Familiarity and within-person facial variability: The importance of the internal and external features

Robin SS Kramer

Department of Psychology, Trent University, Canada; Department of Psychology, University of York, UK

Zoi Manesi

Department of Experimental and Applied Psychology, VU Amsterdam, the Netherlands

Alice Towler

Department of Psychology, University of York, UK

Michael G Reynolds

Department of Psychology, Trent University, Canada

A Mike Burton

Department of Psychology, University of York, UK

Corresponding author:

Robin Kramer, Department of Psychology, Trent University, Peterborough ON, K9J 7B8, Canada.

Email: remarknibor@gmail.com

Abstract

As faces become familiar, we come to rely more on their internal features for recognition and matching tasks. Here, we assess whether this same pattern is also observed for a card sorting task. Participants sorted photos showing either the full face, only the internal features, or only the external features into multiple piles, one pile per identity. In Experiments 1 and 2, we showed the standard advantage for familiar faces – sorting was more accurate and showed very few errors in comparison with unfamiliar faces. However, for both familiar and unfamiliar faces, sorting was less accurate for external features, and equivalent for internal and full faces. In Experiment 3, we asked whether external features can *ever* be used to make an accurate sort. Using familiar faces and instructions on the number of identities present, we nevertheless found worse performance for the external in comparison with the internal features, suggesting that less identity information was available in the former. Taken together, we show that full faces and internal features are similarly informative with regard to identity. In comparison, external features contain less identity information and produce worse card sorting performance. This research extends current thinking on the shift in focus, both in attention and importance, towards the internal features and away from the external features as familiarity with a face increases.

Keywords

unfamiliar faces, card sorting, within-person variability, internal features, external features, familiarity

Introduction

Research in face recognition is increasingly focussed on the distinction between familiar and unfamiliar faces (Burton, 2013). Face matching is one task that highlights the stark contrast between these two face categories (or more accurately, two ends of a continuum of familiarity; e.g., Clutterbuck & Johnston, 2002). While highly accurate for familiar faces, even under relatively challenging conditions (Bruce, Henderson, Newman, & Burton, 2001; Burton, Wilson, Cowan, & Bruce, 1999), face matching for unfamiliar faces is significantly more difficult and error-prone (Bruce et al., 2001; Bruce et al., 1999; Henderson, Bruce, & Burton, 2001; Megreya & Burton, 2006, 2008). Although the gradual change in face processing that comes with increased familiarity is still not well understood, evidence suggests a shift in the importance we place on the internal and external facial features. Here, we investigate this component of familiarity using a card sorting task (Jenkins, White, Van Montfort, & Burton, 2011).

In early research, Ellis and colleagues found that the internal features of the face (eyes, nose, mouth) proved more important than the external features (hair, facial outline, etc.) when recognising familiar faces. In contrast, internal and external features were equally important for unfamiliar face recognition (Ellis, Shepherd, & Davies, 1979). Building upon this finding by considering face matching, there is now a growing body of evidence that matching faces using internal facial features is performed faster and more accurately for familiar in comparison with unfamiliar (or less familiar) faces. However, researchers typically find no effect of familiarity when matching using external features (Clutterbuck & Johnston, 2002, 2004, 2005; Young, Hay, McWeeny, Flude, & Ellis, 1985; for a review, see Johnston & Edmonds, 2009). Further, this internal feature advantage for familiar faces seems to emerge later in childhood, with young children instead demonstrating a benefit for the external features (Bonner & Burton, 2004; Campbell, Walker, & Baron-Cohen, 1995).

Familiarity is graded, rather than being an "all or none" dichotomous variable, and so we should expect to see more efficient processing of the internal features as we gradually become more familiar with a face. Several studies have found this to be the case (Bonner, Burton, & Bruce, 2003; Clutterbuck & Johnston, 2002, 2004, 2005; Osborne & Stevenage, 2008), supporting both the notion that we process familiar and unfamiliar faces differently, and that this shift in familiarity is a continuous one that can be tracked using indirect measures like face matching performance.

If people show a processing advantage for the internal features of familiar (in comparison with unfamiliar) faces, can unfamiliar face matching be improved by directing participants towards the internal features of the face? Recently, researchers have found benefits for unfamiliar faces when displaying only the internal features of the two faces during face matching (Kemp, Caon, Howard, & Brooks, 2016). However, this advantage was limited to the most difficult trials only, and so it remains unclear how robust this finding is, or whether this effect would generalise to a different task.

