipped to the sheet to record their vocal responses (as in the videos). They were informed their participant number was 11 to increase experiment believability.

In the “belief of being recorded” conditions, participants were told they were being recorded during the learning phase and were shown the webcam recording. In the “live video feed” conditions, participants were informed that the video was being transmitted in real time to 19 other participants taking part in the same experiment at different universities and were shown this on Skype. In condition 1 (live video, being recorded), participants were informed their video was being transmitted live and they could see the other participants at collaborating universities. In condition 2 (live video, not being recorded), participants were informed they were not being recorded, but they were viewing a live feed of other participants. In condition 3 (not live, being recorded), participants viewed pre-recorded videos but were being recorded for use in future studies. In condition 4 (not live, not being recorded), participants viewed pre-recorded videos in the learning phase and were not recorded doing so. This is the control condition.

Following instructions, all participants completed the same procedure. During the learning phase, participants viewed 19 5s videos sequentially in a random order. Participants counted aloud from 1 to 5 at the same time as the person in the video. Between each face was a 5s interval during which the participants were presented with the following participant number and a count-down screen.

After this, the distractor phase began, consisting of two 5min irrelevant activities. In the first, participants viewed screens with multiple (1-10) randomly positioned squares and triangles and were required to indicate, using the keyboard (S-square, T-triangle), which was the majority shape. In activity two, participants were presented with simple maths problems and were required to indicate whether the answer provided was true or false, by pressing the appropriate key (Z-true, M-false). The order of trials within these tasks was randomised. Immediate feedback was provided for both activities with a screen with the words “correct” or “incorrect”.

Following this, a forced-choice recognition test was completed, in which 38 still face images were presented in a random order. Nineteen were target images and 19 were non-targets. Participants were required to indicate whether they had seen the image before or not using the keypad (S-seen, N-not seen). Images persisted on screen until a response was made. Between each face there was a white screen for 150ms.

Finally, participants were asked to complete a demographic questionnaire in which their belief in the experimental instructions was established. Participants who were cognisant of the experimental deception were excluded. The experiment took a total of 30mins to complete. The study protocol was approved by University's Faculty of Science and Technology Research Ethics Board.

Results

Four participants were excluded as they did not believe the experiment was genuine (all of whom did not believe the videos were live and two were in condition 2). Participants responses were converted into the Signal Detection Theory (e.g., Green & Swets, 1966) measures of *d'* (accuracy) and *C* (response bias) using the MacMillan & Creelman (2005) method. A 2x2 between-subjects analysis of variance (ANOVA) on the recognition accuracy data (summarised in Figure 1) was conducted with the factors: recording belief (recorded or not) and video status (live or not). Faces were recognized significantly more accurately if participants believed they were not being recorded (mean difference=.29) than if they believed they were, $F(1,80)=4.65$, $p=.034$, $\eta_p^2=.06$. The main effect of video status was not significant, $F(1,80)=0.45$, $p=.50$, $\eta_p^2=.01$, nor was the interaction, $F(1,80)=0.57$, $p=.45$, $\eta_p^2=.01$.

Parallel analyses were conducted on response bias and response time revealing no significant effect, largest $F(1,80)=1.29$, smallest $p=.26$, $\eta_p^2=.02$. We decomposed the effect on accuracy to establish the locus of this social pressure effect and established that it affects hit rate rather than false alarm rate. Parallel 2x2 ANOVAs were conducted and these revealed a significant main effect of recording belief on hit rate, $F(1,80)=4.06$, $p=.047$, $\eta_p^2=.05$, but not false alarm rate, $F(1,80)=0.61$, $p=.435$, $\eta_p^2=.01$.

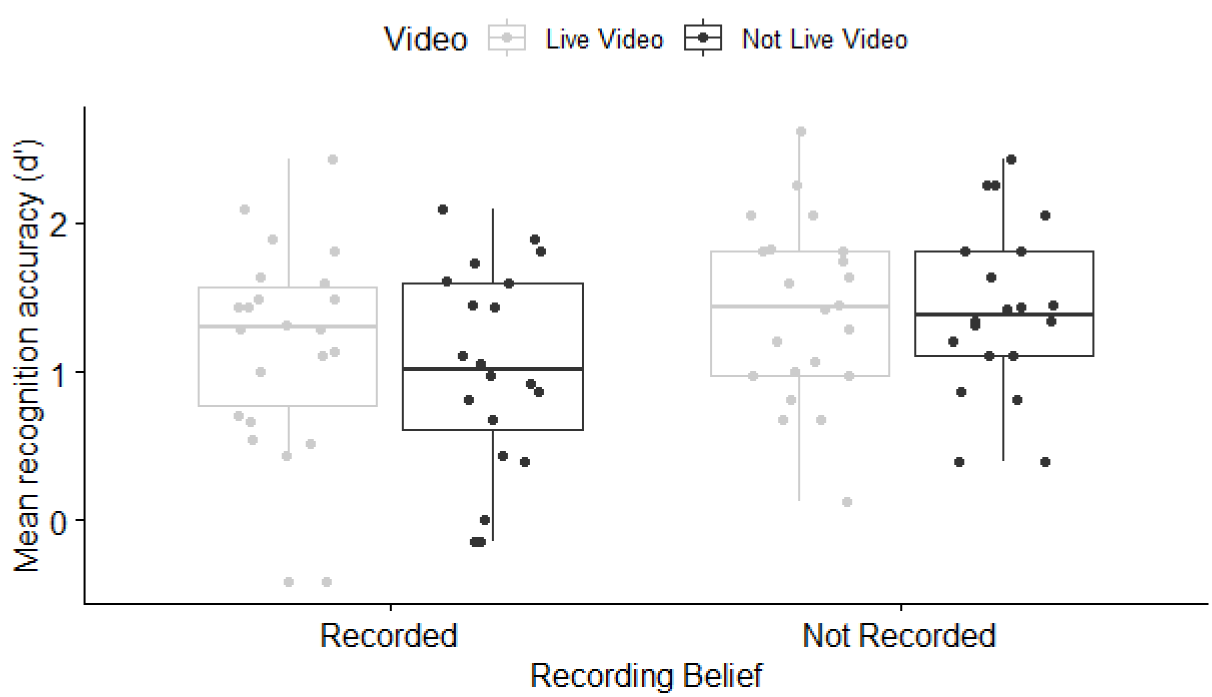

Figure 1. Mean recognition accuracy when participants believed they were being observed split by whether they believed they were interacting with another person during learning.

Table 2.

*Mean (standard error in parentheses) response bias (*C*), response time (*ms*), hit rate, and false alarm rate for participants who believed they were being observed split by whether they believed they were interacting with another person during learning for Experiment 1.*

			Recording Belief	
			Recorded	Not Recorded
Response Bias (*C*)	Video Status	Viewing Live	.03 (.09)	.05 (.09)
		Not Viewing Live	.17 (.10)	.01 (.08)
Response Time (*ms*)	Video Status	Viewing Live	1304 (125)	1336 (133)
		Not Viewing Live	1422 (136)	1528 (153)
Hit Rate	Video Status	Viewing Live	.70 (.03)	.72 (.03)
		Not Viewing Live	.62 (.04)	.74 (.03)
False Alarm Rate	Video Status	Viewing Live	.29 (.03)	.25 (.03)
		Not Viewing Live	.28 (.03)	.26 (.03)

Discussion

Participants who believed they were subjected to recording were significantly less accurate at recognising faces compared to those who did not have this belief. No corresponding difference in response time was observed. Believing that the faces being viewed were live did not impact on recognition accuracy significantly. Perceived social pressure reduces accuracy for face recognition, consistent with work in other fields that has demonstrated performance deterioration in socially pressured conditions (Markus, 1978), including memory accuracy (Echterhoff & Hirst, 2009; Edelson, Sharot, Dolan, & Dudai, 2011).

If we consider this face recognition task to be novel or difficult for our participants, then these results are consistent with evaluation apprehension (Cottrell et al, 1968) and drive theory (Zajonc, 1965): Performance deterioration is observed for novel or unfamiliar tasks that fall outside of a participant's behavioural repertoire. In this task, participants were tested in a novel context (many of the participants would not have undertaken a face recognition test before). Furthermore, we used unfamiliar faces. Unfamiliar face recognition is not as expert a process as familiar face recognition (Megreya & Burton, 2006; Young & Burton, 2018). These reasons lead us to believe that unfamiliar face recognition is a difficult enough task to be negatively affected by social pressure caused by observation.

The results may reflect the effect that social pressure caused by being observed caused participants to engage in shallow and/or categorical processing. Such processing is based on heuristics rather than controlled processes (Lamont et al., 2003). Human face recognition is a highly controlled and expert ability involving highly stereotyped eye movements devoted to encoding faces (Althoff & Cohen, 1999), dedicated brain processing mechanisms (Haxby et al., 2001), and specialist cognitive architecture. If social pressure interrupts this controlled processing, forcing participants to engage in shallower heuristic processing, then the faces might be recognised less accurately (e.g., Levin, 2000). This was explored in Experiment 2.

The results of Experiment 1 also indicated that the effect of observation was primarily on target faces rather than lure faces. This suggests that social pressure affected the encoding and/or the storage of the faces, rather than the recognition phase. To adequately explore this, a condition needs to be implemented whereby the social pressure is only present during the recognition test phase, this was tested in Experiment 2. Furthermore, while the effect sizes observed in Experiment 1 were small, they did meet the standard requirement for significance. Nevertheless, it is important to replicate the present findings to ensure we do not have a false positive.

Experiment 2

Being observed caused participants to be less accurate at the face recognition task than not being observed. However, all the faces were of the participants' own-ethnicity. Participants are better able to recognise faces of their own ethnicity relative to other ethnicities (Meissner & Brigham, 2001) potentially due to experience, categorisation mechanisms, or the differential use of expert processing (Young et al., 2012). Nevertheless, it is clear from behavioural data that the recognition of other-ethnicity faces is more difficult than own-ethnicity faces. Therefore, according to drive theory (Zajonc, 1965), participants should be more affected by observation in the recognition of other-ethnicity faces than own-ethnicity faces. Of course there are many ways that the face recognition task can be made more difficult (by adding noise to the images, for example), however, research has established processing differences between own- and other-ethnicity faces based on the use of heuristic versus controlled processing (e.g., Hugenberg et al., 2010). We are, therefore, able to test between a strict drive theory explanation for the effects of observation on face recognition performance and other explanations based on the type of processing engaged in.

