Title: That's my teacher! Children's ability to recognize personally familiar and unfamiliar faces improves with age

Article Type: Empirical Article

Keywords: Face Recognition; Identity Perception; Familiar Faces; Children

Corresponding Author: Dr. Catherine Mondloch, Ph.D.

Corresponding Author's Institution: Brock University

First Author: Sarah Laurence, PhD

Order of Authors: Sarah Laurence, PhD; Catherine Mondloch, Ph.D.

Abstract: Most previous research on the development of face recognition has focused on recognition of highly controlled images. One of the biggest challenges of face recognition is to identify an individual across images that capture natural variability in appearance. We created a child-friendly version of the Jenkins, White, Van Montford & Burton (2011) sorting task to investigate children's recognition of personally familiar and unfamiliar faces. Children between 4 and 12 years of age were presented with a familiar/unfamiliar teacher's house and a pile of face photographs (nine pictures each of the teacher and another identity). Each child was asked to put all the pictures of the teacher inside the house while keeping the other identity out. Children over 6 years of age showed adult-like familiar face recognition. Unfamiliar face recognition improved across the entire age range, with considerable variability in children's performance. These findings suggest that children's ability to tolerate within-person variability improves with age and support a face-space framework in which faces are represented as regions, the size of which increases with age.
Ms. Ref. No.: JECP-D-15-00335
Title: That's my teacher! Children's ability to recognize personally familiar and unfamiliar faces improves with age
Journal of Experimental Child Psychology

Dear Ramesh,

We are submitting a revised version of our manuscript. We were delighted to see that the reviewers were so positive about our paper and appreciated its potential impact. We have addressed each comment raised in our response letter and, where appropriate, the manuscript. We believe that the manuscript is now stronger and hope it is ready for publication in JECP.

We have highlighted all changes in the manuscript, with one exception: Primary Analyses. Because these changes were so extensive (given separate analyses of familiar and unfamiliar faces) we did not highlight them to make it easier for you to read. Our responses are in italics.

Sincerely,

Cathy Mondloch and Sarah Laurence
Ms. Ref. No.: JECP-D-15-00335
Title: That's my teacher! Children's ability to recognize personally familiar and unfamiliar faces improves with age
Journal of Experimental Child Psychology

Dear Dr. Bhatt,

We are submitting a revised version of our manuscript. We were delighted to see that the reviewers were so positive about our paper and appreciated its potential impact. We have addressed each comment raised in our response letter and, where appropriate, the manuscript. We believe that the manuscript is now stronger and hope it is ready for publication in JECP.

Original comments are in bold font, our comments back are in italics, and quotes from the paper are in regular font. Additions to the paper are highlighted both in the letter and the manuscript. The only exception to this is the Primary Results section; because we rewrote that entirely (given separate analyses for familiar vs. unfamiliar faces) the highlighting makes it hard to read.

Editor’s Comments.
A significant issue, noted by Reviewer 1, pertains to the conclusion that recognition of unfamiliar faces improves with age. This conclusion is not supported by the nonsignificant interaction that you report in the Results section. Reviewer 1 suggests separate analyses of performance on unfamiliar and familiar faces as a way of understanding the effects of age. I think this approach is reasonable given the theoretical distinction that has been drawn between representations of familiar versus unfamiliar faces.

As requested we have now conducted separate analyses of d’, hits, and false alarms for familiar and unfamiliar faces. Accuracy increased with age for both familiar and unfamiliar faces when all children were taken into account. However, when we include only children aged 6 to 12 years (an age ranged tested with both familiar and unfamiliar faces) performance only improved with age for unfamiliar faces. Recognition of familiar faces was nearly perfect for children aged 6+, but younger children (aged 4 and 5 years) made several errors.

Also, in varying ways, all three reviewers take issue with several theoretical concepts that you use. For example, Reviewer 1 wonders about the connection between the concept of attractor fields and stimulus familiarity, and Reviewer 3 suggests that the notion of face-space does not contribute much to the current study.

We have moved our entire discussion of attractor fields to the discussion because our study was not explicitly designed to test this model; rather, the model provides one way to think about our findings. In addition, we now focus on attractor fields with respect to development rather than familiarity.

*1a) Detailed Response to Reviewers
Further, I agree with the reviewers that it will be helpful to present data that are currently not included, such as hits and false alarm scores.

*We have added these data.*

When you submit the revision, please include appropriate measures of effect size for all of your statistical tests. Also, include a separate letter (i.e., separate from the general cover letter) that describes your response to the reviewers' concerns.

*We have done so.*

Reviewer #1: In this paper, the authors examined the development of within- versus between-person familiarity in children between 4 and 12 years of age. Children were asked to sort photos of a familiar teacher or photos of a teacher who was unfamiliar to them. For familiar people, 6 year-old children achieved near perfect, adult-like performance. For unfamiliar people, recognition appeared to improve with age. To my knowledge, this is the first study to apply the Jenkins et al. sorting task to a developmental population. Although the approach is novel and the results are potentially informative, the current version of the manuscript requires further work to explore aspects of the data in more detail, to improve the writing, and sharpening the theoretical message of the piece. Below are the major issues as I see them.

1.) At several points in the manuscript (see abstract: page 22, lines 9-17; page 23, lines 41-48), the author(s) make the claim that recognition of unfamiliar faces improves with age. Yet, this conclusion is unsupported by the results where the interaction between age and familiarity failed to reach significance (B = .05, t = .75, p = .45, SR = .04). The authors will need to modify their interpretation of the findings so that it is aligned with the results. Nevertheless, it would be informative to know the potential source of the reliable Age (B = .15, t = 4.85, p < .001, SR = .27) effect. To probe possible age-related effects, I suggest that the authors carry out separate analyses for the Familiar and Unfamiliar groups to investigate whether sorting performance improves as a function of age.

As recommended, we have conducted separate analyses for familiar and unfamiliar faces. The analysis of unfamiliar faces reveals a significant effect of age. The analysis of familiar faces also reveals a significant effect of age when all children (aged 4 to 12 years) are included. However, when only the children between 6 and 12 years (comparable to those tested with unfamiliar faces) are included the effect of age is non significant. Indeed, only 4- and 5- years olds made errors when tested with familiar faces. We have altered our results section accordingly.

2.) In the introduction, there are gaps in the theoretical arguments that need to be further developed. For example, on page 5, additional explanation is needed to describe how a refined norm-based coding informs the debate between general...
cognitive development versus quantitative face processing accounts. Also, the connection between the concept of an attractor field and familiarity also requires further explication. That is, why would familiar faces have larger attractor fields and how does this relate to the Jenkins et al. finding showing that people sort familiar much more accurately than unfamiliar faces.

As noted above, we have moved information about attractor fields and norm-based coding to the discussion. Although we find the relation between this influential model and our study compelling, we did not intend to test the model. We hope that the connection is clearer. In addition, we have taken out discussion of the relation between attractor fields and familiarity as this is not really central to our work.