In the current work, we investigate whether previous findings with face matching and recognition generalise to a card sorting task. Recent work with card sorting has shown that it is useful for examining the significant shift in behaviour that comes with increased familiarity with specific faces. In a typical experiment, participants are presented with a set of cards, each depicting a different face photograph, and they are instructed to sort the cards into piles, creating one pile for each identity. Researchers find striking effects of familiarity on sorting behavior. For two sets comprising two identities, with twenty photos each, unfamiliar participants typically produce around seven to nine separate piles. However, when familiar with the identities, cards are correctly sorted into two piles (Jenkins et al., 2011). Importantly, all the necessary pictorial information is present in the cards, with unfamiliar participants often performing perfectly when informed that there are only two identities to sort (Andrews, Jenkins, Cursiter, & Burton, 2015).

If the card sorting task utilises the same processes used in face recognition and matching tasks then it should show similar reliance on external features for familiar and unfamiliar faces, but greater reliance on internal features for familiar compared to unfamiliar faces. We also wanted to assess whether there would be similar performance for the internal and external features, with full faces (where both sources of information are present) resulting in higher performance. For participants who are familiar with the faces, we might predict that performance with the internal features will be similar to full face accuracy (given the shift in focus that comes with familiarity), with the external features resulting in notably worse performance than these two conditions. In contrast, for unfamiliar faces, performance with the external features may be similar to, or even surpass, sorting with the internal features.

Experiment 1

In this first experiment, participants were given a card sorting task, where images were sorted into piles, one pile for each identity. Three different types of cards were used – full (complete face photographs), internal (cropped to show only the internal features), and external (cropped to show only the external features). We investigated the behaviours both of participants who were familiar with the two featured identities and those who were unfamiliar in order to compare familiarity across the three types of cards. Based on prior research, we expected to see a difference in the pattern of behaviour across the types of cards for those who were familiar with the faces versus those who were unfamiliar.

Methods

Participants

For our ‘familiar’ condition, 60 students (39 women; age *M* = 21.93 years, *SD* = 5.56; age missing for one participant) at VU Amsterdam, the Netherlands, volunteered to take part in the experiment and received chocolate biscuits as compensation. All participants in the three experiments presented in this article provided either written or verbal informed consent, and were given either a written or verbal debriefing (or both) at the end of the experiment. Regarding our sample size, we continued to collect data until we had 20 usable participants (based on Jenkins et al., 2011) for each of the three card types. Data from an additional 22 participants were discarded because they reported being unfamiliar with either one or both of the identities.

For our ‘unfamiliar’ condition, 60 students (32 women; age *M* = 20.12 years, *SD* = 1.09) at the University of York, UK, volunteered to take part in the experiment and received chocolate biscuits as compensation. All participants reported being unfamiliar with the identities. For unfamiliar card sorting in particular, previous research has shown a difference in behaviours when cards depict other-race faces (Laurence, Zhou, & Mondloch, 2016; Zhou & Mondloch, 2016). As such, care was taken with this sample to test only self-reported White participants (the same race as the depicted identities).

The University of York's psychology department ethics committee approved the experiments, which were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Stimuli

Following Jenkins and colleagues (2011), we downloaded 20 images of each of two Dutch female celebrities (Chantal Janzen and Bridget Maasland) using Google Images searches. Both identities are well known in the Netherlands but unfamiliar to the majority of people living in other countries. All images were high quality, showed the face in approximately frontal view, and were free from occlusions. Images were cropped so that the face filled most of the frame, and were then printed on to laminated cards measuring 70 x 44 mm. All images were presented in full colour.

These 40 'full' images were also cropped using Adobe Photoshop CS4 software to produce 'internal' and 'external' card sets, which showed only the internal or external features respectively. Examples of the image manipulation are shown in Figure 1.