As highlighted in the introduction, the effect of observation might cause participants to engage in heuristic processing (Lamont et al., 2003). In Experiment 1, participants should have engaged expert and controlled processing for the faces because they were own-ethnicity (e.g., Michel et al., 2007). When participants are presented with faces that they are not expert at processing, they may engage in more categorical processing (Young et al., 2012). If social pressure is causing participants to engage in more shallow or categorical processing (Lamont et al., 2003), then this should have a disproportionately larger effect on own-ethnicity faces relative to other-ethnicity faces as own-ethnicity faces are not typically processed in this way. Experiment 2, therefore, was designed to assess whether social observation caused more heuristic processing. Such a prediction is the opposite of what drive theory would predict.

Given the results of Experiment 1 suggest that it is the effect of being observed that affected face recognition accuracy, we modified and simplified the procedure for Experiment 2. Instead of using videos to create a belief of interaction, we used a more standard recognition procedure, recognising still images. Nevertheless, we maintained the belief about recording as this was found to be critical to recognition. In Experiment 2, we also manipulated when the recording instruction was delivered. We either gave this at learning or at test. Based on the results from Experiment 1, we anticipate that the effect of observation will be greater at learning than at test.

Method

Participants

An opportunity sample of 82 (54 females) Anglia Ruskin University students aged 18-60 years (mean age 25 years, SD=12.41) took part in this study. Demographic details across conditions are presented in the design section and did not differ significantly across conditions. All participants self-reported they were ethnically White and had normal or correct-to-normal vision. Participation was part of a course requirement.

Materials

The stimuli set used in this experiment consisted of 80 face images taken from the Minear and Park (2004) face database. Images used were 40 White (20 female) and 40 Black (20 female) faces each displaying a neutral expression (a different image of each face was used at the test phase to the learning phase to reduce pictorial recognition). Each image was cropped such that only the face was visible on a white background, showing no clothing and adjusted to be of equal size (640x480px). A Photoshop tool was used to remove any distinctive features (such as facial piercings) and to hide most of the hair. Each image was standardized for position, such that the eyes of each face were presented in the upper mid centre of the computer screen. The Black faces and the White faces have been used in previous research (Mukudi & Hills, 2019) and data from those studies indicate that they were equivalently recognised by own-ethnicity participants: $t(78)=1.14$, $p=.257$, indicating that the faces were equally difficult for own-ethnicity participants to recognise.

The Experiment was implemented in the same way as Experiment 1, with the same distractor task.

Design

A 2x2x2 mixed factorial design was employed with the factors: Recording belief at learning (webcam recorded, not recorded); recording belief at test (webcam recorded, not recorded); and ethnicity of face (own- and other-ethnicity). Recording belief variables were manipulated between-subjects whereas ethnicity of face was manipulated within-subjects. Participants were randomly allocated to one of the resulting four conditions with the proviso that each condition had an equal number of participants in it. Participant demographics were as follows: recorded at learning and test recorded - 15 Female participants, M_{age}=21.00 years (SD=1.20); recorded at learning but not test - 17 Female participants, M_{age}=20.92 years (SD=1.44); not recorded at learning, but recorded at test - 14 Female participants, M_{age}=21.04 years (SD=1.28); and not recorded at either learning nor test - 18 Female participants, M_{age}= 21.31 years (SD=1.49). Counterbalancing was applied by using each stimulus image as a target and a non-target in each condition an equal number of times. The dependent variables were recognition accuracy (measured using the Signal Detection Theory measure of stimulus discriminability, *d'*), response bias (*C*), and response time (*ms*).

Procedure

The experiment consisted of six consecutive phases, as in Experiment 1. Participants in the "belief of being recorded" conditions were instructed that they were being watched by the experimenter for their facial reaction to the faces and were shown the position of the webcam and an example of the live feed on the experimenter’s computer. Participants were told when the webcam was recording or not immediately before the relevant experimental phases, depending on condition.

Apart from the instructions given, all participants completed the same procedure. During the learning phase, participants viewed 40 (20 Black) face images sequentially in a random order. Participants were required to rate each face for attractiveness on a 1 to 9 scale by responding using the appropriate number keys on the keyboard. This was to ensure adequate attention was paid to the face (Light, Hollander, & Kayra-Stuart, 1981). Each face was on screen for 2000ms and there was an inter-stimulus interval of 800ms.

Immediately following this, the distractor phase as in Experiment 1, began. Following this, a forced-choice recognition test was completed, in which 80 face images were presented sequentially in a random order. Forty were target images and 40 (20 Black) were novel faces. Participants were required to indicate whether they had seen the image before or not using the keypad (S-seen, N-not seen). Test images persisted on screen until a response was made. Between each face there was a plain white screen for 150ms.

As in Experiment 1 participants completed a demographic questionnaire and those who were cognisant of the experimental deception were excluded. The experiment took a total of 30mins to complete.

Results

Two participants were excluded due to not believing they were being recorded at test. Mean recognition accuracy is presented in Figure 2 and was subjected to a 2x2x2 mixed analyses of variance (ANOVA) with the factors: Recording belief at learning; recording belief at test; and ethnicity of face. As expected, this revealed a significant main effect of ethnicity of face on accuracy, $F(1,76)=30.13$, $p=.001$, $\eta_p^2=0.28$, whereby own-ethnicity faces were recognised more accurately than other-ethnicity faces (mean difference=0.35). A significant main effect of recording belief at learning was also found, $F(1,76)=5.87$, $p=.02$, $\eta_p^2=.07$, in which faces were recognised more accurately when participants were not being webcam recorded than when they believed they were (mean difference=0.31). Furthermore, a significant interaction was found between the factors ethnicity of face and recording belief at learning, $F(1,76)=6.65$, $p=.01$, $\eta_p^2=.08$, whereby own-ethnicity faces were recognised with significantly higher accuracy than other-ethnicity faces when not being webcam recorded at learning (mean difference=0.47, $p=.01$), but there was no difference in recognition accuracy between own- and other-ethnicity faces when being recorded (mean difference=0.19). No main effect for recording belief at test was found, $F(1,76)=.03$, $p=0.863$, $\eta_p^2< .01$, nor were any further interactions significant, largest $F(1,76)=0.65$, smallest $p=.42$, $\eta_p^2=.01$.

Response bias data, shown in Table 3, were subjected to a parallel mixed-ANOVA. The only significant effect in this analysis was a more conservative response bias when recognising other-ethnicity faces than own-ethnicity faces (mean difference=.336), $F(1,76)=50.93$, $p<.001$, $\eta_p^2=.40$. The response time data were also subjected to a parallel mixed ANOVA, which revealed a significant main effect for recording belief at test, $F(1,76)=6.35$, $p=.01$, $\eta_p^2=.08$, in which participants were 140.6ms slower at making a recognition judgement if being webcam recorded at test. No further main effects were found, largest $F(1,76)=1.35$, smallest $p=0.25$, $\eta_p^2=.02$. However, a significant three-way interaction was present, $F(1,76)=4.18$, $p=.04$, $\eta_p^2=.05$: While participants were slower at making recognition judgements at test when they believed they were being recorded than when they did not, this effect was not present if the participant had been recorded during learning. No further interactions were found, largest $F(1,76)=2.41$, smallest $p=0.13$, $\eta_p^2=.03$.

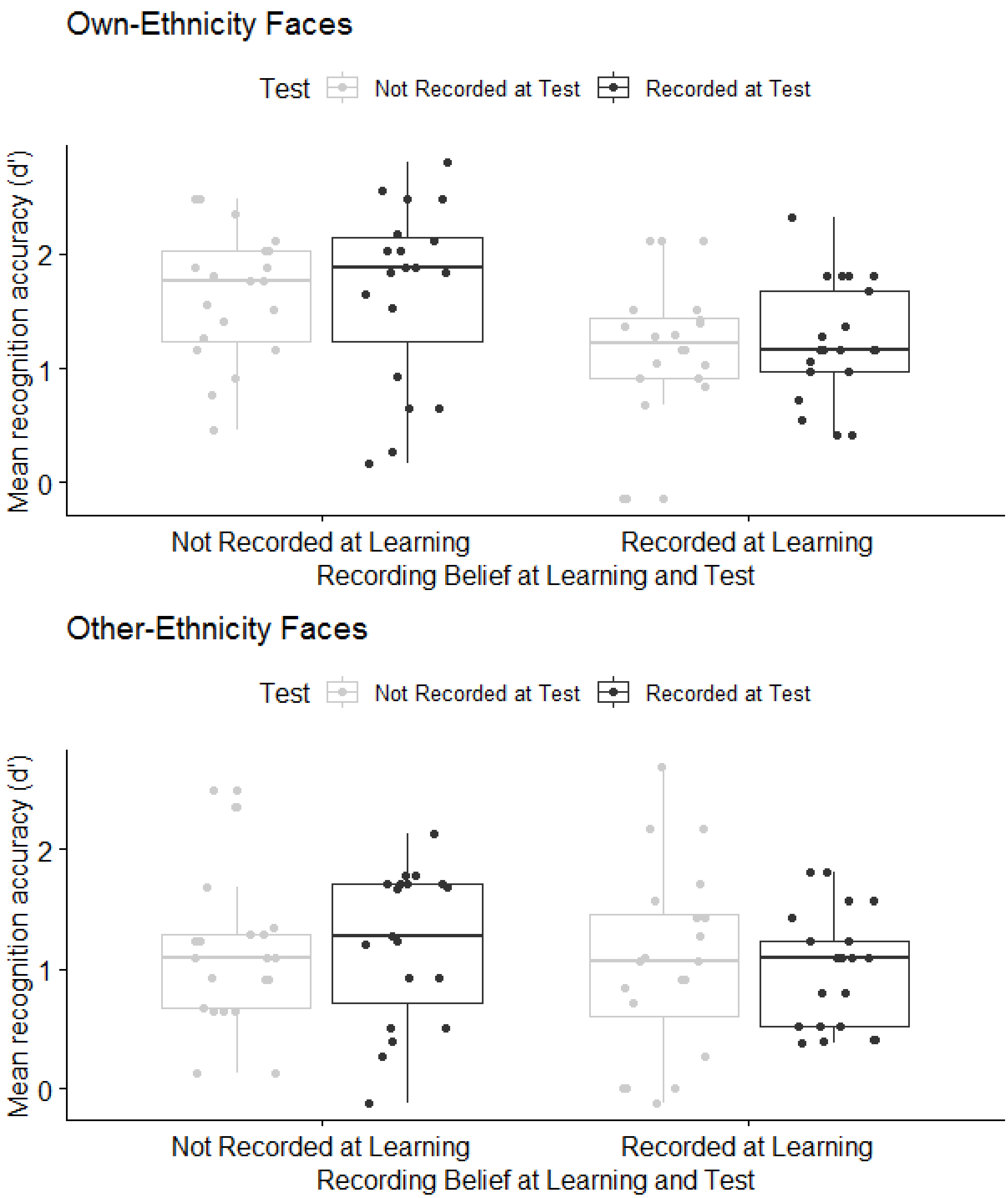

Figure 2. Mean recognition accuracy (d') when participants believed they were being observed split by whether they believed they were being observed by another person during learning for other-ethnicity faces and own-ethnicity faces.