There seem to be critical differences between the Jenkins et al. study with adults and the current study with children. In the Jenkins et al. study, the participants performed an open-ended sorting task where the number of piles was underdetermined. In the current experiment, children performed a binary sorting between "Mrs. Smith" and not "Mrs. Smith". This difference is critical and requires a careful explanation of how it might influence sorting behaviors and interact with development. In general, the Introduction fails to provide a coherent, compelling justification for the current study and the predictions do not logically follow based on the discussion of previous studies and theories.

We opted to use a child-friendly adaptation of Jenkins’ method. We agree that the two approaches could yield different outcomes but both tap sensitivity to identity in the context of within-person variability. In the Discussion we now say: Finally our task had a different structure than that of Jenkins et al. (2011). An advantage of our task is that it is likely less cognitively demanding because children did not need to keep track of 40 photographs simultaneously; rather, they made a yes/no decision for each image. Nonetheless, an advantage of Jenkins’ sorting task is that participants are free to make as many or as few piles as they like. Future research should examine whether these two tasks yield similar outcomes.

3.) Similarly, I did not find the arguments presented in the General Discussion to be very persuasive. For example, the authors make a distinction between recognition and discrimination processes. They claim that their paradigm avoids an elimination strategy where the child chooses Face B because she knows that it is NOT Face A. However, children in the Familiar Face condition can employ a similar elimination strategy. That is, if the children do not recognize a given face as their teacher (not A), they know by exclusion that face is unfamiliar and therefore, should be left outside the house. This is true, but knowing that one picture in the pile was not the teacher would not lead to the conclusion that the next face was the teacher. In contrast, in a match-to-sample task, knowing that one stimulus does not match the target leads one to select the alternative regardless of whether that alternative was recognized.
We hope we have clarified our point on page 21 where we now say: A number of other studies investigating children’s view-invariant face recognition used choice-from-array (e.g., delayed match-to-sample) tasks (e.g., Bruce et al., 2000; Mondloch et al., 2003; de Heering et al, 2012; Anzures et al., 2013; Crookes & Robbins, 2014). A limitation of such tasks is that under conditions of uncertainty, children can rely on a process of elimination to complete the task (e.g., “the face on the left is not the target so I’ll select the face on the right”) because on each trial they are forced to choose one face as the correct response. Thus accurate performance might not reflect recognition per se; it might only reflect discrimination. Children use a similar elimination strategy when it comes to identifying novel facial expressions (e.g., Nelson & Russell, 2015) and words that they have never heard before (e.g., Carey & Bartlett, 1979). In our task a single face was presented on each trial. On each trial on which the teacher’s face was presented children had an opportunity to demonstrate recognition; the number of hits thus reflects their sensitivity to identity despite within-person variability in appearance. On each trial on which the similar distractor was presented children had an opportunity to demonstrate discrimination; the number of false alarms reflects children’s ability to tell two identities apart. Accuracy increased with age for unfamiliar faces; neither hits nor false alarms were significantly correlated with age indicating that changes in both recognition and discrimination contributed to improved performance.

However, recognition and discrimination distinction does seem to apply the divergent strategies that children employ in the familiar and unfamiliar conditions. In the familiar face condition, participants were employed a recognition process in which a given photo was compared to a stored memory of their teacher. In contrast, participants in the unfamiliar face condition had to based their decisions purely of perceptual discrimination (i.e., does the face match the target face?) without the benefit of a visual face memory. It would be useful for the authors to consider the strategy differences between children in the familiar and unfamiliar test conditions.

The reviewer is correct. All children could use the photo on the front of the house when making their decision and, in addition, children for whom the face was familiar could rely on previous experience with their teacher. Regardless, when the teacher’s face was presented the task was to recognize her despite a change in appearance. We feel that a discussion of strategies is beyond the scope of this paper although we do an image analysis. We also highlight this point on page 22 under the subsection Perception versus Memory: Our task was designed to be perceptual in nature; memory demands were completely eliminated (although children for whom the target face was familiar were able to rely on previous experience with their teacher’s face).

4.) Critical aspects of the data were not reported. For example, what was the accuracy for the control stimuli of the identical target faces and unrelated non-target faces? In addition to the d’ results, it would be useful to know about the within-category versus between-category errors. That is, did children make more inclusion errors (ie. putting non-target faces in the house) or exclusions (i.e., leaving target faces outside of the house). This value is especially important in the
unfamiliar face condition where mistakes were more prevalent. We now report the accuracy (perfect) for control stimuli on page 13: Every child recognized all four of the images identical to the image on front of the house, and none of the children put any of the four images of the dissimilar distractor inside the house. We agree that distinguishing hits and false alarms is of both applied and theoretical interest. Given the request to analyze accuracy for familiar and unfamiliar faces separately we have also done separate analyses of hits and false alarms. We made a new figure in which d', hits, and false alarms are presented in separate panels and we specifically compared number of inclusion vs. exclusion errors (page 14): The mean number of misses (M = 4.41, s.e. = .35) was greater than the mean number of false alarms (M = 1.61, s.e. = .27, t (72) = 5.17, p < .001, r = .97) indicating that, like adults, recognizing an identity despite variability in appearance was more challenging than discriminating the two identities present. Thank you for this suggestion!

Other suggestions:

Page 5, Line 31:

Page 6, line 4: in "however" after semi-colon. Done.
Page 6, line 5: what is meant by "ambient" images?
We have removed this term. We now say; ability to recognize identity in natural images across which appearance varies naturally.

Page 6, line 36-38: Instead of describing the "faces" are familiar or unfamiliar, it would be clearer to state that it was the "participants" who were familiar or not familiar with the faces. That is, familiarity is not a characteristic of the face stimulus but a characteristic of the participant.
We prefer to leave this as is only because it is more economical. We hope that because each teacher’s face was shown to children from the same school and children from a different school that it is clear what we mean by ‘familiar face’ vs. ‘unfamiliar face’.

Page 7, Line 58: - "en" face?
We replaced this with looking ahead.

Page 8 - Were the same teachers tested in the two different groups? How were the faces selected? If so, for Teacher X’s class of children, Teacher X would be the familiar person and Teacher Y would be the unfamiliar person and for Teacher Y’s class, the converse would be true. Teacher Y would be the familiar person and Teacher X would be the unfamiliar. This would control stimulus confounds due to the relative discriminability of the photos.

This information was included in Table 1. One teacher (A) was familiar to 11 children and never shown to children who were unfamiliar with her. The remaining five teachers were shown both to children in the same school (for whom the teacher was a familiar face) and to children in a different school (for whom the teacher was unfamiliar).
was done to control for differences in relative discriminability. Whereas the age range of children for whom the face was familiar was constrained by the grade(s) the teacher taught, the age range of children tested with each face when it was unfamiliar was comparable. Thus we are confident that stimulus effects are not an issue. We have highlighted this throughout the manuscript.