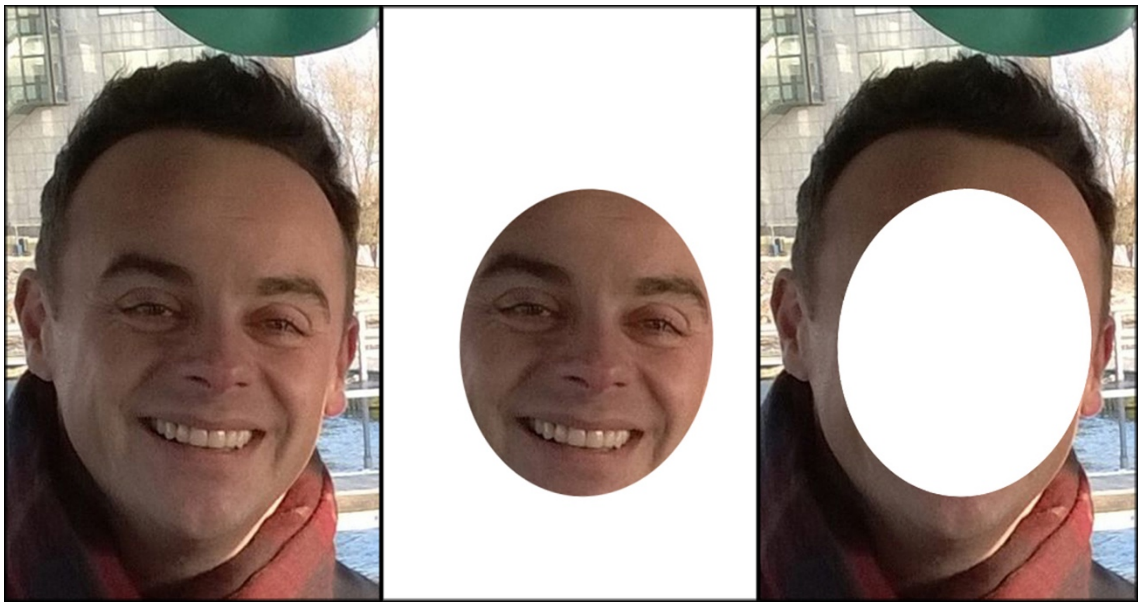

Figure 1. Example image of Anthony McPartlin (used in Experiment 2), showing 'full' (left), 'internal' (middle), and 'external' (right) card types. Image attributed to Ben Salter (Own work) [CC BY 2.0].

Procedure

Each participant was given one set of 40 cards to sort, showing full face, internal or external features. Allocation to conditions was determined by the order in which they took part in the experiment, cycling participants who satisfied the familiarity criteria through conditions. Cards were shuffled beforehand, and participants were instructed to sort them into piles – one pile per identity. We explained that they would not be told how many different identities appeared in the set, and that they were free to make as many or as few piles as they wanted. The task was self-paced.

Upon completion, participants were shown a printed sheet with one ‘full’ image of each identity, and were asked if they were familiar with either of the two celebrities depicted. If yes, they were then asked to write down the names or some other identifying information to prove that they knew who the two people were. We required that our participants were able to identify both the celebrities for the familiar sample, and neither for the unfamiliar sample.

All instructions were given in spoken English for both samples. In addition, participants in the Netherlands were provided with written instructions and debriefing information in both English and Dutch.

Results

Typically, analyses of card sorting data focus on the number of piles created by participants (Jenkins et al., 2011; Zhou & Mondloch, 2016). However, more fine-grained analyses are possible after calculating the proportions of ‘different person/same group’ errors (grouping images of different people in the same pile) and ‘same person/different group’ errors (separating images of the same person into different piles; Balas & Saville, 2017). These two types of error are the result of participants mistakenly sorting images of the two identities into the same pile, as well as images of the same identity into different piles. Using these two types of error, sensitivity indices (*d'*) can be calculated using the conventional formula [*z*(Hits) – *z*(False alarms)] and the following definitions: *Hits* = 1 - 'different person/same group' error rate; *False alarms* = 'same person/different group' error rate (Balas & Pearson, 2017).

The data for this experiment are summarised in Figure 2.

