Natural variability is essential to learning new faces

Journal:	*Visual Cognition*
Manuscript ID	Draft
Manuscript Type:	Full Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Kramer, Robin; University of York Jenkins, Rob; University of York Young, Andy; University of York Burton, Mike; University of York,
Keywords:	face learning, face recognition, variability

Natural variability is essential to learning new faces

Robin S. S. Kramer, Rob Jenkins, Andrew W. Young & A. Mike Burton

Department of Psychology, University of York

Correspondence to:

Mike Burton, Department of Psychology, University of York, York YO10 5DD, UK

Tel. +44 1904 3231409 Email mike.burton@york.ac.uk

Acknowledgements

We would like to thank Renée Lefebvre for taking part in the initial piloting of the video procedure. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.323262 to AMB, and from the Economic and Social Research Council, UK [ES/J022950/1] to AMB.

Running Head: Natural variability is essential to learning new faces

Abstract

We learn new faces throughout life, for example in everyday settings like watching TV. Recent research has shown that *image variability* is key to this ability: if we learn a new face over highly variable images, we are better able to recognise that person in novel pictures. Here we asked people to watch TV shows they had not seen before, and then tested their ability to recognise the actors. Some participants watched TV shows in the conventional manner, whereas others watched them upside down or contrast-reversed. Image variability is equivalent across these conditions, and yet we observed that viewers were unable to learn the faces upside down or contrast-reversed – even when tested in the same format as learning. We conclude that variability is a necessary, but not sufficient condition for face learning. Instead, mechanisms underlying this process are tuned to extract useful information from variability falling within a critical range that corresponds to natural, everyday variation.

Key words: Face recognition, face learning

Introduction

Our ability to learn new faces is poorly understood. This is perhaps surprising, because there are large differences between perception of familiar and unfamiliar faces (see Johnston & Edmonds, 2009, for a review). Familiarity allows us to generalise over a very large range of conditions, and we can recognise the people we know, even in highly degraded images (Burton, Wilson, Cowan & Bruce, 1999; Clutterbuck & Johnston, 2005). However, unfamiliar face perception is more closely image-bound, and does not generalise well to new pictures of the same person (Bruce et al, 1999; Hancock, Bruce & Burton, 2000).

Early studies of face learning emphasised the duration or frequency of encounters (e.g. Shapiro & Penrod, 1986) and achieved only limited success. In part, this limited success may be attributable to over-reliance on highly standardised images. It is now clear that different images of the same person can look very dissimilar (Adini, Moses & Ullman, 1997; Jenkins, White, van Montfort & Burton, 2011), which limits the usefulness of standardised images for testing face recognition. In everyday life the appearance of a particular face is determined not only by changes in viewpoint and lighting, but also by non-rigid motion arising from speech and emotional expression. Exposure to these changes could be central to the process of face learning because the presence of variability provides the visual system with information about what is constant over different encounters with an individual (Bruce, 1994). In support of this hypothesis, recent experiments have shown that exposure to such variability is a key predictor of learning (Andrews, Jenkins, Cursiter & Burton, 2015; Murphy, Ipser, Gaigg & Cook, 2015; Ritchie & Burton, in press). Exposure to greater variability in images of a particular face leads to a more robust facial representation that can generalise to novel images of that face.

In this study we address the fundamental question of whether image variability *per se* is sufficient to explain everyday face learning. As this issue is bound up with the natural range over which people are encountered, we are concerned not to use artificial lab-based learning settings. We have argued previously that the use of ambient images (Burton, 2013; Jenkins et al, 2011) is critical to understanding face recognition, and so here we ask people to learn new faces in a normal setting: watching TV. Participants watched episodes from a sitcom, which they had never seen prior to the experiment, and were then tested on novel photos of the actors. Under these conditions we expect good face learning. However, we asked some

participants to watch the TV episodes upside down (inverted) or contrast-reversed. These manipulations have no effect on the variability exhibited by the characters – however, they clearly disrupt the typical ways in which faces are encountered in daily life.