Page 7: The same stimuli were shown to both groups of children.

Page 9: All were shown to at least three children from their own school and five were shown to children from another school (i.e., to children for whom the teacher was an unfamiliar adult).

Page 12, Line 53-58: Were the photos equated for within-person variability? This should be clarified.

Within-person variability is idiosyncratic (see Burton et al., 2015) so there is no obvious way of equating the variability across individuals. We tried to match our stimulus selection procedure as closely to Jenkins et al. as possible. Because we wanted to study children’s ability to recognize facial identity despite natural variation both for familiar and unfamiliar faces we used every possible identity who agreed to share their photos; six teachers volunteered to provide pictures and all were used. Each teacher was asked to provide 10 photographs that met the following criterion: roughly frontal aspect, bigger than 150 pixels in height, free from occlusions. All the pictures were to have been taken on different days (page 12).

A strength of our study is that we used multiple identities which is an improvement on previous work (including our own) in which adults were tested with only 1 or 2 sets of photographs.

Page 16, Age is a continuous variable and familiarity is dichotomous variable. This analyses has been eliminated.

Page 16, Line 41-46: Please clarify the sentence "These results suggest that accuracy improved with age, was higher for familiar faces compared to unfamiliar faces, but the effect of age on accuracy did not vary as a function of familiarity."

This sentence has been deleted because we have now analyzed familiar and unfamiliar face data separately.

Page 18 - I suggest omitting the subheading supplemental analysis and include those analyses that are central to the study. I think the "general cognitive development" or block factor can be included in the regression analysis add much to the paper and the "duration" can be omitted.

We would like to keep our analyses organized as they are. The other reviewers seemed to like our organization. We think it is important to address the issue of general cognitive versus face-specific development but the primary question was whether d’, hits, and false
alarms changed with age. We do not want to remove duration. Readers will wonder for how long children knew their teacher and whether that differed with age; we want to be clear that this is not a concern in this paper and that they had known their teacher for several months, making data from the youngest children quite remarkable!

Page 19 - At the same time, this is a drawback in the study because each age group saw a DIFFERNT unfamiliar identity. Hence, stimuli is confounded with age where the some unfamiliar identities might be more difficult (or easier) to discriminate from the familiar identity than others.

This was not the case for unfamiliar identities. As noted above, for each of the 5 identities used as unfamiliar faces the youngest children tested were 5 to 6 years of age and the oldest children were 12 (see Table 1). It is the case that the age range of children who were familiar with each identity varied; however given the results with unfamiliar faces the effect of age cannot be attributed to stimulus effects.

Page 20, Line 46: I suggest using the term duration of "familiarity" rather than "exposure".
Thank you for this suggestion. Done.

Reviewer #2: This study examines children's face recognition with images that capture natural variation in appearance, using a version of Jenkins et al.'s (2011) sorting task. Children over 6 years of age showed adult-like familiar face recognition, whereas unfamiliar face recognition improved across the entire age range.

This is an important study that will attract enormous attention from researchers in this field. It is now clear that it is inadequate to study face recognition without taking account of the variation that different identities can exhibit in appearance. To my knowledge this is the first developmental study to explore this. I can find very little fault with this work. The manuscript is written well and also provides a good overview of the literature. I would like to see this work published and only have a few minor comments.

1. The introduction refers to norm-based encoding several times. However, for a wider readership it would help to include a sentence or two to define what that is. On page 4 we now say: a process by which each face is individuated based on how much each face deviates from an average or internal norm.
2. While d' and criterion is an appropriate analysis for this type of data, I also like to see % accuracy alongside these measures to get a full overview of participant's performance. This would also strengthen some statements in the text, such as "older children performed essentially without error" (p. 18). 

We have added % accuracy to our descriptive statistics; thank you for that suggestion. We think that figure 2 more clearly shows the pattern of performance.

3. The inclusion of the control images (i.e. the identical images and the very dissimilar identity images) is a strong aspect of the experimental design. However, while these are discussed in the General Discussion, I could not find these mentioned in the Results section.

We have added this statement: Every child recognized all four of the images identical to the image on front of the house, and none of the children put any of the four images of the dissimilar distractor inside the house.

4. In the analysis of criterion, it is stated that children were conservative in their responses (p. 17). It would help comprehension to include a sentence that states in plain terms what this means.

We elected to remove our analysis of criterion to avoid the paper becoming too lengthy. It is more intuitive to talk about hits and false alarms.

Reviewer #3: This is an interesting paper, which I think is quite significant. The authors show how children's ability to tolerate within-face variability improves with age, and also demonstrate important differences between familiar and unfamiliar face recognition. They use a novel design to test children's ability with naturally varying ('ambient') images, and the results are clear and compelling.

The authors do a very good job of explaining why this is important research. The Discussion on pages 21 to 26 captures this very well. I found some of the other theoretical material in the Intro and Discussion less compelling (as I will elaborate below). I also had a couple of minor issues to raise about the reporting of results.

Results: I realise that it is not straightforward to decide how to present these results, but I would ask the authors to consider whether presenting only d' is optimal. We have stats for Hits and FAs (p. 17), but I would have liked to see the data for these too (figures like Fig 2). We never actually see these numbers, and I would like to have an idea of just how accurate the children are in the two sub-components of the task. I also think it is interesting that the age-related changes were more pronounce in the Hits than the FAs. This is possibly more interesting than the authors acknowledge. Hits and FAs dissociate for unfamiliar (but not familiar) face tasks. That association might be worth reporting here too - I suspect they will be uncorrelated, but either result is interesting. I wonder whether there is any theoretical interest in the fact that the developmental trend is carried mostly by
Hits?
As noted above, we have now analyzed hits and false alarms separately. In adults FAs (perceiving that pictures of two different people belong to the same identity) are rare for both familiar and unfamiliar people (Jenkins et al., 2011; Laurence et al., 2015). In contrast, HITS (perceiving that different pictures of the same person belong to the same identity) are more likely when a face is familiar.

Our new analyses show that improved accuracy for unfamiliar faces between 6 and 12 years of age is attributable to more hits and fewer false alarms. But misses (failing to put a picture of the teacher into the house) are more common than false alarms. Children between 6 and 12 years rarely made an error when tested with familiar faces; younger children did make errors, most of which were misses. We explore this in the revised paper.

Theoretical development: I found the face-space material uncompelling in the Intro (p. 6-8) and Discussion (p. 26-7). It also seems unnecessary. The authors say that unfamiliar face recognition is more closely tied to image-properties than familiar face recognition, and they track children's ability to overcome some (but not all) of the image variability as they become adults. That seems to give a fairly complete understanding of their study. By invoking face-space explanations they do not seem to add anything, except some speculation.