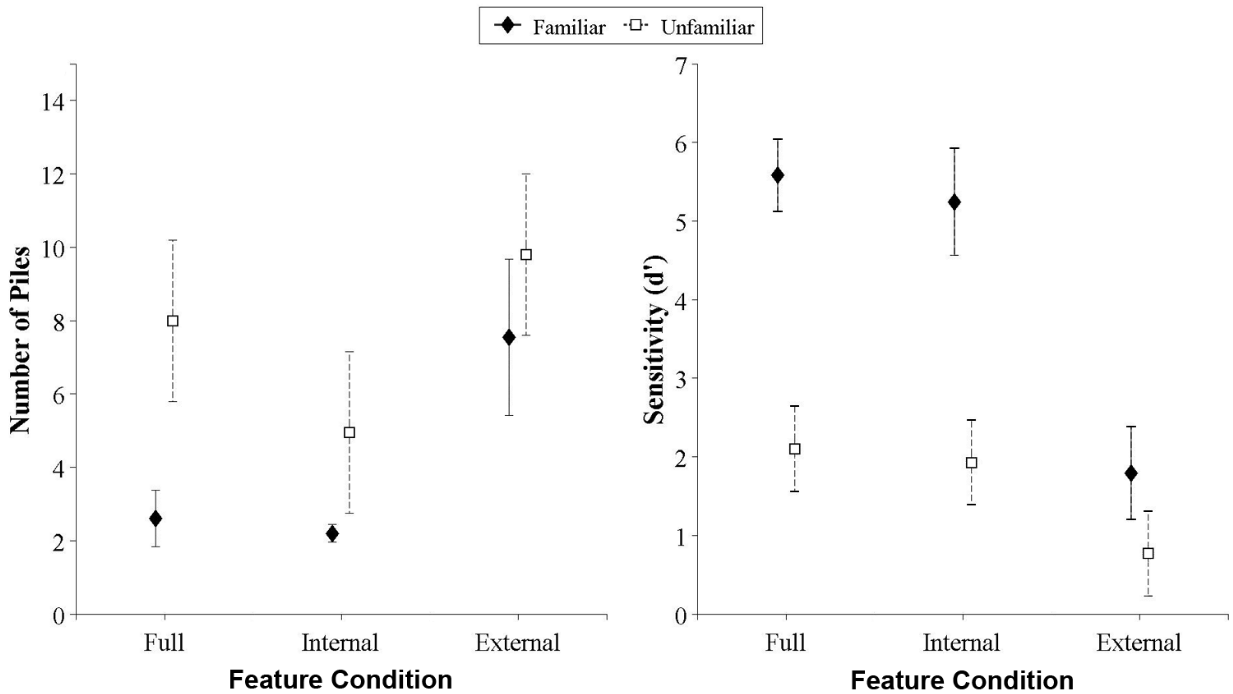

Figure 2. A summary of the data from Experiment 1. Error bars represent 95% confidence intervals.

Data for the number of piles formed and sensitivity indices were analysed separately using 2 x 3 between-subjects analyses of variance (ANOVA), where both Familiarity (familiar, unfamiliar) and Feature Condition (full, internal, external) varied between participants. All pairwise comparisons were Dunn-Šidák corrected.

For the number of piles, we found a significant main effect of Familiarity, $F(1, 114) = 22.50$, $p < .001$, $\eta^2_p = .17$, with fewer piles for familiar faces ($M = 4.12$, $SD = 3.69$) compared with unfamiliar faces ($M = 7.58$, $SD = 5.24$). We also found a significant main effect of Feature Condition, $F(2, 114) = 16.80$, $p < .001$, $\eta^2_p = .23$. Pairwise comparisons showed that more piles were made in the 'external' condition ($M = 8.68$, $SD = 5.33$) compared with both 'full' ($M = 5.30$, $SD = 4.77$) and 'internal' conditions ($M = 3.57$, $SD = 2.53$) (both *p*s < .001). However, the number of piles for 'full' and 'internal' conditions did not differ ($p = .160$). The Familiarity x Feature Condition interaction was not statistically significant, $F(2, 114) = 1.79$, $p = .172$, $\eta^2_p = .03$.

For sensitivity indices (*d'*), the Familiarity x Feature Condition interaction was statistically significant, $F(2, 114) = 12.57$, $p < .001$, $\eta^2_p = .18$. We therefore considered the simple main effects of Feature Condition at each level of Familiarity. For the familiar faces, we found a simple main effect of Feature Condition, $F(2, 114) = 58.55$, $p < .001$, $\eta^2_p = .51$. Pairwise comparisons showed that lower sensitivity was found for the 'external' feature condition ($M = 1.80$, $SD = 1.26$) compared with both 'full' ($M = 5.59$, $SD = 0.98$) and 'internal' feature conditions ($M = 5.24$, $SD = 1.45$) (both *p*s < .001). However, sensitivity for 'full' and 'internal' feature conditions did not differ ($p = .762$). For unfamiliar faces, we found the same pattern of results. There was a simple main effect of Feature Condition, *F*(2,

114) = 6.98, p = .001, η^2_p = .11. Pairwise comparisons showed that lower sensitivity was found for the 'external' feature condition (M = 0.77, SD = 0.46) compared with both 'full' (M = 2.10, SD = 0.89) and 'internal' feature conditions (M = 1.93, SD = 1.83) (both ps ≤ .01). However, sensitivity for 'full' and 'internal' feature conditions did not differ (p = .958). The interaction we found was driven by the larger decrease for 'external features' sensitivity for familiar in comparison with unfamiliar faces.

Discussion

We found no difference between the 'full' and 'internal' feature conditions across the two measures of performance, whereas the 'external' feature condition led to worse performance in both cases. That we see the same pattern of results for both familiar and unfamiliar faces (simply with higher levels of performance for familiars) is surprising, given that previous work using recognition and face matching tasks has repeatedly shown that an internal advantage (in comparison with the external features) is found only for familiar faces.

To consider our initial questions, we find no evidence that external features produce greater sensitivity (M = 0.77) than internal features (M = 1.93) for unfamiliar faces. In fact, the opposite was true. We also expected to see similar levels of sensitivity for the external features for unfamiliar ($M = 0.77$) in comparison with familiar faces ($M = 1.80$). Again, we found a different pattern in our data. We also see no suggestion that presenting unfamiliar faces with only internal features can lead to benefits above full faces, as suggested by recent work with face matching by Kemp and colleagues (2016). However, we do find support for our prediction that familiar faces show similar performance for full and internal faces, and worse performance with external features. As mentioned, what is surprising is that this same pattern is seen for unfamiliar faces.