Inversion and negation are, of course, well-known to disrupt familiar face recognition (Yin, 1969; Galper, 1970), but these effects are usually measured for faces that are learned in a normal format and then tested in a different format in the laboratory. However, there is some evidence that extensive training (hundreds of exposures) to relatively small numbers of images, can reduce the well-known effect of inversion (Laguesse, Dormal, Biervoye, Kuefner & Rossion, 2012) though this tends to be based on image-specific learning (Hussain, Sekuler & Bennett, 2009; Robbins & McKone, 2003). In our experiments, we were able to expose viewers to exactly the same naturalistic range of faces (TV episodes) in typical or atypical (inverted or contrast-reversed) formats. If variability is the key to face learning, then there should be equivalent learning in each condition – provided recognition tests are consistent with learning format (i.e., a face learned inverted should be well-recognised inverted). Furthermore, one might predict equivalent costs of format change: for example, a face learned inverted might be hard to recognise upright. Alternatively, if the variability that leads to face learning is not a simple statistical summary of exposure history, but is tied more closely to the ways in which variability across the *class* of faces is typically encountered - i.e. to natural everyday variability - then we would expect much better learning when faces are shown in typical format.

Experiment 1

Our first experiment investigated whether people were able to learn the faces of new identities when these identities were experienced inverted. This involves asking participants to watch TV upside down. As we were initially unsure whether this would be a disturbing experience, we first ran a pilot study in which two observers (RSSK plus one other) watched, upside down, several episodes of a TV programme which they had not previously seen. This pilot convinced us that the experience was not an uncomfortable one, and that it was easy to follow the plot and engage with the programme. We therefore proceeded to the formal experiment.

Method

Participants

Eighty students (57 women; age $M = 20.70$ years, $SD = 3.10$; 98.75% self-reported ethnicity as White or White mixed) at the University of York took part in exchange for either course credits or a small payment. The data from two additional participants were excluded because they reported being familiar with one of the actors after viewing the videos.

Stimuli

Two episodes of the US television sitcom 'Cristela' (2014-2015) were selected for the face learning phase of the experiment. This series was chosen because the actors featured were expected to be unfamiliar to UK participants. The total playing time for both episodes was 42 minutes 7 seconds. These particular episodes (Season 1, Episodes 2 and 8) were selected because the eight main characters appeared in one or more scenes in both episodes. No other characters featured other than tangentially.

For each of the eight characters, two different facial photographs were downloaded from a Google Images search and used to test how well each face had been learnt. Note that these test photographs were not images of the actors as they appeared in the show (e.g., screenshots). In addition, two photographs of 'foils' (other people who resembled the actors) were downloaded for each actor using descriptive search terms that matched the actor's general appearance (e.g., 'blonde woman', 'grey haired man'). All images were high quality, colour, unconstrained, naturalistic photographs, and were cropped around the person's head, as in Figure 1. Images were resized to 380 x 570 pixels.

Procedure

Participants watched both episodes of the sitcom, with no delay in between and no commercial breaks, using VLC media player. They wore on-ear headphones throughout and adjusted the volume to their own preferences. The purpose of the experiment was not revealed, and the face recognition task was not mentioned.

Instead, participants were simply instructed to "focus on the social relationships between the characters" while watching the videos, and told that there would be "a short computer task" afterwards. Before watching the videos, participants were asked if they were familiar with either the sitcom or any of the actors who starred in it (their names were presented on a printed sheet), and all responded that they had no prior familiarity.

For half of the participants the episodes were viewed in colour in the normal upright format. For the remaining participants, both episodes were played inverted (rotated by 180 degrees) onscreen, using VLC's video effects options. Allocation of participants to these conditions was made by alternation, i.e. the first participant viewed upright, the second upside down, etc.