Here are some examples:

p. 5 'The location of each face is determined by its value on the dimensions underlying face space, along which faces vary (e.g. the distance between the eyes, nose length)'. Although common, that seems an odd assertion when we have no idea what the dimensions are. Certainly, they cannot be distance between the eyes or nose length. Have a look at figure 1 - for the same person both distance between the eyes and nose length varies rather a lot. Of course, these physical measures are constant for the person, but they are very different in the different photos. Furthermore, we cannot recover their 'true' value by mentally transforming the images to the same pose. In order recover an accurate measure of distance between the eyes (say) we would need to know the distance between the camera and the face, and the perspective settings (focal length) of the lens. We cannot know this, and yet all the images will be perfectly recognisable for an familiar viewer.

p. 6 '... consistent with distinctive faces having bigger attractor field because they reside in less populated area of face space'  It is very hard to make that argument unless one knows what these dimensions are. Furthermore, arguments based on exemplar-density break down in spaces which are larger than two dimensions (Burton & Vokey, 1998, QJEP) - a point which is acknowledged in recent descriptions of face space (Valentine, Lewis & Hills, 2015, QJEP).
There are plenty of other examples throughout. I should be clear in saying that the invocation of face space is very common, and many researchers are happy to base their explanations on this structure, in the absence of any proposal about what the dimensions of such a space actually are. This is the authors' paper, and they are free to interpret the data as they see fit. However, from my own reading, the data are perfectly compelling as they stand, and do not benefit from explanation in such hypothetical terms.

Despite these quibbles, my overall impression is that this is a very good paper. In my view, it represents significant progress in the field and has the potential to be very influential.

We have removed the theory from the introduction. Our hypotheses now stem from the literature regarding familiar/unfamiliar face recognition. We’ve also reduced our discussion of norm-based coding and elected, based on the reviewers’ suggestion, to eliminate discussion of how norm-based coding and attractor fields relate to familiar vs. unfamiliar faces.

As the reviewer points out, face space has been used to explain much past research on the development of face perception. To ignore it entirely would limit the impact of our findings. By using face space as a means of explaining our data, we wish to highlight that variability should be incorporated into our models, and provide a testable (and perhaps falsifiable) framework for future research.

We have also highlighted a major point of our paper on page 23: Our data indicate that the developmental trajectory for recognizing unfamiliar faces is different than that for familiar faces. The ability to recognize unfamiliar identities improved between 6 and 12 years of age. In contrast, recognition of familiar identities continued to improve until 6 years of age, with no evidence for age-related improvement after that. This is consistent with Burton’s argument that there is an important difference between familiar and unfamiliar face recognition (Burton, 2013).

Signed: Mike Burton

Collectively we think these changes have clarified the paper, highlighted the most important points, and yet provide enough speculation to stimulate debate about the development of face recognition, the difference in familiar vs. unfamiliar face recognition, the relevance of norm-based coding, and the importance of incorporating within-person variability.

Thank you again for your careful attention to our work.
That’s My Teacher! Children’s ability to recognize personally familiar and unfamiliar faces improves with age

Sarah Laurence and Catherine J. Mondloch

Brock University

Author Note

Sarah Laurence and Catherine J. Mondloch, Department of Psychology, Brock University.

Correspondence should be addressed to C. Mondloch – cmondloch@brocku.ca
That’s My Teacher! Children’s Ability to Recognize Personally Familiar and Unfamiliar Faces Improves with Age

Word count (text + references): 9887
Abstract

Most previous research on the development of face recognition has focused on recognition of highly controlled images. One of the biggest challenges of face recognition is to identify an individual across images that capture natural variability in appearance. We created a child-friendly version of the Jenkins, White, Van Montford & Burton (2011) sorting task to investigate children’s recognition of personally familiar and unfamiliar faces. Children between 4 and 12 years of age were presented with a familiar/unfamiliar teacher’s house and a pile of face photographs (nine pictures each of the teacher and another identity). Each child was asked to put all the pictures of the teacher inside the house while keeping the other identity out. Children over 6 years of age showed adult-like familiar face recognition. Unfamiliar face recognition improved across the entire age range, with considerable variability in children’s performance. These findings suggest that children’s ability to tolerate within-person variability improves with age and support a face-space framework in which faces are represented as regions, the size of which increases with age.

Keywords: Face Recognition; Identity Perception; Familiar Faces; Children
That’s My Teacher! Children’s Ability to Recognize Personally Familiar and Unfamiliar Faces Improves with Age

Adults are experts in face recognition (see Maurer, Le Grand, & Mondloch, 2002). They can recognize individual faces at a glance and are able to do so under poor lighting conditions, across variability in expression and point of view, and even after a face has aged several years. Adults’ expertise has been attributed to norm-based coding (Rhodes, Jeffery, Taylor, Hayward, & Ewing, 2014; Valentine, 1991), holistic processing (Young, Hellawell, & Hay, 1987; Hole, 1994), sensitivity to feature shape and spacing (Freire & Lee, 2001; Mondloch, Le Grand, & Maurer, 2002), and specialized neural mechanisms (e.g., Allison, McCarthy, Nobre, Puce & Belger, 1994; Bentin, Allison, Puce, Perez, & McCarthy, 1996; Kanwisher, McDermott, & Chun, 1997). At what age face recognition and the underlying mechanisms become adult-like is a matter of ongoing debate.

To date, the focus has been on the development of underlying mechanisms rather than an absolute measure of children’s face recognition per se. Qualitatively, the mechanisms underlying face recognition appear to be adult-like by early childhood. Holistic processing, most directly measured by the composite effect whereby perception of one (e.g., the top) half of the face is influenced by the other half (Young et al., 1987; Hole, 1994), is evident by 4 to 6 years of age (e.g., de Heering, Houthuys, & Rossion, 2007; Macchi Cassia, Picozzi, Kuefner, Bricolo, & Turati, 2009; Mondloch, Pathman, Maurer, Le Grand, & de Schonen, 2007). Evidence from the part-whole task, whereby recognition of a face part (e.g., Joe’s eyes) is more accurate when the part is presented in the context of the face than when presented in isolation, is consistent with early emergence of holistic processing (Pellicano & Rhodes, 2003; Tanaka, Kay, Grinnell, Stansfield, & Szechtier, 1998). Likewise, young children are sensitive to feature shape and spacing (e.g., Baudouin, Gallay, Durand,
& Robichon, 2010; Gilchrist & McKone, 2003; Macchi Cassia, Turati, & Schwarzer, 2011; McKone & Boyer, 2006; Pellicano, Rhodes, & Peters, 2006; Mondloch & Thomson, 2008; Mondloch et al., 2002), and show adult-like patterns of norm-based coding (Anzures, Mondloch & Lackner, 2009; Hills, Holland & Lewis, 2010; Jeffery et al., 2010; Short, Hatry, & Mondloch, 2011; Nishimura, Maurer, Jeffery, Pellicano, & Rhodes, 2008), a process by which each face is individuated based on how much each face deviates from an average or internal norm.