Experiment 2

Given that the results found in Experiment 1 appeared to contradict some well-established findings in the literature from recognition and matching tasks, we decided to perform a full replication using two new identities and new participant samples. We hoped to confirm that this pattern was not simply the result of the particular images or identities chosen in the first experiment. As such, our revised prediction is that performance on full and internal conditions will be comparable, with worse accuracy for external features. Importantly, this pattern will be evident for both familiar and unfamiliar faces.

Methods

Participants

For our ‘familiar’ condition, 60 students (52 women; age $M = 19.72$ years, $SD = 2.01$) at the University of York, UK, volunteered to take part in the experiment and received course credits as compensation. Data from an additional 10 participants were discarded because they reported being unfamiliar with either one or both of the identities. (There was no overlap between this sample and the UK-based sample used in Experiment 1.)

For our ‘unfamiliar’ condition, 60 students (58 women; age $M = 18.87$ years, $SD = 1.55$) at Trent University, Canada, volunteered to take part in the experiment and received course credits as compensation. Data from an additional 31 participants were discarded because they reported being familiar with either one or both of the identities, or because they self-reported as being an ethnicity other than White. As in Experiment 1, care was taken with this ‘unfamiliar’ sample to test only self-reported White participants (the same race as the depicted identities).

The University of York and Trent University’s psychology department ethics committees approved the experiments, which were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Stimuli

Images were collected, and stimuli created, in the same way as in Experiment 1. The only difference is that here, we used two British male celebrities (Anthony McPartlin and Declan Donnelly). Both identities are well known in the UK but unfamiliar to the majority of people living in other countries.

Procedure

The procedure was identical to Experiment 1.

Results

The data for this experiment are summarised in Figure 3.

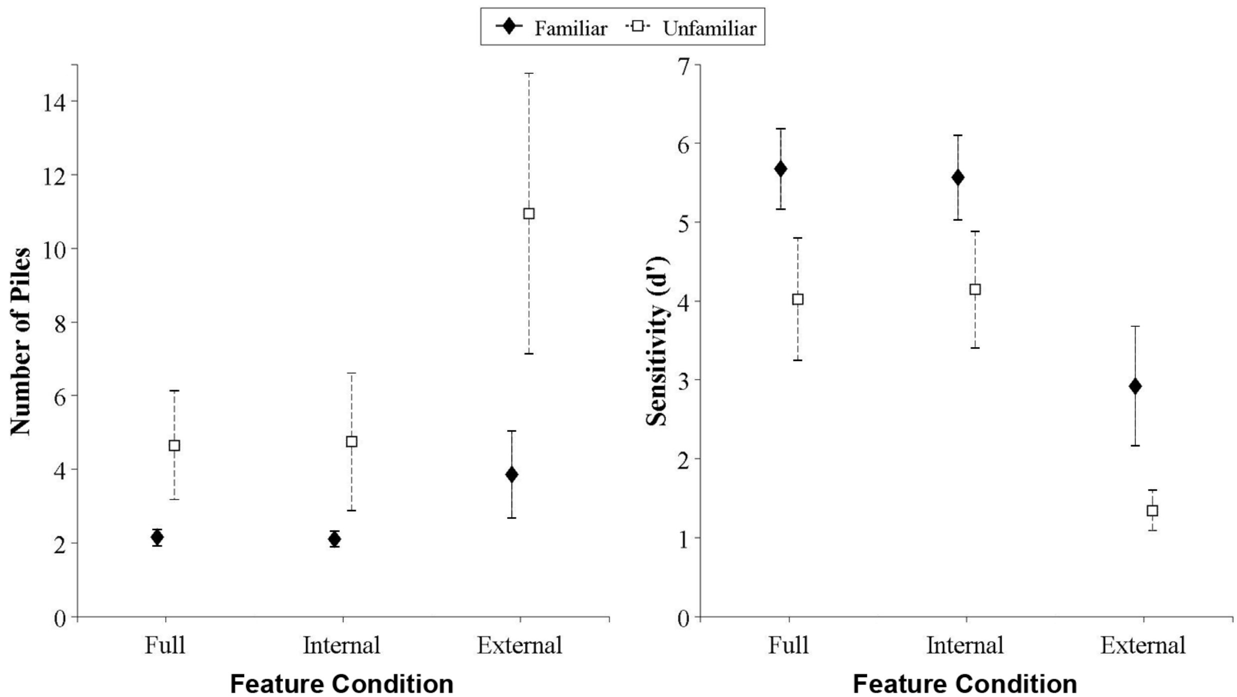

Figure 3. A summary of the data from Experiment 2. Error bars represent 95% confidence intervals.

Data were analysed using the same approach as in Experiment 1.