After watching the two episodes, participants carried out an old/new decision task on individual face images, presented in random order. In the first test block of 16 trials, participants were shown one image and one foil (both randomly selected from the two pairs collected) for each of the eight actors. These were presented inverted (rotated by 180 degrees – see Figure 1) in a random sequence, and participants were asked onscreen, "Did you see this actor in the TV show?" A Likert scale appeared below each image, with labels ranging from -3 ("definitely no") through 0 ("unsure") to 3 ("definitely yes"). A continuous scale was used, rather than asking for a simple yes/no response, in order to provide more fine-grained information, (e.g. Bonner, Burton & Bruce, 2003). Responses were made using the mouse, with no time limit. In the second block of 16 trials, the remaining images (the other image and foil for each actor) were presented upright onscreen and the same question was posed and responses collected.

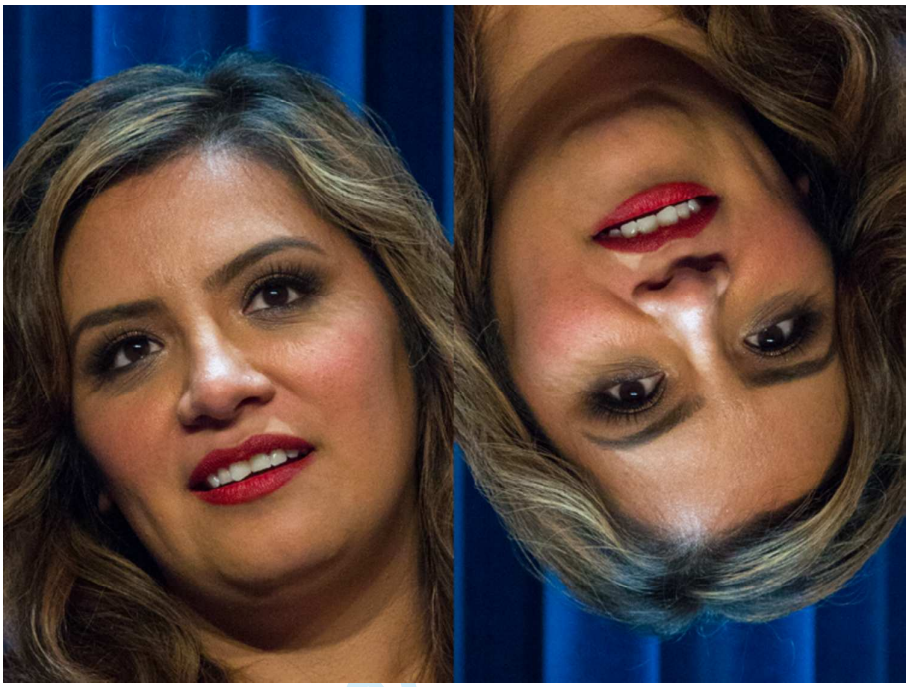

Figure 1. An example test image from Experiment 1, presented either upright (left) or inverted (right). Image attributed to Dominick Dusseault (Own work) [CC BY-SA 2.0].

Throughout the viewing of inverted videos (for participants in that condition), and inverted images during the test, participants were asked not to rotate their heads or try to see the faces upright as this would be considered "cheating".

Results

Although we limited our investigation to the recognition of only the eight core actors, we found that several participants failed to learn and later recognise one of these actors. Across both episodes, it turned out that the actress Justine Lupe appeared onscreen (including in the background of scenes) for a total of only 2 minutes 49 seconds. Considering only recognition of upright images after learning in the upright viewing condition (where accuracy should be at its highest), participants were noticeably worse with Lupe in comparison with the remaining seven actors (see Figure 2). Therefore, recognition trials involving Lupe and her foils were excluded from subsequent analyses.