Table 3.

*Mean (standard error in parentheses) response bias (*C*), response time (*ms*), hit rate, and false alarm rate for participants who believed they were being observed split at learning and test for own- and other-ethnicity faces for Experiment 2.*

			Recording belief			
			Recorded at learning		Not Recorded at Learning	
			Recorded at Test	Not Recorded at Test	Recorded at Test	Not Recorded at Test
Response Bias (*C*)	Face Ethnicity	Own-ethnicity	.42 (.07)	.45 (.09)	.51 (.07)	.42 (.07)
		Other-ethnicity	.07 (.08)	.18 (.07)	.14 (.09)	.07 (.09)
Response Time (*ms*)	Face Ethnicity	Own-ethnicity	1150 (231)	1015 (283)	1164 (303)	990 (138)
		Other-ethnicity	1137 (209)	1065 (324)	1170 (138)	988 (324)
Hit Rate	Face Ethnicity	Own-ethnicity	.57 (.04)	.54 (.04)	.62 (.04)	.64 (.04)
		Other-ethnicity	.65(.04)	.63 (.04)	.65 (.04)	.67 (.04)
False Alarm Rate	Face Ethnicity	Own-ethnicity	.16 (.02)	.17 (.02)	.11 (.02)	.13 (.02)
		Other-ethnicity	.30 (.03)	.27 (.03)	.24 (.03)	.29 (.03)

Discussion

Recognition accuracy for own-ethnicity faces was higher than for other-ethnicity: the classic OEB. Furthermore, being observed made participants less accurate at face recognition, however, this was only if the observation occurred at learning and only for own-ethnicity faces. These results indicate that the influence of social pressure due to observation affects the encoding of faces rather than recollection or recognition. However, this impact of social pressure, induced by intensely watching participants on a webcam whilst they were initially encoding the face stimuli, appears to have had a significant detrimental effect on accuracy when recognising own-ethnicity faces only. The recognition accuracy of own-ethnicity faces when being recorded at learning was brought down to the level of the recognition accuracy of other-ethnicity faces in all conditions. According to drive theory (Zajonc, 1965), the recognition of other-ethnicity faces should have been more affected by social pressure caused by observation. Therefore, we can rule out drive theory as an explanation for the present findings.

To interpret these results within the framework described in the introduction, we could suggest that social pressure induces a reliance on heuristic processing. The recognition of other-ethnicity faces is already based on heuristics (e.g., Levin, 2000; Sporer, 2001) and thus is not going to be affected by social pressure. This is the pattern that we found. Specifically, being observed lowered the hit rate for own-ethnicity faces but had no effect on other-ethnicity face recognition. Social pressure may have caused own-ethnicity faces to be processed in a similar categorical manner which leads to poorer recognition. However, this explanation would assume that the metric in which performance detriment is observed through the recognition of other-ethnicity faces would be the same metric as when being observed. This is not the case: The OEB is driven by false alarms, that is heuristic processing causes participants to more easily confuse faces that they haven't seen for ones that they have; However, the effects of being observed are on hit rate, that is the effect of encoding.

Alternatively, the finding that observation affects own-ethnicity face processing but not other-ethnicity face processing may indicate that social pressure interferes with the expert face processing system. This is based on the notion that other-ethnicity faces are processed in an inexpert manner (Michel et al., 2007) relative to own-ethnicity faces. This alternative explanation stems from the notion that face recognition is often considered "special" (Moscovitch, Winocur, & Behrmann, 1997) and relies on different cognitive architecture than object recognition (Haxby et al., 2002). If observation affects the processing of faces by disrupting this specialist cognitive architecture then it is likely to affect upright own-ethnicity face recognition more so than the recognition of other-ethnicity faces, objects, and inverted faces. This alternative prediction was assessed in Experiment 3.

Being watched during the retrieval stage also appears to increase the time it takes to make a recognition judgement of previously seen faces, of both own- and other-ethnicity. Additionally, recognition judgements for own-ethnicity faces were made much slower (and notably less accurate) if participants believed they were being recorded during encoding. The effect of slower judgements, whilst under pressure at the retrieval stage may be due to participants taking more time to deliberate over whether they have seen the face before, rather than making an automatic judgement. It may also indicate that there has been a disruption to the expert face processing system and a slower more piecemeal processing system must be engaged, consistent with the previous theorising.

Experiment 3

We have found that the recognition of own-ethnicity faces is detrimentally affected by being observed during encoding of those faces. We suggest that this is caused by a decreased use of expert processing mechanisms when being observation. It is possible that observation disrupts the expert, and potentially dedicated, processing mechanisms employed for face recognition. This was tested in Experiment 3, in which we ran an old/new recognition experiment in which the recognition of faces was compared with the recognition of objects and inverted faces when either under observation or not under observation. If the social pressure caused by being observed affects the expert mechanisms involved in face processing uniquely then observation will not affect the recognition of inverted faces nor objects as much as upright faces. This is because there is no evidence to suggest that inverted faces are processed heuristically. Once again, drive theory (Zajonc, 1965) would predict that the recognition of inverted faces and objects to be more affected by observation than upright faces because such stimuli are more difficult to recognise.

Method

Participants

An opportunity sample of 108 (43 male) Bournemouth University students aged between 18 and 26 years (mean age 22 years, SD=4.99) naïve to the purpose of the study were recruited. Participants self-reported they were ethnically White and had normal or correct-to-normal vision were and participated as partial fulfilment for a course requirement. Thirty five female participants (out of 52, M_{age}=19.88 years, SD=2.20) were observed and 30 female participants (out of 56, M_{age}=23.96 years, SD=5.98) were not observed.

Materials, Design, and Procedure

Two images of 80 faces were taken from the Minear and Park (2004) database: one image was used in the learning phase and the other in the testing phase, counterbalanced across participants. These faces were ethnically White aged between 18 and 34 years. Each image was cropped and standardised in the same way as Experiment 2.

Images of 30 chairs were taken from the Object Databank (CNBC, 2012). These were cropped to be the same size as the face stimuli and standardised for position, so the centre of the image was presented in the upper mid centre of the computer screen. All images were presented in 72 dpi resolution. The experimental procedure was implemented in the same way as Experiment 1, except that participants responded with "z" (seen before) and "m" (not seen before) in the recognition phase. Moreover, there were two old/new recognition tests: one for faces and one for objects (each one involved a learning, distractor, and test phases). The object test involved 15 chairs presented at learning and 30 at test. The face test involved 40 faces (20 inverted) presented at learning and 80 (40 inverted) at test. Facial orientation was matched from learning to test.

A 3x2 design was employed with the factors: stimuli type (upright faces, inverted faces, or objects) and observation belief. Observation belief was manipulated between subjects and in the same manner as that described in Experiment 2, whereas stimuli type was manipulated within subjects. All other design aspects were the same as Experiment 1.

Results

Old/new responses were converted into *d'* using the Macmillan and Creelman (2010) method. Mean recognition accuracy across conditions is shown in Figure 3 and was subjected to a 3x2 mixed subjects ANOVA with the factors observation and stimuli type. This analysis revealed an anticipated main effect of stimulus type, $F(2, 212)=31.39$, $p<.001$, $\eta_p^2=.23$. Bonferroni-corrected pairwise comparisons revealed that upright faces were recognised more accurately than inverted faces (mean difference=0.89, $p<.001$), thus demonstrating the standard face-inversion effect, Yin (1969), and objects (mean difference=0.65, $p<.001$), thus demonstrating the superiority of face recognition relative to object recognition. There was no significant difference in the recognition of inverted faces and objects (mean difference=0.24, $p=.077$). This effect interacted with observation condition, $F(2, 212)=3.46$, $p=.036$, $\eta_p^2=.03$. Because there was a significant difference in age between the two groups (with the not observed participants being significantly older than the observed participants, $t(106)=4.77$, $p<.001$, and older participants being significantly more accurate at face recognition than younger participants, Germine, Duchaine, & Nakayama, 2011), this variable was entered into a secondary analysis as a covariate. The key interaction remained significant, $F(2, 210)=3.78$, $p=.024$, $\eta_p^2=.04$, with a slightly increased effect size. Post-hoc comparisons (adjusted due to a significant Levene's Test for equality of variance), revealed that observation impaired upright face recognition, $t(106)=2.06$, $p=.043$, Cohen's $d=0.40$, but had no significant effect on the recognition of inverted faces, $t(106)=0.25$, $p=.807$, Cohen's $d=0.05$, nor objects, $t(106)=0.27$, $p=.790$, Cohen's $d=0.05$. These effects were not driven by floor effects as the recognition of all stimuli was significantly greater than 0, all t's(107)>11.23, all ps<.001, all Cohen's ds>2.17. The main effect of observation condition was not significant, $F(1, 106)$, $MSE=2.47$, $p=.201$, $\eta p^2=.02$.

In order to assess whether the recognition of upright faces under observation was brought to the same level as the recognition of inverted faces or objects (thereby removing all face-specialist processing), we used a series of Bonferroni-corrected t-tests ($\alpha=.01$). Upright face recognition under observation was not significantly greater than inverted face recognition, *t*(51)=2.40, *p*=.040, Cohen's *d*=0.67, and object recognition, *t*(51)=1.37, *p*=.354, Cohen's *d*=0.38, indicating a reduction in the deployment of expert processing mechanisms while under observation.

Because the recognition of inverted faces is lower than that of upright faces, detecting effects in the recognition of inverted faces is more difficult because of floor effects. In order to adequately show that observation is selectively affecting the recognition of upright faces, a relative measure of the face-inversion effect is required (see Hills & Lewis, 2018, for further explanation and calculation). This calculation equates performance of upright and inverted faces and allows for a direct comparison. Even with this statistical procedure, the recognition of upright faces was affected by observation more so than the recognition of inverted faces, *t*(106)=2.01, *p*=.047.