For the number of piles, the Familiarity x Feature Condition interaction was statistically significant, $F(2, 114) = 4.13$, $p = .018$, $\eta^2_p = .07$. We therefore considered the simple main effects of Feature Condition at each level of Familiarity. For familiar faces, we found no simple main effect of Feature Condition, $F(2, 114) = 1.20$, $p = .305$, $\eta^2_p = .02$. In contrast, for unfamiliar faces, we found a simple main effect of Feature Condition, $F(2, 114) = 15.76$, $p < .001$, $\eta^2_p = .22$. Pairwise comparisons showed that more piles were made in the 'external' feature condition ($M = 10.95$, $SD = 8.15$) compared with both the 'full' ($M = 4.65$, $SD = 3.17$) and 'internal' feature conditions ($M = 4.75$, $SD = 4.00$) (both *p*s $< .001$). However, the number of piles for 'full' and 'internal' feature conditions did not differ ($p > .999$).

For sensitivity indices (*d'*), we found a significant main effect of Familiarity, $F(1, 114) = 40.48$, $p < .001$, $\eta^2_p = .26$, with familiar faces ($M = 4.72$, $SD = 1.82$) showing higher sensitivity than unfamiliar faces ($M = 3.17$, $SD = 1.87$). We also found a significant main effect of Feature Condition, $F(2, 114) = 55.40$, $p < .001$, $\eta^2_p = .49$. Pairwise comparisons showed a lower sensitivity in the 'external' feature condition ($M = 2.13$, $SD = 1.44$) compared with both 'full' ($M = 4.85$, $SD = 1.62$) and 'internal' feature conditions ($M = 4.86$, $SD = 1.54$) (both *p*s $< .001$). However, the sensitivity for 'full' and 'internal' feature conditions did not differ ($p > .999$). The Familiarity x Feature Condition interaction was not statistically significant, $F(2, 114) = 0.08$, $p = .926$, $\eta^2_p = .00$.

Discussion

Replicating the findings of Experiment 1, we see that participants showed similar levels of performance (as measured by sensitivity) for full and internal faces, and worse performance for external features. Importantly, this pattern was seen for both familiar and unfamiliar faces.

Here, we find that the number of piles created did not differ across feature conditions for familiar faces. In Experiment 1, we found that more piles were produced when only the external features were provided, a trend which did not reach significance in Experiment 2. Although the reason for this difference between experiments is uncertain, it may be due to more variation across the external features in the first experiment (e.g., greater changes in background and the women's hairstyles/colours). This result highlights why considering only the number of piles (a relatively coarse measure) may obscure underlying behaviours, while comparing sensitivity across feature conditions allows a more fine-grained analysis of performance.

In line with Experiment 1, we again find that the external features produce lower sensitivity (*M* = 1.34) than internal features (*M* = 4.14) for unfamiliar faces. We also found lower sensitivity for the external features for unfamiliar (*M* = 1.34) in comparison with familiar faces (*M* = 2.92). Finally, we saw no suggestion that presenting unfamiliar faces with only internal features can lead to benefits above full faces (cf. Kemp et al., 2016).

Given that our results fully replicated in this second experiment, we can say with confidence that these findings, which appear to contradict those of previous work, are due to the nature of the task (card sorting rather than recognition/matching) rather than the particular images or participants used.

Experiment 3

Experiments 1 and 2 found that sensitivity in the full and internal feature conditions did not differ, whereas worse sensitivity was found with the external features. This was true for both familiar and unfamiliar faces. Importantly, and perhaps surprisingly, unfamiliar face sensitivity was *lower* with the external in comparison with the internal features. This raises an interesting question – is the information required to accurately sort identities even present in the external features?

Note that in the experiments presented here, the external features included the face outline, ears, hairstyles, and some clothing and background information (see Figure 1). The definition of 'external features', however, is not set, and researchers have sometimes chosen to remove both clothing and backgrounds (e.g., Clutterbuck & Johnston, 2002, 2004, 2005).

To address this question of the informational content present in the external features, we asked participants who were familiar with the two identities to carry out a card sort as before, but this time informing them that exactly two identities were present in the set. Previous research has shown that for unfamiliar faces, this “two-sort” condition (with ‘full’ images) typically results in perfect or almost perfect performance (Andrews et al., 2015). Therefore, for familiar faces, sorting ‘full’ images should be error-free since the identities are familiar and participants know to create exactly two piles.

Here, we focussed on the ‘internal’ and ‘external’ feature conditions since we already know (from previous research, and Experiments 1 and 2) that the ‘full’ images provide sufficient information for perfect card sorting for familiar faces in a “free sort” task, and even unfamiliar faces can be sorted perfectly in a “two-sort” task. If familiar face sorting shows worse performance in the external feature condition, even after participants are told that two identities are present, then this would be strong evidence that the external features simply provide less identity information than the internal features. This is an important point since (ideally) information content should dictate which features are used to sort unfamiliar and familiar faces.