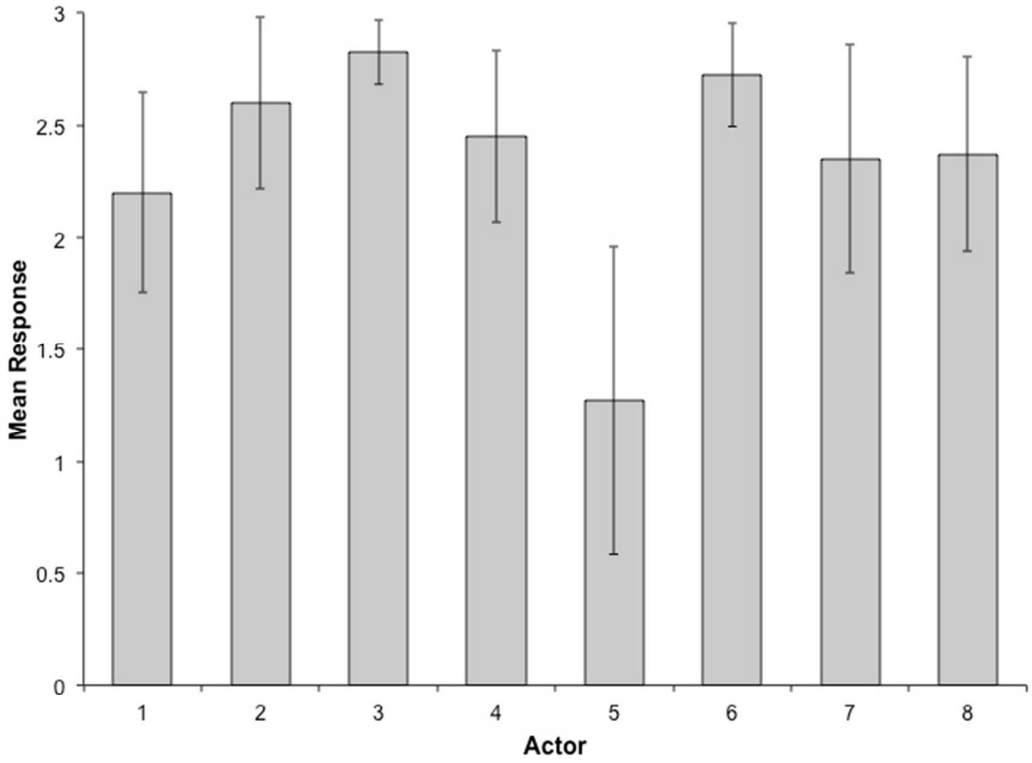

Figure 2. Recognition of each actor's face from upright test images after learning in the upright viewing condition. Performance is expressed in terms of mean ratings on the -3 to +3 Likert scale. Actor 5 is Justine Lupe. Errors bars represent 95% confidence intervals.

In order to determine whether the remaining actors were learned and subsequently recognised, we investigated signal detection measures as follows. We converted our participants' responses (ranging from -3 to +3) to 'absent' (-3 to -1) and 'present' (+1 to +3), and discarded the remaining 'unsure' (0) responses. We then calculated sensitivity indices (*d'*) using: *Hit:* test image is of an actor from the show and participants responded 'present'; *False alarm:* test image is of a foil and participants responded 'present'. Upright and inverted test images were considered separately.

These *d'* values were analysed using a 2 (Learning Condition: upright vs. inverted) x 2 (Test Image: upright vs. inverted) mixed ANOVA. Learning Condition varied between subjects while Test Image varied within subjects. A significant

interaction, $F(1, 78) = 22.88$, $p < .001$, $\eta^2_p = .23$, prompted us to carry out tests for simple main effects (see Figure 3).