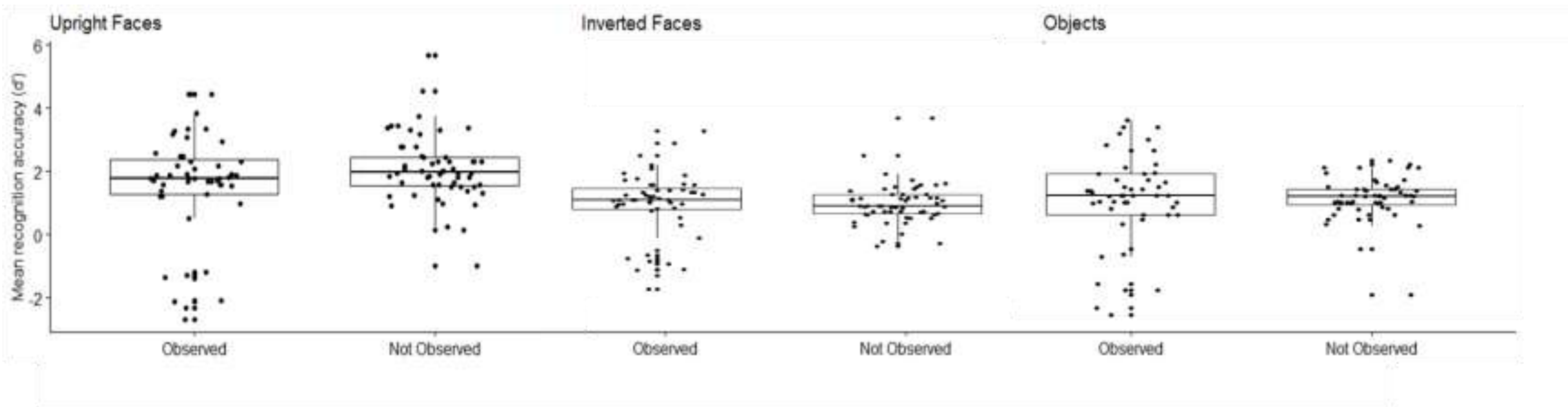

Figure 3. Mean recognition accuracy for upright and inverted faces and objects (chairs) when being observed and when not being observed. Since all accuracy measures were significantly above zero, there were no floor effects in these measures.

In the preceding analysis, we did not control for the overall difference in the unobserved conditions. Because the performance in the upright face condition is that much greater than the other conditions, it is easier to see performance decrements due to a larger range of data points available than in the other conditions. In other words, this may represent a scalable interaction (Loftus, 1978; Wagenmakers, Krypotos, Criss, & Iverson, 2012). In order to resolve this, we constructed a relative measure of the effect of observation[1]:

$$Obs=(d'_{not\ under\ observation} - d'_{under\ observation}) / (d'_{not\ under\ observation} + d'_{under\ observation})$$

The resulting data were subjected to a univariate ANOVA, revealing an effect stimuli type, $F(2, 153)=3.78$, $p=.025$, $\eta_p^2=.05$. Upright faces were recognised less accurately when under observation relative to not when under observation (.14, *SE*=.03) than inverted faces (.02, *SE*=.03) and objects (.03, *SE*=.03).

Response bias data were analysed using a parallel ANOVA, with summary means shown in Table 4. This revealed a main effect of stimulus type, $F(2, 198.03$[2]$)=15.39$, $p<.001$, $\eta_p^2=.13$. Bonferroni-corrected pairwise comparisons showed that response bias was more conservative for objects than upright (mean difference=.34, $p<.001$) and inverted faces (mean difference=.22, $p<.001$). No other effects were significant.

Response time data were analysed using a parallel ANOVA, shown in Table 4. This revealed a marginal main effect of stimulus type, $F(2, 212)=3.24$, $p=.068$, $\eta_p^2=.03$. Bonferroni-corrected pairwise comparisons showed the recognition of upright faces was faster than that of inverted faces (mean difference=102.93, $p<.001$), replicating the standard face-inversion effect (Yin,

[1] Note, that because of the design, this could only be done in a by-item analysis.
[2] Mauchley's Test of sphericity was significant, therefore the Huynh-Feldt correction was applied.

1969). The interaction was not significant, $F(2, 212)=0.21$, $p=.810$, $\eta_p^2<.01$. The main effect of observation condition was significant, $F(1, 106)=3.24$, $p<.001$, $\eta_p^2=.11$: Observed participants (M=1445ms, SD=43) were slower than non-observed participants (M=1229, SD=41).

Table 4.
*Mean (standard error in parentheses) response bias (*C*), response time (*ms*), hit rate, and false alarm rate for participants who believed they were being observed for the recognition of upright faces, inverted faces, and objects (chairs) for Experiment 3.*

			Recording Belief	
			Recorded	Not Recorded
Response Bias (*C*)	Stimuli Type	Upright Faces	.13 (.08)	.11 (.07)
		Inverted Faces	-.02 (.07)	.02 (.07)
		Objects	-.23 (.06)	-.22 (.06)
Response Time (*ms*)	Stimuli Type	Upright Faces	1374 (47)	1181 (45)
		Inverted Faces	1503 (58)	1257 (56)
		Objects	1459 (60)	1250 (58)
Hit Rate	Stimuli Type	Upright Faces	.69 (.03)	.78 (.03)
		Inverted Faces	.64 (.03)	.65 (.03)
		Objects	.72 (.03)	.77 (.03)
False Alarm Rate	Stimuli Type	Upright Faces	.27 (.03)	.18 (.03)
		Inverted Faces	.36 (.03)	.33 (.03)
		Objects	.39 (.02)	.36 (.02)

Discussion

The results from this study indicate that upright face recognition was less accurate under observation, whereas object recognition and inverted face recognition was unaffected by observation. This is inconsistent with drive theory. Such results demonstrate that the dedicated expert perceptual mechanisms involved in face recognition are affected by the social pressure of being under observation. The reason why expert processing might be more easily disrupted by social pressures might be due to the ease with which the specialist architecture is disrupted by external forces. The specialist face recognition mechanisms may be housed in the left fusiform gyrus and nearby regions (Haxby et al., 2002). Damage to these regions causes face recognition deficits (e.g., Barton, 2008), and it may be that this region is simply more susceptible to interference than the more general cognitive architecture employed for object recognition. This explanation may be plausible, however, further tests of how observation disrupts the expert face processing system is required.

In Experiments 2 and 3, we found that observation causes participants to respond slower than when not under observation. While the effects were different across Experiments (in Experiment 3, the effect was at test) and quite small, they indicate that the effect of being observed was causing participants to process faces in a less efficient manner or that their cognitive capacity was somehow occupied by the mere presence of others. During other cognitive tasks, anxiety in participants has been related to slower responding (Mandler & Sarason, 1952) due to altered cognitive processing strategies being employed (Hockey, 1997). Such altered cognitive processing might included increased selective attention (Easterbrook, 1959; Chajut & Algom, 2003) or not using expert face recognition systems.

It is harder to explain the results of Experiment 3 using the notion that social pressure causes participants to rely more on heuristic processing (Lamont et al., 2003). This is partially due to the lack of an operationalisable definition of heuristic processing and partially due to the lack of evidence suggestion that objects and, in particular, inverted faces are processed using heuristics. While this explanation cannot be ruled out (since there is a lack of evidence either way), we prefer to explain the current findings as expert face recognition being more easily disrupted by social pressure than inexpert perceptual processing.

Typically, the expert processing mechanisms deployed for face recognition are considered to be configural processing (Maurer et al., 2002). However, Maurer et al. (2002) highlight that configural processing is a generic term used to describe three distinct processing types: First-order relational processing (coding the basic arrangement of faces); second-order relational processing (detecting and utilising the idiosyncratic differences in the basic arrangement of the facial features, Diamond & Carey, 1986); and holistic processing (processing faces as a Gestalt, Rossion, 2008). Extensive evidence suggests that expert face recognition is not based on fine-grained second-order relational processing (e.g., by stretching the face; Hole, George, Eaves, & Rasek, 2002; Mondloch & Desjarlais, 2010; for a review see Burton, Schweinberger, Jenkins, & Kaufmann, 2015). First-order relational information is disrupted by inversion, suggesting that the face-inversion effect (Yin, 1969) is one measure of expert processing. However, the ability to discriminate between many hundreds of faces cannot be based on first-order relational processing alone. Indeed, it has been theorised that first-order relational processing is necessary to engage expert holistic processing but is not the mechanism of expert discrimination. Therefore, to adequately explore if observation affects expert face processing mechanisms, further tests of holistic processing are required. These were conducted in Experiment 4.

There are alternative explanations for how being observed may affect face perception based on eye movements hitherto not considered. It has been well documented that social anxiety affects the way participants view faces (Schneier, Rodebaugh, Blanco, Lewin, & Liebowitz, 2011). Specifically, eye-contact is reduced in those with social anxiety (Daly, 1978, Horley, Williams, Gonsalvez, & Gordon, 2003, but see Wieser, Pauli, Alpers, & Mühlberger, 2009), or those under social pressure. Participants actively avoid looking at anxiety-inducing stimuli, such as angry faces (Mogg & Bradley, 2002). Social pressure may cause anxiety and this may encourage participants to look away from the eyes as they are the most anxiety-inducing visual features during social interactions. If participants view faces differently, they may be encoding them in an atypical manner (which may or may not be related to holistic processing). Indeed, the eyes are the most important visual feature for recognising White faces (e.g., Althoff & Cohen, 1999; Ellis, Deregowski, & Shepherd, 1975). Increased attention to the eyes correlates with face recognition accuracy (Hills, Cooper, & Pake, 2013b). This offers another explanation for why observation might affect face recognition accuracy and was also tested in Experiment 4.

Experiment 4

Experiment 3 demonstrated that being observed has a detrimental effect on expert face processing mechanisms afforded to upright faces but not objects nor inverted faces. Expert face processing is assumed to be based on some form of configural or holistic processing (Maurer et al., 2002). However, there are at least three forms of configural processing. First-order relational information is used to engage the expert face recognition systems and is indexed by the face-inversion effect (Yin, 1969). Expertise, assumed to be based specifically on holistic processing (Burton et al., 2015), is measured by the composite face-effect (Young, Hellawell, & Hay, 1987), and the parts-and-wholes test (Tanaka & Farah, 1993). These tests measure the ability of participants to fuse the parts of faces together. While these tests conceptually measure configural processing, they do not reliably correlate with each other (Rezlescu, Susilo, Wilmer, & Caramazza, 2017). If observation is disrupting holistic processing, then we should see reductions in these measures whilst under observation.