What continues to be debated is the extent to which there is quantitative improvement in face processing during childhood. Some researchers (Crookes & Robbins, 2014; McKone, Crookes, Jeffrey, & Dilks, 2012; Weigelt, Koldewuyn, Dilks, Balas, McKone & Kanwisher, 2014) argue for quantitative maturity in face perception by 5 years of age with any further improvements being attributed to general cognitive development. Others (de Heering, Rossion, & Maurer, 2012; Short, Lee, Fu, & Mondloch, 2014; Tanaka et al., 2014) argue that quantitative improvements in face processing continue beyond this age. For example, Short et al. showed that norm-based coding continues to be refined after 5 years of age.

Although understanding the development of the processes underlying face recognition is of theoretical importance, recent developments in the field of adult face perception highlight an aspect of face recognition that has been largely ignored in the literature: The ability to recognize a face’s identity across a set of images that incorporate natural variability in appearance (Burton, 2013). The vast majority of studies investigating face recognition in adults and children have used tightly controlled stimuli. For example, Mondloch and colleagues (Mondloch, Geldart, & Maurer, 2003; Mondloch et al., 2002) created a carefully controlled set of stimuli in which all images were taken from the same distance, with the same camera, and under identical lighting conditions. Hair was covered with surgical caps, clothing was covered with a cape, and blemishes were removed.
Presenting such images minimizes the observer’s ability to use non-face cues to identity and provides important information about our ability to discriminate between images/identities; however, it ignores the ability to recognize identity in natural images across which appearance varies naturally. This aspect of face recognition is crucial for daily interactions. It allows us to recognize our neighbor when she returns disheveled from a camping trip or our uncle after several years of aging.

In a seminal paper by Jenkins, White, Van Montfort, & Burton (2011) adults were given 40 photographs and asked to sort them into piles such that each pile contained all of the images of one person. Images were downloaded from the Internet and incorporated natural variation in appearance (in hairstyle, lighting, expression, viewpoint, makeup). Participants were not told that there were 20 pictures of each of two people (i.e., that the correct solution was two piles). Nonetheless, participants who were familiar with the identities performed without error. In contrast, participants who were unfamiliar with the identities made about seven piles (i.e., perceived seven different identities) but rarely placed pictures of the two identities into the same pile. These findings highlight the challenge associated with recognizing identity across natural variation in images when faces are unfamiliar.

Although errors in discrimination, telling faces apart, were rare for both familiar and unfamiliar faces, errors in recognition, telling faces together (Jenkins et al.) were prevalent for unfamiliar faces, but rare for familiar faces.

Almost nothing is known about children’s ability to recognize familiar versus unfamiliar faces across images containing natural variability in appearance. The vast majority of studies investigating the development of recognition of unfamiliar faces (i.e., faces learned in the context of the experimental paradigm) have used identical images at study and test (Baudouin et al., 2010; Gilchrist & McKone, 2003; Macchi Cassia et al, 2011; McKone & Boyer, 2006; Pellicano et al., 2006; Mondloch & Thomson, 2008; Mondloch et al., 2002; Pellicano & Rhodes, 2003; Tanaka et al.,
CHILDREN’S FACE RECOGNITION

1998). A few have incorporated some variability, but even in these the variability was limited. de Heering et al. (2012) found improved identity matching between 6 and 12 years of age in the Bentin face recognition task when images of the target identity varied in viewpoint and lighting. A similar pattern of results was observed when images of the target were taken on different cameras (Megreya & Bindemann, 2015). Crookes and Robbins (2014) found no difference in the extent to which changes in viewpoint impaired face recognition in 8-year-olds and adults. However, they used facegen (computer-generated) images and simply showed the same image from a different view. In two previous studies (Bruce et al., 2000; Mondloch et al., 2003) children completed a delayed match-to-sample task in which the target identity was shown from a different point of view or displaying a different emotion at study and test; performance improved between 6 and 10 years of age. Nonetheless, all images were taken under similar lighting conditions, on the same day, and with the same camera. None of these studies challenged children with the degree of variability encountered in the Jenkins et al. sorting task.

Furthermore, only a few studies have contrasted children’s recognition of unfamiliar with familiar faces. Testing children with familiar faces is challenging because, unlike adults who can be tested with famous faces (e.g., politicians and actors), it is difficult to find a set of familiar faces with which a large number of children are familiar. A few studies have tested children with personally familiar faces (e.g., classmates or teachers). These studies showed that children, like adults, rely on internal features to recognize familiar faces but external features to recognize unfamiliar faces (Bonner & Burton, 2004; Ge et al., 2008; Wilson, Blades, Coleman & Pascalis, 2009; but see Newcombe & Lie, 1995) and that the effect of paraphernalia on children's face recognition is reduced for familiar faces (Diamond & Carey, 1977). Again, however, natural variation in appearance was limited. Bonner and Burton, like Newcombe and Lie, tested 7- to 11-year-old
children with two images of each identity; the images were taken from two points of view, but all pictures were taken on the same day and models were asked to smile for each picture. Ge et al. used a single (looking ahead, neutral expression) image of each identity, as did Wilson et al. Collectively, these studies leave open the question as to whether children can tolerate the same variability as adults when viewing familiar faces.

In the current study we examined children's tolerance for natural variability when recognizing identity in unfamiliar and personally familiar faces by creating a child-friendly version of Jenkins et al.'s (2011) sorting task. We collected 10 photographs from each of six female teachers. As in Jenkins et al., the images incorporated natural variability in appearance. One image was placed on the front of a shoebox decorated to look like a house. The remaining nine images were placed in a pile along with nine images of a matched identity—another woman of similar age and hairstyle. Children were asked to place all pictures of the target identity into the house while keeping all other people out. The number of pictures of the target identity placed into the house provides a measure of recognition (perceiving the same identity despite variability in appearance) whereas the number of pictures of the matched identity placed outside of the house provides a measure of discrimination (telling different people apart). Each child was tested with one target identity. Critically, we tested children across a broad age range (4 to 12 years) and half of the children were tested with images of their teacher, with whom they were highly familiar, and the other half were tested with images of a teacher from a different school, with whom they were unfamiliar. The same stimuli were shown to both groups of children.

To validate our child-friendly method, eight additional control images were intermixed with these test stimuli. Four control images were identical to the one on the front of the house and four control images showed another woman who was much older than the target identity and whose hair
differed in both color and style. To be included in our analyses, children were required to place in the house at least three of the identical images and no more than one image of the dissimilar identity.