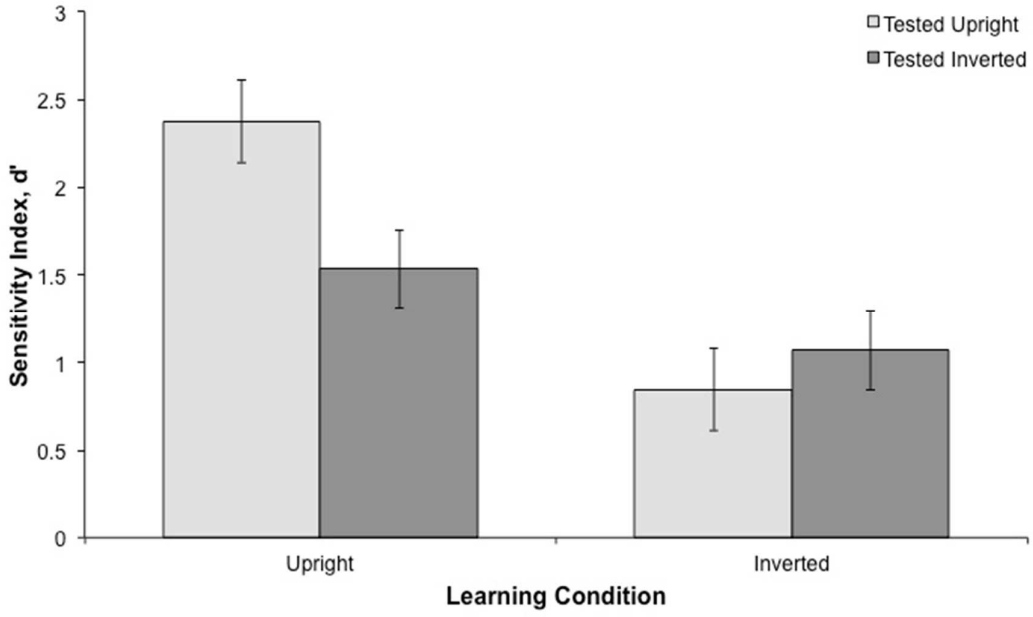

Figure 3. Sensitivity indices (d') for faces learned upright and inverted. Errors bars represent 95% confidence intervals.

After learning faces upright, participants were better at recognising upright (M = 2.37) in comparison with inverted test images (M = 1.53), $F(1, 78) = 28.58$, $p < .001$, $\eta^2_p = .27$, demonstrating the classic inversion effect. In contrast, for those who learned the faces inverted, we found no difference between recognising upright (M = 0.85) and inverted test images (M = 1.07), $F(1, 78) = 2.01$, $p = .160$, $\eta^2_p = .03$.

In addition, participants who learned the faces upright were better at recognising upright test images in comparison with those who learned inverted, $F(1, 156) = 88.08$, $p < .001$, $\eta^2_p = .36$. Interestingly, participants who learned the faces upright were also better at recognising inverted test images in comparison with those who learned inverted, $F(1, 156) = 8.07$, $p = .005$, $\eta^2_p = .05$. However, that particular effect size is small and so may not warrant further interpretation without replication.

Experiment 2

Here, we investigated whether people were able to learn the faces of new identities when these identities were experienced contrast-reversed.

Method

Participants

A new set of 80 students (65 women; age M = 19.76 years, SD = 1.57; 100% self-reported ethnicity as White or White mixed) at the University of York took part in exchange for either course credits or a small payment. The data from one additional participant were excluded because he reported being familiar with one of the actors after viewing the videos.

Stimuli

We used greyscale versions of images and videos used in Experiment 1, presented in either positive or negative contrast.

Procedure

The procedure here was identical to the previous experiment other than the two conditions used. Rather than watching the television episodes upright or inverted, participants watched them in either positive or negative contrast, using VLC's video effects options. Similarly, during the subsequent recognition test, images in the first block were presented in negative contrast, while images in the second block were positive contrast (see Figure 4). All videos and test images were presented upright.

Figure 4. An example test image from Experiment 2, presented either in positive contrast (left) or negative contrast (right). Image attributed to Dominick Dusseault (Own work) [CC BY-SA 2.0].

Results

As in Experiment 1, participants were noticeably worse at learning the face of Justine Lupe (Figure 5). Therefore, recognition trials involving Lupe and her foils were excluded from subsequent analyses.