An alternative view is that observation might lead to faces being viewed in an atypical manner (avoiding eye contact) causing more error during encoding. Typically, when viewing faces, attention is directed toward the most diagnostic features for recognition. The eyes have been found to be the most important facial feature for face processing of White faces by White participants through a variety of methods (Haig, 1985, 1986; Hosie, Ellis, & Haig, 1988; Janik, Wellens, Goldberg, Dell'Osso, 1978). When viewing faces, the eyes receive the most amount of fixation (Althoff & Cohen, 1999; Arizpe, Kravitz, Yovel, & Baker, 2012; Barton, Radcliffe, Cherkasova, Edelman, & Intriligator, 2006; Blais, Jack, Scheepers, Fiset, & Caldara, 2008) with the eyes typically the first feature scanned (Bindemann, Scheepers, & Burton, 2009; Sæther, van Belle, Laeng, Brennen, & Øvervoll, 2009). Disruption to eye scanning can be detrimental to face recognition (Hills et al., 2013b) therefore, we might expect observation to cause a reduction in eye scanning and this cause a reduction in face recognition performance.

Eye movements can also reveal the amount and type (Bombari, Mast, & Lobmaier, 2009) of processing effort (Goldinger, He, & Papesh, 2009) engaged in. By exploring the eye movement patterns, it is possible to establish the type of coding employed (Chan & Ryan, 2012; Guo, 2012). When processing holistically, each fixation is longer (Blais et al., 2008; Miellet, Vizioli, He, Zhou, & Caldara, 2013). Heuristic processing can also be revealed through eye movements. Higher fixation numbers indicate active scanning (Goldinger et al., 2009) which should be observed for non-heuristic processing. Therefore, if observation is disrupting holistic processing, we would expect to see shorter fixations under observation relative to not under observation. If observation is causing heuristic processing we would expect to see lower fixation numbers whilst under observation relative to not under observation.

Experiment 4 was designed, therefore, to assess whether observation affects the type and depth of face coding engaged in. We utilised the three standard measures of configural processing: the face-inversion effect (Yin, 1969); the composite face effect (Young et al., 1987); and the parts and wholes test (1993). Given that these measures do not converge (Rezlescu et al., 2017) and there is significant debate as to what each metric measures (Richler & Gauthier, 2014), we take a fairly exploratory approach as to whether observation will affect each measure of configural processing. Further, the effect sizes of each measure will differ (for example, the effect size for the complete design in the composite face effect is nearly three times that of the partial design, Richler & Gauthier, 2014). In addition, we measured eye movements during the face-recognition task to establish whether observation affects scanning behaviour.

Method

Participants

An opportunity sample of 102 (75 female) Bournemouth University staff and students aged between 18 and 35 (mean age 21.95 years, *SD*=4.22) naïve to the purpose of the study were recruited. Participants had normal or correct-to-normal vision and were either given course credit or compensated £10.00 for their participation. Forty one female participants (out of 50,

M_{age}=19.90 years, *SD*=2.10) were observed and 34 female participants (out of 52, M_{age}=24.10 years, *SD*=2.10) were not observed. These demographic details did not differ across conditions.

Procedure and Materials

Participants completed the three tasks (face recognition, composite face task, and parts-and-wholes task) in a random order. Participants also completed three questionnaires measuring their fear of negative evaluation and their anxiety. Overall, the experiment took one hour to complete.

Face recognition task

The experimental procedure was implemented using E-Prime Professional 2 program (Psychology Software Tools Inc., 2010) and eye movements were tracked using the Eyelink 1000. The Eyelink 1000 camera was mounted on the desktop and positioned 50cm away from the participant. Participants rested their head in the SR Research head support which kept the participant's head still and increased their comfort. Stimuli were displayed on a 15' LCD high resolution colour monitor (1280x1024dpi, refresh rate 50hz). Prior to completion of the task, participants' eyes were calibrated to the eye-tracker. Due to a failure to calibrate to the eye-tracker, four participants data for this task was not collected. The task itself continued in the same way as the face recognition task in Experiment 3.

Questionnaires

The Brief Fear of Negative Evaluation (Leary, 1983) consists of 12 questions in which participants rate their fears of negative evaluation from others on a 5-point Likert scale ranging from 1 ("not at all characteristic of me") to 5 ("very characteristic of me"). The scale has excellent inter-item reliability (α=97) and 2-week test-retest reliability (*r*=94; Collins, Westra, Dozois, & Stewart, 2005). The Classroom Anxiety Measure (Nist & Diehl, 1990) requires participants to rate their agreement with statements about their feeling relating to a classroom situation on a 5-point Likert (1="strongly disagree", 5="strongly agree"). Some questions were reversed to decrease biased responding. The Evaluation Apprehension Questionnaire (Richmond, Wrench, & Gorham, 2001) has an expected reliability of α=.85. It measures participants' feelings of being evaluated in classroom situations on a series of 20 items on a 5-point ("strongly disagree" to "strongly agree") Likert-type scale. These measure feelings of anxiety in specific contexts.

The composite face test

Twelve (6 female) frontal-view faces were taken from the Minear and Park (2004) database. Instead of combining parts of different faces, like the standard task (Young et al., 1987), different identities were created by moving features to increase the chance of participants using holistic processing (Lewis & Hills, 2018) as the isolated features remain constant in all but position. Therefore, identifying the face by the position of the features cannot be completed using featural processing (Lewis & Hills, 2018). Indeed, different eye positions is akin to different face identities, therefore this task is conceptually the same as the standard composite face task. Twelve new images were created by adjusting the distance between the eyes, and a further 12 by moving the mouth up. A further 12 images included both the changes. An additional 48 images were created by splitting the image through the nose and moving the lower half of the face to the right. This resulted in 96 images (8 versions of 12 different identities, see Lewis & Hills, 2018, for the images). Stimuli were 562x762px presented in 320dpi.

Participants were presented with 385 trials. Each trial consisted of a probe face (for 600ms), a blank inter-stimulus interval (300ms), and the target face that remained on screen until participants responded to it. Participants had to decide whether there was a difference in the distance between the face's eyes between the first and second presentation of that face, responding using the "C" for same and "M" for different. Pairs of images corresponded to the conditions in the composite face test (both the partial and the full version).

Parts-and-wholes test

Stimuli were 266 images of which there were 206 whole faces and 60 part (40 eyes, 10 noses and 10 mouths) faces. Only eye trials were analysed but the other features were included at test to prevent participants only focusing on the eyes (Michel et al., 2006). All the faces were ethnically White with neutral expression in frontal poses and presented 240x256px in 72dpi. Hair and necks were cropped and the faces were grey scaled and pasted onto a white background.

Participants were presented with 480 trials. Each trial consisted of a probe face (for 1508ms), a blank screen (305ms), followed by a pair of faces. One was the target identity (a different image of the same face) and one was a different face. Participants were required to decide, using the left or right arrow keys, as to which option corresponded with the target face. Images remained on the screen until a response had been made. Trials were split equally into four blocks. Each block contained a different combination of images (learning-test): whole-whole (WW), whole -

part (WP), part-whole (PW), part-part (PP). The order of the blocks and images within the blocks were randomised across participants.

Results

To assess whether observation affected the magnitude of holistic processing engaged in, we computed metrics of holistic processing. For the composite face task, we computed holistic processing using both the partial (Rossion, 2013) and complete analysis (Richler, Cheung, & Gauthier, 2011) because of the ongoing debate[3] as to which calculation best represents holistic processing. The face-inversion effect was calculated as the difference in recognition accuracy for upright and inverted faces. For the parts and whole task, holistic processing was calculated as follows: (WW-WP) + (PP-PW). This gives a more complete measure of holistic processing (Leder & Carbon, 2005). One-sample *t*-tests confirmed holistic processing as measured by the composite face test (partial), $t(100)=8.08$, $p<.001$, Cohen's $d=1.62$, composite face test (complete), $t(100)=5.25$, $p<.001$, Cohen's $d=1.05$, parts-and-wholes test, *t*(101[4])=6.60, $p<.001$, Cohen's $d=1.32$[5], and the face-inversion effect, $t(97)=21.35$, $p<.001$, Cohen's $d=4.33$.

Figure 4 presents the means for all metrics of holistic processing. There was no effect of observation on holistic processing as measured by the part-and-wholes test, $t(100)=1.27$,

[3] In the partial design, the same top of a face is always paired with a different bottom leading to an incongruence in the response, whereas when the different top is presented, there is a congruence in the response. This suggests that there might a congruence effect in the partial design. Due to the statistical calculation of the composite face effect in the complete design, Rossion (2013) argues that this is measuring response bias rather than holistic processing. Since we are not able to resolve this debate, we have employed a task where both measures can be calculated.

[4] Differences in degrees of freedom across this Experiment represent data loss due to programming error or failure to calibrate to the eye-tracker.

[5] Incidentally, using the simpler WW-WP measure of holistic processing, there was a significant result, $t(100)=4.43$, $p<.001$, Cohen's $d=0.89$. Similar to the reported measure, this was not significantly affected by observation condition, $t(99)=0.97$, $p=336$, Cohen's $d=0.19$.

p=.208, Cohen's *d*=0.25, nor using the partial composite face task, *t*(99)=1.38, *p*=.172, Cohen's *d*=0.28, nor the full complete composite face task, *t*(99)=0.44, *p*=.664, Cohen's *d*=0.09. However, there was a difference in the magnitude of the face-inversion effect between observed and not observed conditions (mean difference=0.60), *t*(96)=2.54, *p*=.013, Cohen's *d*=0.52, driven almost exclusively by the effect of observation on upright face recognition: when observed, upright face recognition was lower than when not observed, *t*(96)=3.42, *p*=.001 Cohen's *d*=0.70; this pattern was not observed for inverted faces, *t*(96)=0.50, *p*=.616, Cohen's *d*=0.10. These data indicate that being observed does not alter the use of holistic processing (as measured by the composite face task and the parts-and-wholes task), but does alter the engagement of it (as measured by the face-inversion effect).