We made three predictions. First, we predicted improved performance with age. Children’s ability to match unfamiliar faces in images incorporating limited variability improves with age (e.g., Bruce et al. 2000; Mondloch et al., 2003; Heering et al., 2012; Megreya & Bindemann, 2015), a pattern likely due to improvements both in children’s ability to tell two faces apart, and their ability to recognize an identity despite variability in appearance (Megreya & Bindemann, 2015 found that correct identifications and correct rejections improve during childhood). This hypothesis would be supported if older children both put more pictures of the teacher and fewer pictures of the distractor into the house.

Second, we hypothesized that children, like adults, would make fewer errors when tested with a familiar than with an unfamiliar identity. As noted above, familiar faces can be characterized as having a sufficiently robust representation to allow recognition across a range of inputs (see Burton, Jenkins, Hancock & White, 2005; Johnston & Edmonds, 2009; Burton, Jenkins & Schweinberger, 2011; Jenkins et al., 2011). Evidence also suggests that lower-level image properties contribute more to unfamiliar than familiar face recognition (Hancock, Bruce & Burton, 2000). As a result children, like adults, might tolerate less variability for unfamiliar faces because they rely more on pictorial cues for recognition.

Third, we hypothesized that children, like adults, would make more errors in recognition (i.e., failing to recognize that multiple images belong to the same identity) than in discrimination (i.e. in misidentifying non-target images). This would be consistent with adults splitting images of the same person into multiple piles in the sorting task (Jenkins et al., 2011) and with one previous study (Saltz & Sigel, 1967) in which children’s ability to recognize identity across small variations in appearance
was examined. In that study, 6- and 9-year-old children were shown a target face followed by three images; on any given trial either zero, one, two, or all three images matched the target in identity. Across all levels of task difficulty children were less likely than adults to find all matches, despite all pictures being taken on the same day and varying only in facial expression or slight differences in head position. In contrast, false alarms (misidentifying a foil as the target identity) did not change with age. Saltz and Sigel concluded that children overdiscriminate identity, as do adults when tested with unfamiliar faces in the sorting task (Jenkins et al.).

Method

Participants

Teachers. Six teachers from four schools agreed to donate their photographs to be used as stimuli. All teachers were Caucasian females. Of the six teachers that donated their pictures, all were shown to at least three children from their own school and five were shown to children from another school (i.e., to children for whom the teacher was an unfamiliar adult).

Children. A total of 139 children participated in the study and were included in the final analysis. Children were recruited from local elementary schools and received a small toy and a certificate for their participation. Sixty-six children ($M_{age} = 7.04$ years; range = 3.92 to 12.08) from four schools were familiar with one of the six teachers; a student was considered familiar with the teacher if they were currently in the teacher’s class, were in the teacher’s class the previous year, and/or had regular daily contact with the teacher. All students had known their teacher for at least 3 months.
We also tested 73 children (Mage = 9.26; range = 5.00 to 12.17) from two schools who were unfamiliar with the teachers. Given the performance of the youngest children tested with unfamiliar faces (see results) we did not test 4-year-olds in this condition.

Table 1 shows the number of children tested who were familiar versus unfamiliar with each of the six teachers, and their ages (mean and range). One additional child was excluded because they failed to successfully complete the training task (see procedure).

Table 1

*The numbers, mean age, and age range of children familiar and unfamiliar with each of the six teachers.*

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Familiar</th>
<th>Unfamiliar</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>n = 11</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>M age = 5.64 (4.25 to 7.17)</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>n = 4</td>
<td>n = 12</td>
</tr>
<tr>
<td></td>
<td>M age = 6.48 (3.92 to 12.08)</td>
<td>M age = 9.67 (6.25 to 12.17)</td>
</tr>
<tr>
<td>C</td>
<td>n = 3</td>
<td>n = 14</td>
</tr>
<tr>
<td></td>
<td>M age = 5.08 (4.92 to 5.33)</td>
<td>M age = 9.56 (5.67 to 12.08)</td>
</tr>
<tr>
<td>D</td>
<td>n = 12</td>
<td>n = 14</td>
</tr>
<tr>
<td></td>
<td>M age = 6.12 (4.08 to 11.83)</td>
<td>M age = 9.76 (6.58 to 12.17)</td>
</tr>
<tr>
<td>E</td>
<td>n = 9</td>
<td>n = 12</td>
</tr>
<tr>
<td></td>
<td>M age = 4.76 (4.00 to 5.58)</td>
<td>M age = 9.26 (6.42 to 11.67)</td>
</tr>
<tr>
<td>F</td>
<td>n = 27</td>
<td>n = 21</td>
</tr>
<tr>
<td></td>
<td>M age = 9.09 (6.42 to 12.00)</td>
<td>M age = 8.50 (5.00 to 11.83)</td>
</tr>
<tr>
<td>Overall</td>
<td>n = 66</td>
<td>n = 73</td>
</tr>
<tr>
<td></td>
<td>M age = 7.04 (3.92 to 12.08)</td>
<td>M age = 9.26 (5.00 to 12.17)</td>
</tr>
</tbody>
</table>
**Stimuli**

**Teacher Stimuli.** Collection and preparation of stimuli were based on the methods used by Jenkins et al. (2011). For each of the six teachers, the experimenters selected a matched identity. The matched identities were individuals known to the experimenters who matched in age, skin tone, and hair colour the teacher with whom they were paired. Ten pictures of each of six teachers and nine pictures of six matched identities were used as test stimuli. The teachers and the matched identities were instructed to provide pictures that met the following criteria: roughly frontal aspect, bigger than 150 pixels in height, free from occlusions. All the pictures were to have been taken on different days (see Figure 1).

*Figure 1:* An example of the variability contained in the images.

In addition to the above test stimuli we included control stimuli designed to ensure that each participant understood the task and remained attentive throughout. For each of the teachers, an unfamiliar dissimilar identity was selected (e.g., celebrities from XXXX who are not famous in XXXX). The dissimilar identities looked very different from the teachers (e.g., if the teacher had dark hair and a long, thin face then the dissimilar identity had light blonde hair and a round, plump
CHILDREN’S FACE RECOGNITION

face). A single image of each dissimilar identity was obtained via a Google image search. Four copies were embedded among the test stimuli. In addition one image of the teacher was selected to go on the front of a toy house and four copies of that image were embedded among the test stimuli. Children were required to include at least three of these exact image matches and no more than one instance of the dissimilar identity to be included in the final sample; all participants met this criterion.

All images were cropped so that only the head was visible, converted to greyscale and printed on card measuring 38 x 50 mm in size.

**Training stimuli.** Five pictures of Buzz Lightyear and four pictures of Noddy were obtained using a Google image search. The images were selected and prepared using the same criteria as Jenkins et al. (2011) and one image of Buzz Lightyear was selected to go on the front of a toy space ship.