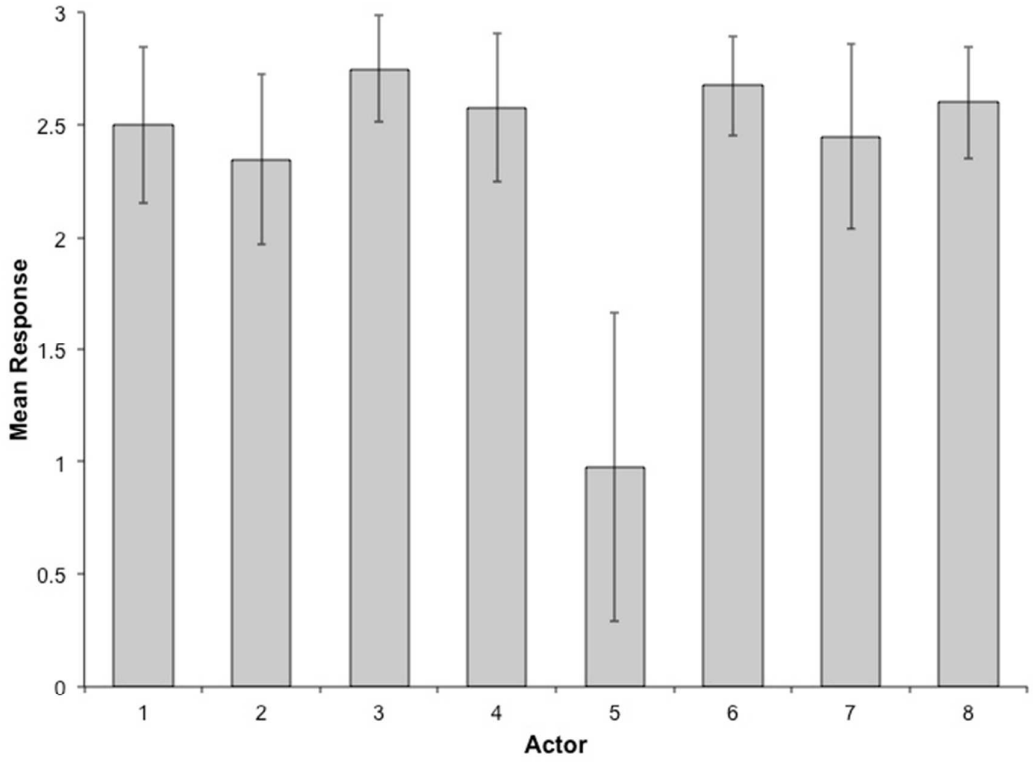

Figure 5. Recognition of each actor's face from positive test images, after learning in the positive viewing condition. Performance is expressed in terms of mean ratings on the -3 to +3 Likert scale. Actor 5 is Justine Lupe. Errors bars represent 95% confidence intervals.

Data were analysed using signal detection measures as in Experiment 1. The *d'* values were analysed using a 2 (Learning Condition: positive vs. negative contrast) x 2 (Test Image: positive vs. negative contrast) mixed ANOVA. Learning Condition varied between subjects while Test Image varied within subjects. A significant interaction, $F(1, 78) = 27.13$, $p < .001$, $\eta^2_p = .26$, prompted us to carry out tests for simple main effects (see Figure 6).