A secondary analysis was run on upright face recognition accuracy data in which we controlled for anxiety and evaluation apprehension (as measured by the chosen questionnaires). Firstly, we established that observation led to higher anxiety scores as measured with the Brief Fear of Negative Evaluation, *t*(100)=1.81, *p*=.035, Cohen's *d*=0.36, one tailed, and the Evaluation Apprehension Questionnaire, *t*(100)=1.73, *p*=.044, Cohen's *d*=0.35, one tailed, but not the Classroom Anxiety Meaure, *t*(100)=0.64, *p*=.523, Cohen's *d*=0.13. We ran a between-subjects ANCOVA, revealing no change in the effect of observation when controlling for anxiety and evaluation apprehension, *F*(1, 92)=12.03, *p*=.001, η_p^2=.12, compared to not controlling, *F*(1, 96)=11.71, *p*=.001, η_p^2=.11. These data indicate that the detrimental effects of being observed on face recognition are not caused by evaluation apprehension nor anxiety.

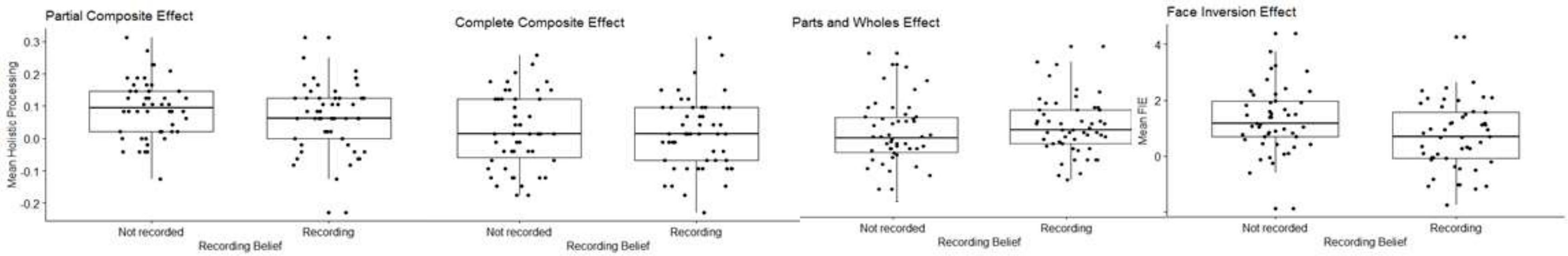

Figure 4. Mean holistic processing as measured by the composite face test, parts-and-wholes test, and the face-inversion effect for participants who were observed and who were not observed. Given that all effects are significantly above zero, there are no floor effects in these results.

Because there is an upright face in all of the tasks, it is possible that the performance in this condition was affected across all tests. Figure 5 shows the means for each condition for each measure. This highlights that the effect of observation was not present in any task other than the recognition paradigm. In other words, no conditions were affected by observation significantly except for the recognition of upright faces.

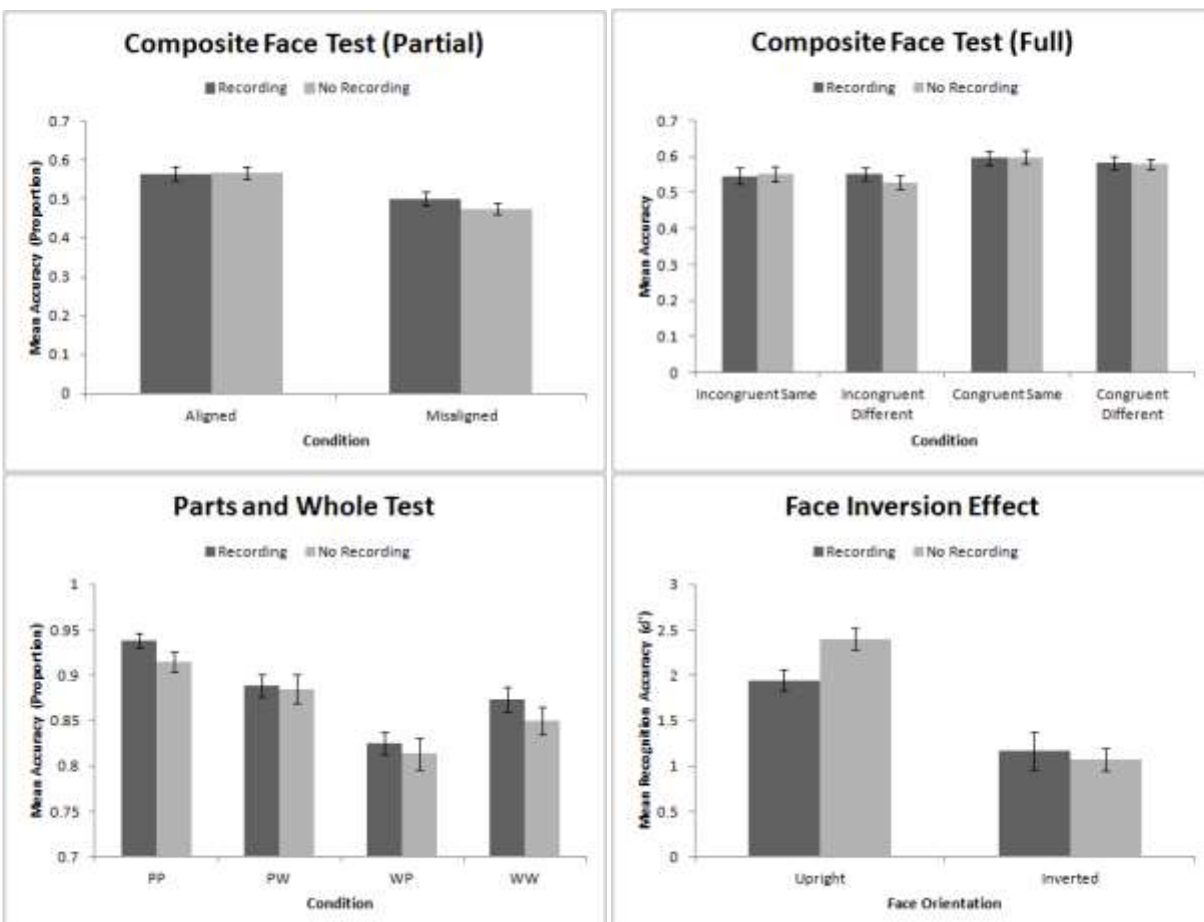

Figure 5. Mean accuracy for the conditions in the composite face test, parts-and-wholes test, and the face-inversion effect. Error bars represent standard error of the mean.

Finally, we ran an analysis on the eye movement data. Critically, we explore the effects of observation on fixations to each feature during the learning of the faces as Experiments 1 and 2 seemed to indicate the effect is one of encoding. We entered the data as 2x2x7 mixed-subjects ANOVA with the factors: observation belief; face orientation; and facial feature (hairline, forehead, left eye, right eye, nose, mouth, and chin and cheeks). We first analysed the number of fixation data, exploring which features were explored more. If observation causes stress, we would expect potential eye movement differences under observation relative to not under observation. This is what we found, as revealed by a significant interaction between observation and feature viewed, $F(6, 163.83$[6]$)=3.72$, $p=.031$, $\eta_p^2=.04$. Between-subjects *t*-tests revealed that under observation, the bottom half (chin and cheeks) of the face, $t(93)=2.18$, $p=.031$, Cohen's $d=0.45$, and the nose, $t(93)=2.27$, $p=.026$, Cohen's $d=0.47$, were viewed more than when not under observation.

Consistent with previous research (Hills, et al., 2013b), we also found that orientation interacted with feature viewed, $F(6, 117.04)=64.39$, $p<.001$, $\eta_p^2=.41$. In upright faces, the left eye, $t(94)=4.03$, $p<.001$, Cohen's $d=0.83$, and the nose, $t(94)=6.37$, $p<.001$, Cohen's $d=1.31$, were viewed more than in inverted faces. Conversely, the hairline, $t(94)=2.93$, $p=.004$, Cohen's $d=0.60$, and the forehead, $t(94)=9.24$, $p<.001$, Cohen's $d=1.91$, were viewed more in inverted faces than upright faces.

To assess the activeness of the scanning behaviour across groups, we compared the average fixation number (with higher fixation numbers potentially indicating more active scanning,

---

[6] Mauchley's test of sphericity was significant, therefore we employed the Greenhouse-Geisser correction (Field, 2010).

Goldinger et al., 2009) and average fixation length (with high length indicating deeper encoding, Bombari et al., 2009) for those under observation and those not under observation. These revealed no significant differences across conditions, $t(93)=1.72$, $p=.089$, Cohen's $d=0.37$ (fixation number mean difference=0.5) and $t(93)=0.62$, $p=.538$, Cohen's $d=0.13$ (fixation length mean difference=9ms).

Discussion

While the results of Experiment 4 indicate that being observed affects the recognition of upright faces, it is not affecting holistic processing as measured by the composite face test, nor the parts-and-wholes task. By affecting only the face-inversion effect, it can be assumed that being observed influences the engagement of configural processing, but not the extent of its use (Maurer et al., 2002). Given that the presentation of inverted faces causes a delay in the 'face-specific' ERP N170 (Rossion, Gauthier, Tarr, Despland, Bruyer, Linotte, & Crommelinck, 2000), it has been considered as a measure of participants engagement of configural processing rather than a direct measure of holistic processing. In other words, we would expect that observation has notable effects on the time-course of face processing without affecting the coding employed - expert processing mechanisms will be employed under observation but later than they would be when not under observation. These results were not the result of anxiety nor fear of negative evaluation as measured by the brief fear of negative evaluation (Leary, 1983) nor classroom anxiety (Nist & Diehl, 1990) nor the evaluation apprehension (Richmond et al., 2001) questionnaires. These results are due to direct effects of observation.

Our eye-tracking results further this interpretation: while the features viewed were dependent on being observed, the way information was extracted from faces did not. Scanning behaviour was not more active nor was more information extracted from each fixation when not being observed compared to being observed. Instead, observation caused participants to look more at the nose and areas of the face (chin and cheeks) that are less diagnostic for face recognition. This suggests that the reason why observation causes participants to be less accurate at face recognition is because they look at the faces in an atypical manner. Of course, these results are somewhat exploratory in nature and do need to be replicated for confirmation.