Methods

Participants

Forty students (30 women; age $M = 21.75$ years, $SD = 6.43$) at the University of York, UK, volunteered to take part in the experiment and received course credits or payment as compensation. Data from an additional 10 participants were discarded because they reported being unfamiliar with either one or both of the identities. (There was no overlap between this sample and the UK-based sample used in Experiments 1 and 2.)

The University of York's psychology department ethics committees approved the experiment, which were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Stimuli

The 'internal' and 'external' images presented in Experiment 2 (depicting the two British male celebrities) were used here.

Procedure

The procedure was identical to Experiments 1 and 2, with one important difference. Here, participants were informed before carrying out the task that two identities featured in the set of cards. This “two-sort” version of the task has been used in previous research (Andrews, Burton, Schweinberger, & Wiese, 2017; Andrews et al., 2015).

Results

Participants produced two piles in all cases since this is a requirement of the “two-sort” task. As such, the number of piles was not analysed here. For sensitivity indices (*d’*), we found a significant difference between the ‘internal’ ($M = 5.86$, $SD = 0.77$) and ‘external’ ($M = 4.45$, $SD = 1.90$) feature conditions, $t(38) = 3.07$, $p = .004$, $d = 0.97$.

Discussion

Here, all participants were informed that only two identities appeared in the set of cards, and were familiar with both celebrities. We found better performance when only the internal features were available in comparison with the external features. This result shows that less information regarding identity is available in the external features.

General Discussion

In three experiments, we investigated how familiarity affects card sorting behaviour when participants are presented with full faces or are limited to only the internal or external features. Across Experiments 1 and 2, we find two general results: 1) higher sensitivity was found when sorting familiar faces compared with unfamiliar faces; and 2) sensitivity is worse when only the external features are provided, while no difference is found between full faces and internal features.

That familiarity improves performance in card sorting is no surprise (Jenkins et al., 2011). Participants who were familiar with the two identities depicted on the cards simply recognised the face on each card and sorted with few or no errors. This was the case when both the full face and just the internal features were provided. However, performance (as measured by sensitivity) was significantly worse, even for familiar participants, when only the external features were shown. During debriefing, it was clear that participants sometimes failed to recognise one or both identities from the external features alone, resulting in poorer sorting.

In contrast with previous findings in face matching and recognition, our results demonstrated higher sensitivity for the external features when sorting familiar (vs. unfamiliar) faces. That the external features provided more information regarding identity when these faces were familiar highlights a surprising pattern of results that may differentiate card sorting from other face tasks. In addition, we found that the internal features led to better performance than the external features for unfamiliar faces. Again, previous research has shown a different pattern of results (Ellis et al., 1979). Here, we suggest that participants are simply sensitive to the image statistics present in the card sets, and were able to utilise the most informative sources of identity information. Regardless of familiarity, that seems to be the internal and not the external features. Of course, we could imagine situations in which the external features were equally informative, or more informative than the internal features (e.g., two men with similar internal features but very different hair colours/styles). Given such pairings, we would predict (in line with the patterns we see here) that participants would be sensitive to these image statistics and adapt their behaviours accordingly.

This result of the internal features contained more identity information was further clarified in Experiment 3, where participants were informed that only two identities were present in the set of cards when sorting familiar faces. We know that this additional information dramatically improves 'full' face sorting for unfamiliar viewers, resulting in typically error-free sorting (Andrews et al., 2015). Here, we found that familiar faces were sorted less accurately with only the external features available in comparison with the internal features, providing strong evidence that the external features simply contain less identity information.

While there is some evidence suggesting that unfamiliar face matching can be improved through the use of the internal features alone (Kemp et al., 2016), this result was not replicated in the current work, perhaps due to task differences. This lack of a difference between full face and internal feature performance suggests that unfamiliar participants were not misled by the external features (which are typically less informative with regards to identity), since this would result in worse performance with full face images. That internal features alone produced comparable sorting behaviours to full faces perhaps supports an account whereby participants developed some familiarity with the identities simply by carrying out the task (Andrews et al., 2015). This would explain why we see no 'external features detriment'.

Previous work has shown that faces can be learned through card sorting, in particular during a "two-sort", where participants are aware that only two identities are present (Andrews et al., 2015). This learning is evident in subsequent face matching performance with the two learned identities. Here, we found that unfamiliar participants sorted equally well using the full images and only the internal features. One interpretation might be that the internal features contain equivalent identity information, therefore producing similar performance. Alternatively, participants may have started to learn the identities during the card sorting task, meaning that they were no longer completely unfamiliar viewers. With this familiarity, we could expect an increasing reliance on the internal features, resulting in performance levels similar to full face sorting. Our current data do not allow us to differentiate between these two accounts, and this avenue provides an interesting focus for further investigation.