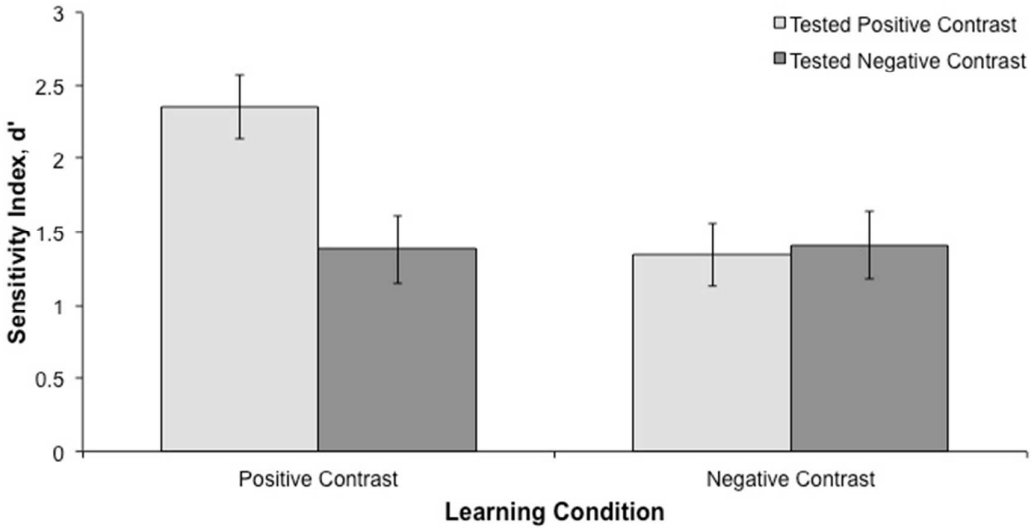

Figure 6. Sensitivity indices (d') for faces learned in positive and negative contrast. Errors bars represent 95% confidence intervals.

After learning faces in positive contrast, participants were better at recognising positive images (M = 2.35) than negative test images (M = 1.38), $F(1, 78) = 47.84$, $p < .001$, $\eta^2_p = .38$, demonstrating the classic contrast-reversal effect. By comparison, those who learned the faces contrast-reversed showed no difference in recognition performance for positive (M = 1.35) and negative test images (M = 1.41), $F(1, 78) = 0.20$, $p = .656$, $\eta^2_p = .00$.

In addition, participants who learned the faces in positive contrast were better at recognising positive test images compared with those who learned the faces in negative contrast, $F(1, 156) = 39.90$, $p < .001$, $\eta^2_p = .20$, whereas participants who learned the faces in positive contrast were no better at recognising negative test images than those who learned negative images, $F(1, 156) = 0.03$, $p = .863$, $\eta^2_p = .00$.

Discussion

First, we note that the paradigm of incidental learning of faces from watching episodes of sitcoms successfully emulates the main characteristic of familiar face recognition. After watching the episodes in normal video format, participants showed good recognition of previously unseen photographs of the characters' faces and correct rejection of similar-looking distractors (see Figures 3 and 6). This level of

performance is actually rarely achieved in laboratory studies using such a stringent test of face familiarisation.

In both experiments, participants showed little ability to learn faces presented in atypical formats that involved watching inverted or contrast-reversed videos. Even when recognition of the faces was tested with images in the same format as the watched videos, it remained poor. Indeed face learning from these atypical formats was so poor that there was no reliable effect of switching format between the studied videos and the test images. In comparison, recognition of the faces learnt from normal videos was substantially impaired if the test images were inverted or contrast-reversed, showing the well-documented susceptibility of familiar face recognition to these transforms.

Inversion and contrast-reversal are transforms that have no impact on the underlying physical variability of the images themselves, yet create images of a type with which our visual systems have little experience. But in standard accounts of face recognition, each transform is thought to have a different locus. Inversion is usually thought to affect holistic or configural encoding, (e.g. Maurer, Le Grand & Mondloch, 2002; Young, Hellawell & Hay, 1987). The effect of contrast-reversal has been more debated, but opinion is converging on the view that it impairs the encoding of critical texture patterns (Bruce & Langton, 1994; Gilad, Meng & Sinha, 2009; Russell, Biederman, Nederhouser & Sinha, 2007; Sormaz, Andrews & Young, 2013). Despite these putative differences between inversion and contrast-reversal, their impact on face learning proved equivalent. The key variable here seems to be whether or not the format of the studied videos was itself unusual.