General Discussion

Across four Experiments, we have shown that face recognition for own-ethnicity upright faces is detrimentally affected by being observed. This is in contrast with the recognition of other-ethnicity faces, inverted faces, and objects which are not affected by such observation. The effect is one of encoding and appears to cause a change in eye movements (focusing more on the nose, hairline, and areas of the face with no facial features). Being observed has no effect on the amount of holistic processing engaged in, nor the amount of heuristic processing engaged in, as measured by eye movements. Here, we shall interpret these results within the theoretical frameworks put forward throughout this article.

In the introduction, we presented drive theory (Zajonc, 1965) as a model to explain these results. If one considers that unfamiliar face recognition, at least in the settings here, is not an easy task (because participants are attempting to recognise a number of previously unfamiliar faces in a laboratory situation), it is sufficiently difficult as to cause social inhibition. We can relate these results to those of Mesout & Valentine (2009) who demonstrated that stress negatively affects face recognition performance. Critically, they too caused stress during encoding, and this caused subsequent recognition deficits to unfamiliar face recognition. It seems that our manipulation caused stress in our participants and this had a detrimental effect on face encoding.

While this explanation seems plausible, we did not find an effect of observation on the recognition of inverted faces nor of objects. Recognising inverted faces and objects is more difficult than recognising upright faces (as indicated by our data in Experiments 2 and 3), therefore drive theory would predict that these would be more detrimentally affected by observation than upright face recognition. While performance was lower in the recognition of these stimuli, it was still above chance and therefore differences across conditions would have been detectable should there have been any (i.e., there were no floor effects). In other words, the current results cannot be explained simply by drive theory. Therefore, we must turn to a different explanation of the results.

We presented the notion that social inhibition may lead to more heuristic style cognitive processing (Lamont et al., 2003). This would cause the recognition of own-ethnicity faces to be recognised in a similar manner to other-ethnicity faces (which are typically processed in a heuristic manner, Levin, 2000). Such an explanation is broadly consistent with the results from Experiment 2. However, the effect of being observed is on hit rate, whereas the OEB is based on differences in false alarm rate, suggesting different mechanisms underlie the two factors. Further, heuristic processing is typically revealed by less active scanning of faces as measured by an eye-tracker (Goldinger et al., 2009). There were no differences in active scanning across participants who were observed compared to those that were not. This, coupled with the difficulties in interpreting how inverted faces are processed in a heuristic manner, leads us to rule out this explanation.

An alternative hypothesis was that observation disrupted the expert face processing systems as these are more susceptible to interference than more generalised cognitive architecture. We demonstrated that the magnitude of the face-inversion effect (Yin, 1969) did indeed depend on whether participants were observed or not. This indicated that there was a disruption to the engagement of configural processing. However, no other measures of holistic processing were affected by observation. This suggests that observation does not affect the amount of expert holistic processing engaged in.

Finally, we must turn to a rather more simple explanation for why face recognition was disrupted by being observed. Observation caused our participants to view faces in a suboptimal manner. When viewing faces, participants typically view the eyes more than any other feature (Althoff & Cohen, 1999). Participants under observation viewed the eyes less than those not under observations. Fixations on the eyes has been correlated with face recognition accuracy (Hills et al., 2013b). Indeed, this explanation can account for the results in all four Experiments. For our participants, own-ethnicity faces would be better recognised following fixations to the eyes. Other-ethnicity faces would not be, therefore observation would have no effect on the recognition of other-ethnicity faces. Inverted faces are typically viewed with less eye fixation than upright faces (Barton et al., 2006), therefore are unlikely to be further affected by being observed. The effect of being observed whilst encoding and recognising faces causes participants to view them in an atypical manner, with fewer fixations to the eyes, and this reduces their recognition accuracy. Indeed, stress invoking activities are known to alter cognitive and perceptual processing in other domains (Hockey, 1997).

This explanation is simple, but we hope it will prompt further work to explore the limits of this explanation. While our exploratory eye tracking results have shown that observation causes participants to fixate less on the eyes, they did not report being anxious in any of our questionnaire measures. Indeed, none of the questionnaire measures impacted on the results suggesting that anxiety was not causing participants to make fewer fixations to the eyes. It might be that an individual difference variable similar to anxiety might be behind this factor, such as self-consciousness or fear of being judged. In any case, such a line of reasoning suggests that different people will be affected by being observed to different degrees.

In the preceding paragraph, we highlighted that potentially a fear of being judged or having an incentive might mediate the current findings akin to evaluation apprehension (Cottrell, et al., 1968). Evaluation apprehension makes very similar predictions to drive theory regarding the effects of being observed. Thereby, we discount a simple explanation of evaluation apprehension causing the detrimental effects of being observed on face recognition. However, if we link arousal to how observation might be affecting participants behaviour, then we can extend this theory. Lambourne and Timporowski (2010) have established that exercise-induced arousal enhanced cognitive performance on speed mental tasks, automatised behaviours, memory storage and retrieval. However, higher test anxiety and arousal causes detriments to performance in academic performance when anxiety is measured by questionnaire (Cassady & Johnson, 2002) or by salivary alpha amylase (Jamieson, Mendes, Blackstock, Schmader, 2010).

This highlights the Yerkes-Dodson (Yerkes & Dodson, 1908) law in the effect that arousal and stress can have on performance: performance is impaired under conditions above or below optimal stress levels (see also Broadbent 1965; Mendl 1999).

One caveat with our study is that the method may not be as ecologically valid as we had hoped, given the novelty of the situation: While many of our students will be familiar with laboratory testing, most will not be familiar with face recognition tasks. Nevertheless, we could test face recognition where participants believed they were being observed (all Experiments) and interacting with the target faces (Experiment 1). Our results also highlight that being observed during the learning of the faces has more of an effect on face recognition than the interaction per se. Nevertheless, a more realistic social pressure situation may produce different results, such as an interview setting. Variations of this task will need to be investigated. As previously mentioned, when created social pressure induces a social facilitation effect studies have mainly employed activity-based tasks, such as playing a musical instrument (Markus, 1978).

There are two important future directions for this work. Firstly, many of the effects in face recognition need to be tested under more socially pressured situations. Secondly, the extent to which social pressure affects participants needs to be explored by including degrees of anxiety or pressure and exploring individual differences in the ability to cope with this pressure. It may be task performance was affected by participants finding the task novel, difficult or even anxiety provoking. Combining this analysis with a series of performance tasks, graded in social difficulty, would also enable a deeper analysis into the relationship between social pressure and its effect on face recognition.

These results have significant implications to face recognition research conducted in the laboratory: If experimenters are in the room and directly observing the participants, the results obtained may differ to if the experimenters are not in the room or obviously not directly observing them. While these effects are small, they provide further evidence for the often subtle experimenter effects that can alter participant behaviour (Rosenthal & Fode, 1963). Such results serve as a warning to experimenters to ensure that they behave in an appropriate and consistent manner across all participants.

These results potentially have important implications for identity checkpoints, CCTV operators, and police practice for eye-witness testimony if the results of these laboratory studies transfer to other paradigms and into other contexts. If those working at identity checkpoints and in CCTV booths believe they are being observed, their abilities to encode faces may be poorer than if they were not being observed. Of course, further work is needed to explore if the effects of this study replicate into these applied settings. Furthermore, standard police practice is to remain with a witness while they are completing a line-up. It may be that such practice affects the recognition of the witness. However, further work is necessary to explore whether these results transfer to a line-up procedure.

In summary, storing and retrieving memories for human faces in social situations is a complex task and the current research study shows how social pressure can reduce the rate of face recognition. We have shown that during encoding, social pressure in the form of being observed causes participants to be less accurate at face recognition tasks involving own-ethnicity upright faces, but not inverted faces, other-ethnicity faces, or objects. This, we suggest, is the result of stress altering eye movements over the faces, leading to poorer encoding.

Acknowledgements

This work was funded by the British Academy Small Grants Scheme to the first author. The authors would like to thank David White and an anonymous reviewer for helpful comments on a previous version of this manuscript. Further, the authors would like to thank Gizem Arabacı for help in constructing the figures.

References

Allport, F. H. (1920). The influence of the group upon association and thought. *Journal of Experimental Psychology, 3*(3), 159-182.

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25(4),* 997-1010.

Arizpe, J., Kravitz, D. J., Yovel, G., & Baker, C. I. (2012). Start position strongly influences fixation patterns during face processing: Difficulties with eye movements as a measure of information use. *PloS one*, *7*(2), e31106.

Barton, J. J. (2008). Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *Journal of Neuropsychology*, *2*(1), 197-225.

Barton J. J. S., Radcliffe, N., Cherkasova, M. V., Edelman, J., & Intriligator, J. M. (2006). Information processing during face recognition: the effects of familiarity, inversion, and morphing on scanning fixations. *Perception, 35*, 1089 – 1105.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, *8*(6), 551-565.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14*(2), 143-177.

Bindemann, M., Scheepers, C., & Burton, A. M. (2009). Viewpoint and center of gravity affect eye movements to human faces. *Journal of Vision*, *9*(2), 7-7.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PlosOne, 3,* e3022.

Bombari, D., Mast, F. W., & Lobmaier, J. S. (2009). Featural, configural, and holistic face-processing strategies evoke different scan patterns. *Perception*, *38*(10), 1508-1521.

Broadhurst, P. L. (1957). Emotionality and the Yerkes-Dodson law. *Journal of experimental psychology*, *54*(5), 345.

Burton, A. M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception, 22* (2), 153 – 176.

Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science*, *10*(4), 482-496.

Conway, A. R., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General*, *123*(4), 354.

Case, R. (1985). *Intellectual development: Birth to adulthood*. Academic Press.

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary educational psychology*, *27*(2), 270-295.

Chajut, E., & Algom, D. (2003). Selective attention improves under stress: implications for theories of social cognition. *Journal of personality and social psychology*, *85*(2), 231.

Chan, J. P. K. & Ryan, J. D. (2012). Holistic representations of internal and external face features are used to support recognition. *Frontiers in Psychology, 3,* 87. doi:10.3389/fpsyg.2012.00087

CNBC (2012). CNBC Wiki. Available at: http://wiki.cnbc.cmu.edu/Objects

Collins, K. A., Westra, H. A., Dozois, D. J., & Stewart, S. H. (2005). The validity of the brief version of the Fear of Negative Evaluation Scale. *Journal of anxiety disorders*, *19*(3), 345-359.

Cottrell, N. B., Wack, D. L., Sekerak, G. J., & Rittle, R. H. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of personality and social psychology, 9*(3), 245.