Our results in Experiment 1 are supported by previous research involving the same identities. The number of piles that our unfamiliar participants produced when given full faces (median of 6.5) was comparable with previous results (6.0 – Andrews et al., 2015; 7.5 – Jenkins et al., 2011). Both these prior studies (and all other card sorting studies that we are aware of) utilised greyscale images, and so presenting colour images in the current work appears to confer no noticeable advantage during sorting. This lack of a colour benefit mirrors previous work in face recognition, where colour cues are no more useful than simple (greyscale) luminance information, since both convey shape-from-shading and information regarding reflectance (Bruce & Young, 1998; Kemp, Pike, White, & Musselman, 1996; Russell, Sinha, Biederman, & Nederhouser, 2006).

We found that performance was noticeably better in Experiment 2 in comparison with our first experiment (see the *d'* values in Figures 2 and 3). Success on a card sorting task depends upon how different the two identities look from each other, as well as how much each identity varies across their own images. While we have not attempted to quantify these characteristics in the current work, it appears that fewer 'same person/different group' errors in Experiment 2 suggest that the two men displayed less within-person variability than the two women, causing fewer instances where images of the same person were thought to be two different people. Further, these differences in sensitivity across experiments appear largely confined to the unfamiliar faces, perhaps suggesting that the amount of within-person variability has little influence once faces are familiar. Why some individuals appear to vary less than others provides an avenue that has yet to be explored.

In Experiments 1 and 2, we find far fewer 'different person/same group' errors (often close to zero) in comparison with 'same person/different group' errors. This result has been discussed in previous work (Jenkins et al., 2011), and is a powerful demonstration that coping with within-person variability can be as difficult, if not more so, than dealing with between-person differences. Of course, this depends entirely upon the two identities chosen. One could imagine (and indeed this would be interesting to explore) that using twins, for example, would result in a far higher proportion of 'different person/same group' errors than we see in the current work.

In conclusion, we show across three experiments that card sorting behaviour with full face images is no different from sorting with only the internal features. This result seems initially to contradict previous research in that we know the external features play a significant role in unfamiliar face recognition and matching. We suggest that participants are sensitive to the information available, and both familiar and unfamiliar viewers are better able to utilise the internal features because they provide more identity information than the external features, but no less information than the full face images. These results build on our understanding of internal versus external feature processing with regard to familiarity, and call into question previous conclusions suggesting a simple shift from external to internal feature reliance with increasing familiarity.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

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 A. Mike Burton.

References

Andrews, S., Burton, A. M., Schweinberger, S. R., & Wiese, H. (2017). Event-related potentials reveal the development of stable face representations from natural variability. *The Quarterly Journal of Experimental Psychology, 70*(8), 1620-1632.

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

Balas, B., & Pearson, H. (2017). Intra- and extra-personal variability in person recognition. *Visual Cognition*. Advance online publication.

Balas, B., & Saville, A. (2017). Hometown size affects the processing of naturalistic face variability. *Vision Research*. Advance online publication.

Bonner, L., & Burton, A. M. (2004). 7–11-year-old children show an advantage for matching and recognizing the internal features of familiar faces: Evidence against a developmental shift. *The Quarterly Journal of Experimental Psychology, 57A*(6), 1019-1029.

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., 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., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied, 7*(3), 207-218.

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception*. New York, NY: Oxford University Press.

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

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.

Campbell, R., Walker, J., & Baron-Cohen, S. (1995). The development of differential use of inner and outer face features in familiar face identification. *Journal of Experimental Child Psychology, 59*, 196-210.

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

Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition, 11*(7), 857-869.

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

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

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

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.

Kemp, R. I., Caon, A., Howard, M., & Brooks, K. R. (2016). Improving unfamiliar face matching by masking the external facial features. *Applied Cognitive Psychology, 30*(4), 622-627.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: The role of shape from shading and pigmentation cues. *Perception, 25*, 37-52.

Laurence, S., Zhou, X., & Mondloch, C. J. (2016). The flip side of the other-race coin: They all look *different* to me. *British Journal of Psychology, 107*, 374-388.

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

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied, 14*(4), 364-372.

Osborne, C. D., & Stevenage, S. V. (2008). Internal feature saliency as a marker of familiarity and configural processing. *Visual Cognition, 16*(1), 23-43.

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception, 35*(6), 749-759.

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

Zhou, X., & Mondloch, C. J. (2016). Recognizing “Bella Swan” and “Hermione Granger”: No own-race advantage in recognizing photos of famous faces. *Perception*, *45*(12), 1426-1429.