In essence, our findings show that whilst normal image variability is useful to learning new faces, the unusual types of variability associated with atypical image formats are not. This has important implications for understanding how our visual systems make use of variability to learn new faces. For example, it has been noted that averaged representations offer a way of eliminating irrelevant image variations to arrive at the stable characteristics of a particular face (Burton, Jenkins, Hancock & White, 2005), and averaged images work well in a computer face recognition context (Jenkins & Burton, 2008). Here, we have shown that if some form of averaging is the mechanism that the brain uses to learn faces, it cannot apply this easily to faces presented in unusual formats.

An obvious hypothesis as to why this might happen is that inverted or contrast-reversed faces are not sufficiently 'face-like'. However, inspection of Figures 1 and 4 suggests it is not hard to see that the inverted and reversed images depict faces. We think instead that the explanation is to be found in phenomena such as the narrowing of perceptual mechanisms in favour of faces as encountered in everyday life (Maurer & Werker, 2014; Pascalis, de Haan & Nelson, 2002; Pascalis et al, 2005). Tuning of processing mechanisms leading to perceptual narrowing is of course famously noted for language, where as adults we find it hard to separate meaningful from meaningless variability in the sounds of a foreign tongue (Werker, Gilbert, Humphrey & Tees, 1981). Much the same point may apply to face recognition, where inverted or contrast-reversed faces simply do not fit the mechanisms we have created for finding a consistent representation. Heroic studies have shown that participants can learn *something* from inverted or contrast-reversed faces if given a lot of practice (Larguesse et al, 2012; Hussain et al, 2009), but this is probably no different from the observation that, given sufficient practice, we can also learn to discriminate foreign sounds.

Our conclusion is that exposure to sufficient variability is a key factor in learning faces, but to be fully effective this variability has to be of a type to which our visual systems have become accustomed.

References

Adini, Y., Moses, Y., & Ullman, S. (1997). Face recognition: The problem of compensating for changes in illumination direction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *19*(7), 721–732.

Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *Quarterly Journal of Experimental Psychology*, *68*(10), 2041–2050.

Bonner, L., Burton, A. M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, *10*(5), 527–536.

Bruce, V. (1994). Stability from variation: The case of face recognition. The M.D. Vernon Memorial Lecture. *The Quarterly Journal of Experimental Psychology. A*, *47*(1), 5–28.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, *5*(4), 339–360.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces, *Perception, 23,* 803 – 822.

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, *66(8)*,1467-1485.

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, *51*(3), 256–284.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*(3), 243–248.

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology, 17(1),* 97–116.

Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science*, *19*(4), 207–208.

Gilad, S., Meng, M., & Sinha, P. (2009). Role of ordinal contrast relationships in face encoding. *Proceedings of the National Academy of Sciences, USA, 106*, 5353-5358.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4(9),* 330–337.

Hussain, Z., Sekuler, A. B., & Bennett, P. J. (2009). Perceptual learning modifies inversion effects for faces and textures. *Vision Research*, *49*(18), 2273–84.

Jenkins, R., & Burton, A. M. (2008). 100% accuracy in automatic face recognition. *Science*, *319*(5862), 435.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577-596.

Laguesse, R., Dormal, G., Biervoye, a., Kuefner, D., & Rossion, B. (2012). Extensive visual training in adulthood significantly reduces the face inversion effect. *Journal of Vision*, *12*(10), 14–14.

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

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178.

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). Exemplar variance supports robust learning of facial identity. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 577–581.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life?. *Science*, *296*(5571), 1321-1323.

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(14), 5297-5300.

Ritchie, K.L. & Burton, A.M. (in press). Learning faces from variability. *Quarterly Journal of Experimental Psychology.*

Robbins, R. A., & McKone, E. (2003). Can holistic processing be learned for inverted faces? *Cognition*, *88*(1), 79–107.

Russell, R., Biederman, I., Nederhouser, M., & Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. *Vision Research*, *47*(2), 157–165.

Shaprio, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139-256.

Sormaz, M., Andrews, T. J., & Young, A. W. (2013). Contrast negation and the importance of the eye region for holistic representations of facial identity. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 1667-1677.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63.

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

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747–759.