Daly, S. (1978). Behavioural correlates of social anxiety. *British Journal of Social and Clinical Psychology*, *17*(2), 117-120.

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36*(3), 316-328.

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

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological review*, *66*(3), 183.

Echterhoff, G., & Hirst, W. (2009). Social influence on memory. *Social Psychology, 40* (3), 106–110.

Edelson, M., Sharot, T., Dolan, R. J., & Dudai, Y. (2011). Following the crowd: brain substrates of long-term memory conformity. *Science, 333*(6038), 108-111.

Ellis, H. D. (1981). Theoretical aspects of face recognition. In G.M. Davies, H. D. Ellis, & J. W. Shepherd (Eds.), *Perceiving and Remembering Faces.* London: Academic Press.

Ellis, H. D., Deregowski, J. B., & Shepherd, J. W. (1975). Descriptions of White and Black faces by White and Black subjects. *International Journal of Psychology, 10*(2), 119-123.

Gauthier, I., McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Van Gulick, A. E. (2014). Experience moderates overlap between object and face recognition, suggesting a common ability. *Journal of vision, 14*(8), 7-7.

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition*, *118*(2), 201-210.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia, 45*(1), 32-41.

Goldinger, S. D., He, Y., & Papesh, M. H. (2009). Deficits in cross-race face learning: insights from eye movements and pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1105.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Guo, K. (2012). Holistic gaze strategy to categorize facial expression of varying intensities. *PLoS ONE 7,* e4285. doi:10.1371/journal.pone.0042585

Henry, P. J., & Sears, D. O. (2002). The symbolic racism 2000 scale. *Political Psychology, 23*(2), 253-283.

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013a). Removing the own-race bias in face recognition by attentional shift using fixation crosses to diagnostic features: An eye-tracking study. *Visual Cognition, 21*(7), 876-898.

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013b). First fixations in face processing: The more diagnostic they are the smaller the face-inversion effect. *Acta Psychologica*, *142*(2), 211-219.

Hills, P. J., & Lewis, M. B. (2011). Reducing the own-race bias in face recognition by attentional shift using fixation crosses preceding the lower half of a face. *Visual Cognition*, *19*(3), 313-339.

Hills, P. J. & Lewis, M. B. (2018). The development of face expertise: Evidence for a qualitative change in processing. *Cognitive Development.*

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological psychology*, *45*(1-3), 73-93.

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception, 31*(10), 1221-1240.

Holroyd, K. A., Westbrook, T., Wolf, M., & Badhorn, E. (1978). Performance, cognition, and physiological responding in test anxiety. *Journal of Abnormal Psychology, 87*(4), 442-451.

Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2003). Social phobics do not see eye to eye:: A visual scanpath study of emotional expression processing. *Journal of anxiety disorders*, *17*(1), 33-44.

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology, 43*(2), 334-340.

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological review*, *117*(4), 1168.

Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal of Experimental Social Psychology*, *46*(1), 208-212.

Joseph, R. M., & Tanaka, J. (2003). Holistic and part-based face recognition in children with autism. *Journal of Child Psychology and Psychiatry*, *44*(4), 529-542.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci, 17*(11), 4302-4311.

Lambourne, K., & Tomporowski, P. (2010). The effect of exercise-induced arousal on cognitive task performance: a meta-regression analysis. *Brain research*, *1341*, 12-24.

Leary, M. R. (1983). A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin*, *9*(3), 371-375.

Leder, H., & Carbon, C. C. (2005). When context hinders! Learn–test compatibility in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*(2), 235-250.

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 128*(4), 559-574.

Lewis, M. B. & Hills, P. J. (2018). Perceived race affects configural processing but not holistic processing in the composite-face task. *Frontiers in Psychology, 9*, 1456. doi: 10.3389/fpsyg.2018.01456

Light, L. L., Hollander, S., & Kayra-Stuart, F. (1981). Why attractive people are harder to remember. *Personality and Social Psychology Bulletin*, *7*(2), 269-276.

Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition*, *6*(3), 312-319.

Logie, R. H., Baddeley, A. D., & Woodhead, M. M. (1987). Face recognition, pose and ecological validity. *Applied Cognitive Psychology, 1*(1), 53-69.

Macmillan, N. A., & Creelman, C. D. (2010). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.

Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *The Journal of Abnormal and Social Psychology*, *47*(2), 166.

Marcon, J. L., Meissner, C. A., Frueh, M., Susa, K. J., & MacLin, O. H. (2010). Perceptual identification and the cross-race effect. *Visual Cognition, 18*(5), 767-779.

Marcon, J. L., Susa, K. J., & Meissner, C. A. (2009). Assessing the influence of recollection and familiarity in memory for own- versus other-race faces. *Psychonomic Bulletin & Review, 16*(1), 99-103.

Markus, H. (1978). The effect of mere presence on social facilitation: An unobtrusive test. *Journal of Experimental Social Psychology, 14* (4), 389–397.

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255-260.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational psychology review*, *8*(3), 299-325.

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational psychologist*, *35*(1), 13-23.

Meade, M. L., & Roediger, H. L. (2002). Explorations in the social contagion of memory. *Memory & Cognition, 30*(7), 995–1009.

Megreya, A. M. & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition, 34*(4), 865-876.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: *A meta-analytic review. Psychology, Public Policy, and Law, 7*(1), 3-35.

Meissner, C. A., Brigham, J. C., & Butz, D. A. (2005). Memory for own- and other-race faces: a dual-process approach. *Applied Cognitive Psychology, 19*(5), 545-567.

Mendl, M. (1999). Performing under pressure: stress and cognitive function. *Applied Animal Behaviour Science*, *65*(3), 221-244.

Michel, C., Caldara, R., & Rossion, B. (2006b). Same-race faces are perceived more holistically than other-race faces. *Visual Cognition*, *14*(1), 55-73.

Michel, C., Corneille, O., & Rossion, B. (2007). Race categorization modulates holistic face encoding. *Cognitive Science*, *31*(5), 911-924.

Miellet, S., Vizioli, L., He, L., Zhou, X., & Caldara, R. (2013). Mapping face recognition information use across cultures. *Frontiers in Psychology, 4,* e34. doi: 10.3389/fpsyg.2013.00034

Minear, M. E., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods: Instruments and Computers, 36*(4), 630–633.

Mogg, K. & Bradley, B. P. (2002). Selective orientation to masked threat faces in social anxiety. *Behaviour, Research, and Therapy, 40*9(2)*,* 1403 – 1414.

Mondloch, C. J., & Desjarlais, M. (2010). The function and specificity of sensitivity to cues to facial identity: An individual-differences approach. *Perception*, *39*(6), 819-829.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of cognitive neuroscience*, *9*(5), 555-604.

Mukudi, P. B., & Hills, P. J. (2019). The combined influence of the own-age,-gender, and-ethnicity biases on face recognition. *Acta psychologica*, *194*, 1-6.

Nist, P., & Diehl, M. (1990). PHCC test anxiety questionnaire. *Retrieved Aug*, *20*, 2010.

Pearce, D. (2012). *Roads to commensurability* (Vol. 187). Springer Science & Business Media.

Pessin, J. (1933). The comparative effects of social and mechanical stimulation on memorizing. *The American Journal of Psychology, 45*(2), 263-270.

Reysen, M. B. (2007). The effects of social pressure on false memories. *Memory & Cognition*, *35*(1), 59-65.

Rezlescu, C., Susilo, T., Wilmer, J. B., & Caramazza, A. (2017). The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with varied relationships to face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1961.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Beliefs alter holistic face processing… if response bias is not taken into account. *Journal of Vision*, *11*(13), 17-17.

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin*, *140*(5), 1281.

Richmond, V. P., Wrench, J. S., & Gorham, J.(2001).*Communication, affect, and learning in the classroom.* Acton, MA: Tapestry Press Rosenthal, R., & Fode, K. L. (1963). The effect of experimenter bias on the performance of the albino rat. *Systems Research and Behavioral Science*, *8*(3), 183-189.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica, 128*, 274 – 289. doi:10.1016/j.actpsy.2008.02.003

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, *21*(2), 139-253.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*(1), 69-72.

Sæther, L., Van Belle, W., Laeng, B., Brennen, T., & Øvervoll, M. (2009). Anchoring gaze when categorizing faces' sex: evidence from eye-tracking data. *Vision research*, *49*(23), 2870-2880.

Schneier, F. R., Rodebaugh, T. L., Blanco, C., Lewin, H., & Liebowitz, M. R. (2011). Fear and avoidance of eye contact in social anxiety disorder. *Comprehensive Psychiatry*, *52*(1), 81-87.

Shepherd, J. W., & Deregowski, J. B. (1981). Races and faces - a comparison of the responses of Africans and Europeans to faces of the same and different races. *British Journal of Social Psychology, 20*(2), 125-133.

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law, 7*(1), 36-97.

Swann, W. B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, *91*(4), 457.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly journal of experimental psychology*, *46*(2), 225-245.

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *The Quarterly Journal of Experimental Psychology Section A, 44*(4), 671-703.

Wagenmakers, E. J., Krypotos, A. M., Criss, A. H., & Iverson, G. (2012). On the interpretation of removable interactions: A survey of the field 33 years after Loftus. *Memory & cognition*, *40*(2), 145-160.

Wieser, M. J., Pauli, P., Alpers, G. W., & Mühlberger, A. (2009). Is eye to eye contact really threatening and avoided in social anxiety?—An eye-tracking and psychophysiology study. *Journal of anxiety disorders*, *23*(1), 93-103.

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin, 84*(5), 963-976.

Wright, D. B., London, K., & Waechter, M. (2010). Social anxiety moderates memory conformity in adolescents. *Applied Cognitive Psychology, 24*(7), 1034-1045.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology*, *18*(5), 459-482.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141-145.

Young, A. W. & Burton, A. M. (2018). Are we face experts? *Trends in Cognitive Sciences, 22*, 2, 100 – 110.

Young, A. W., Hellawell, D., & Hay, D. C. (2013). Configurational information in face perception. *Perception*, *42*(11), 1166-1178.

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, *16*(2), 116-142.

Zajonc, R. B. (1965). Social facilitation. *Science, 149*(3681)*,* 269- 274.

Zajonc, R. B., & Sales, S. M. (1966). Social facilitation of dominant and subordinate responses. *Journal of Experimental Social Psychology, 2*(1), 160-168.