**Design and Procedure**

**Teacher task.** The entire procedure received clearance from the Research Ethics Board at XXXX XXXX. We obtained informed written consent from each child’s parent and verbal assent from each child. Each child was presented with a house on which we mounted a single image of either a familiar teacher or an unfamiliar woman. The experimenter then explained: “This is Mrs. Smith’s house. Look, there’s a picture of Mrs. Smith on the front of the house. Have you ever seen Mrs. Smith before? I’ve got a pile of pictures; some of the pictures are of Mrs. Smith and some of the pictures are of other people. All of the pictures of Mrs. Smith need to go in the house but we have to be careful that no one else’s pictures go inside. Do you think you can help me put all of the pictures of Mrs. Smith inside her house? Great!” The experimenter then proceeded to give each picture to the
child and the child was asked, "Does this picture go into Mrs. Smith's house?" When the teacher was familiar to the child we used her real name.

There were a total of 26 pictures (trials): nine different pictures of the teacher; nine different pictures of an identity who looked physically similar to the teacher (matched identity); four images identical to the image on the front of the house (criterion trials); and four identical images of an identity who looked very dissimilar to the teacher (criterion trials). The pictures were divided into three blocks. In each block there were three pictures of the teacher, three pictures of the matched identity, and two criterion trials (one identical image and one image of the dissimilar identity). The images in each block were presented in a random order. After completing the three blocks each child completed two final criterion trials: one identical image and one dissimilar identity. This was to verify that each child had remained attentive through the experiment.

Training task. The teacher task was preceded by a training task in which each child was presented with Buzz Lightyear’s spaceship and a pile of nine pictures comprising five pictures of Buzz (one of which was identical to the image on the spaceship) and four pictures of another character (Noddy). They were instructed to place all images of Buzz into the spaceship but to keep Noddy out. To ensure that children understood both the importance of discriminating between identities and that Buzz’s appearance could change, they were required to put four out of the five images of Buzz inside the house and to place all pictures of Noddy aside to be included in our analyses. All but one child (a 10-year-old) met this criterion.

Results

Primary Analyses

Every child recognized all four of the images identical to the image on front of the house, and none of the children put any of the four images of the dissimilar distractor inside the house.
were based on trials in which the teacher and her matched identity were presented. Signal detection theory (d’) was used to estimate accuracy: a hit was defined as putting a picture of the teacher into the house, a miss as placing a picture of the teacher outside of the house, and a false alarm as putting a picture of the matched identity into the house. Because adults’ performance on a comparable task varies for familiar vs. unfamiliar faces (Jenkins et al., 2011) we analyzed performance separately for familiar and unfamiliar faces. To better understand age-related changes in accuracy we conducted parallel follow-up analyses of hits and false alarms. Figure 2 shows age-related changes in d’, hits, and false alarms as a function of age for unfamiliar and familiar faces.

**Unfamiliar Faces.** Accuracy (M$_{\text{percentage correct}}$ = 66.51%, s.e. = 1.79%) improved with age, as reflected in the significant correlation between age and d’, $r = .36, p = .002$. Children could make two kinds of errors: misses and false alarms. The mean number of misses (M = 4.41, s.e. = .35) was greater than the mean number of false alarms (M = 1.61, s.e. = .27, t(72) = 5.17, $p < .001$, $r = .97$) indicating that, like adults, recognizing an identity despite variability in appearance was more challenging than discriminating the two identities present. Although hits tended to increase with age, $r = .19$ and false alarms tended to decrease, $r = -.17$, neither correlation was significant, ps = .10 and .15, respectively. This indicates that age-related changes in recognition and discrimination both contributed to improved accuracy. This is consistent with our finding that the total number of faces put into the house was not correlated with age, $r = .04, p = .74$. Thus improved accuracy with age cannot be attributed simply to young children putting more or fewer pictures into the house compared to older children.

**Familiar Faces.** Accuracy (M$_{\text{percentage correct}}$ = 92.51%, s.e. = 1.47%) improved with age, as reflected in the significant correlation between age and d’, $r = .44, p < .001$. Both misses (M = 1.03, s.e. = .24) and false alarms (M = .32, s.e. = .15) were rare, with misses occurring more frequently
CHILDREN’S FACE RECOGNITION

than false alarms, \( t(65) = 2.44, p = .02, r = .87 \). As with unfamiliar faces, the correlation between the number of false alarms and age was not significant, \( r = -.18, p = .16 \). However, there was a significant correlation between age and the number of hits, \( r = .38, p = .002 \), suggesting that the ability of children to recognize their teacher increased with age. We note, however, that the age range of children tested with a familiar face included 4- and 5-year-olds whereas the age range of children tested with unfamiliar faces did not. Thus, we conducted another set of analyses including only children aged 6 to 12 years (n = 34) who were tested with a familiar face. Their performance approached ceiling (\( M_{\text{percentage correct}} = 98.20\%, \text{s.e.} = .80\% \)). Almost every picture of the teacher was placed in the house (\( M = 8.82, \text{s.e.} = .13 \)), pictures of the distractor were almost always kept out (\( M_{\text{false alarms}} = .15, \text{s.e.}, = .07 \)), and age did not correlate with any of our measures, \( ps > .24 \). Thus, by 6 years of age children, like adults, perform (nearly) without error when tested with familiar faces, with improvements between 4 and 6 years of age.
Figure 2. Children’s accuracy ($d'$), hits and false alarms for unfamiliar (a, c, e) and familiar (b, d, f) faces. Each dot represents the data for an individual child.
Supplemental Analyses

**General cognitive development.** It is important to ask the extent to which age-related improvements in recognition of a target identity can be attributed to general cognitive development (e.g., a decrease in concentration over trials). For unfamiliar faces, a repeated measures ANOVA with Block as the repeated measure and age as a covariate showed there was no effect of block on accuracy (d'), \( p = .78 \), recognition (hits), \( p = .26 \), or discrimination (false alarms), \( p = .09 \). For familiar faces, older children performed essentially without error and so we analyzed whether performance of younger children (aged 4 and 5 years) changed across blocks. There was no change in performance across blocks: Repeated measures ANOVAs with Block as the within-subjects variable revealed no effect of block on accuracy (d'), \( p = .34 \), recognition (hits), \( p = .40 \) or discrimination (false alarms), \( p = .45 \). Collectively these analyses show that improvements in performance as a function of age cannot be attributed to young children losing concentration over time.

To further verify that deterioration in young children’s performance across blocks (e.g., loss of concentration) did not contribute to a significant effect of age we replicated our primary analyses but restricted our analysis to the first block of trials. We found a significant correlation between age and d’ for unfamiliar (\( r = .35, p = .003 \)) and familiar (\( r = .34, p = .008 \)) faces. The number of hits increased with age for familiar (\( r = .30, p = .02 \)), but not unfamiliar (\( r = .18, p = .14 \)) faces. The correlation between age and false alarms was not significant for unfamiliar (\( r = -.2, p = .09 \)) or familiar (\( r = -.15, p = .23 \)) faces. This pattern of results replicates the findings from the main analysis including all the trials.

**Image analysis.** One strength of our paper is that children were tested with five different unfamiliar identities, each of which also served as a familiar identity for other children. The number
of children tested with each identity varied \((n = 12 \text{ to } 21)\) but the age ranges were comparable. Thus the developmental pattern observed cannot be attributed to any one identity and the difference in performance for familiar and unfamiliar faces cannot be attributed to stimulus effects. To determine whether some images were especially hard to match we calculated the proportion of children who placed each target image in the house and then calculated the mean and standard deviation across images. We identified seven images that fell more than 1SD below the mean; three of these differed from the target image in hairstyle (e.g., hair shorter/longer), one in the presence/absence of glasses, one in both hairstyle and the presence/absence of glasses, and one in other paraphernalia (the test image included a small crown whereas the target face did not). The final image did not deviate from the target in any easily identifiable way and none of the seven images differed in expression.

Although this suggests that such changes may cross a threshold that children use to discriminate identities, we note that five other images differed similarly from the target face and yet were not especially difficult.

Another strength of our study is that children were tested with six different personally familiar adult identities. The number of children tested with each identity varied (between 3 and 27) because we could only test children who returned the consent form. One teacher was tested only with older children \((> 6 \text{ years of age})\), the age range in which performance was (nearly) perfect \((\text{Mean } d' = 3.10)\); she did not instruct 4- and 5-year-olds. It is unlikely that (nearly) perfect performance among older children can be attributed to this identity because 1) we tested children from another school with the same identity and saw improvement with age, and 2) we saw similar performance among the seven children who were more than 6 years of age and tested with two other identities \((\text{Mean } d' = 2.93)\).
Three of the six teachers were familiar to >10 children. To determine whether some images were especially hard to match we calculated the proportion of children who put each image into the house and then calculated the mean and standard deviation. (We did not include the remaining identities in this analysis because sorting of particular images for these identities might reflect the spurious behavior of one child.) Across identities accuracy for two images fell below 1SD of the mean. Although one image differed from the target image in the presence/absence of glasses, so too did one other image that was successfully placed in the house. No feature distinguished the other image.

**Duration of familiarity.** There was some variability in how long children had known their teacher, a potential confound highlighted by Ge et al. (2008). Some 4- to 5-year-olds were tested late in November (i.e., after knowing their teacher for 3 months) in contrast to (all but one) older children who were tested in April or May (i.e., after knowing their teacher for 8 or 9 months). Again, this variability simply reflects the difficulty in obtaining teacher photos and then consents from children in the same classroom. The performance of children tested after knowing their teacher for 3 months ($Mean \, d' = 2.51, s.e. = .21$) was comparable to (or slightly better than) that of the remaining 13 4- and 5-year-olds who were tested after knowing their teacher for 8 or 9 months ($Mean \, d' = 1.99, s.e. = .23$). Thus, the shorter period of familiarity for these children cannot explain improved recognition of familiar faces between 3 and 6 years of age.

**Discussion**

To our knowledge, this is the first time children’s recognition of familiar and unfamiliar faces has been measured using a task that is representative of the challenges encountered in the real world (e.g., in a task measuring recognition of identity in unconstrained natural images that incorporate within-person variability). It is also one of very few studies comparing children’s recognition of
unfamiliar versus personally familiar faces (see also Bonner & Burton, 2004; Diamond & Carey, 1977; Ge et al., 2008; Newcombe & Lie, 1995; Wilson et al, 2009) and no previous study making this comparison has involved multiple images of familiar identities. Our results include two key findings. First, they provide evidence that children, like adults (Jenkins et al., 2011), are able to tolerate more variability for familiar than for unfamiliar faces. Second, they provide evidence that children's ability to recognize identity in natural images that incorporate within-person variability improves with age, with adult-like performance achieved at a younger age for familiar faces. After 6 years of age children performed (nearly) without error when tested with familiar faces whereas performance continued to improve when tested with unfamiliar faces. Although perfect or nearly perfect performance often is problematic because ceiling effects can mask developmental change, such a concern is not warranted here; in this design perfect performance indicates achievement of adult-like expertise. The fact that perfect performance was not achieved in children for whom the identities were unfamiliar suggests that our stimulus set was challenging. These results have important theoretical implications for debates about face-specific development during childhood as well as practical implications (e.g., for eye-witness testimony).

Our finding age-related improvements in unfamiliar identity recognition between 6 and 12 years of age is consistent with evidence from several studies showing age-related improvements in recognition of previously unfamiliar faces across smaller variations in appearance (e.g., Bruce et al., 2000; Mondloch et al., 2003; de Heering et al., 2012; Megreya & Bindemann, 2015; but see Crookes & Robbins, 2014 for computer-generated faces). Our experimental design has several distinct advantages. First, it allows us to separate the development of discrimination (telling people apart) versus recognition (recognizing identity across images that capture natural variation in appearance). Second, our task minimized memory demands because one image of the target identity was visible at
all times. Weigelt et al. (2014) claim that although there is face-specific improvement in memory throughout childhood, there is no face-specific perceptual development after 5 years of age. Our task is perceptual in nature and thus contributes to this debate. Third, it directly compares performance for familiar versus unfamiliar faces because images of all but one teacher were shown to children in both the unfamiliar and familiar condition; this controls for stimulus effects. We discuss each of these points below.

**Discrimination versus Recognition**

A number of other studies investigating children’s view-invariant face recognition used choice-from-array (e.g., delayed match-to-sample) tasks (e.g., Bruce et al., 2000; Mondloch et al., 2003; de Heering et al., 2012; Anzures et al., 2013; Crookes & Robbins, 2014). A limitation of such tasks is that under conditions of uncertainty, children can rely on a process of elimination to complete the task (e.g., “the face on the left is not the target so I’ll select the face on the right”) because on each trial they are forced to choose one face as the correct response. Thus accurate performance might not reflect recognition per se; it might only reflect discrimination. Children use a similar elimination strategy when it comes to identifying novel facial expressions (e.g., Nelson & Russell, 2015) and words that they have never heard before (e.g., Carey & Bartlett, 1978). In our task a single face was presented on each trial. On each trial on which the teacher’s face was presented children had an opportunity to demonstrate recognition; the number of hits thus reflects their sensitivity to identity despite within-person variability in appearance. On each trial on which the similar distractor was presented children had an opportunity to demonstrate discrimination; the number of false alarms reflects children’s ability to tell two identities apart. Accuracy increased with age for unfamiliar faces; neither hits nor false alarms were significantly correlated with age indicating that changes in both recognition and discrimination contributed to improved performance. Our finding that the
number of misses exceeded the number of false alarms suggest that recognizing faces when appearance varies is especially challenging, consistent with data from adults (Jenkins et al., 2011; Laurence, Zhou, & Mondloch, 2015).

Our results contradict those of Crookes and Robbins (2014) who found no difference in view-invariant face recognition between 8-year-olds and adults. In an old/new recognition task they matched adults’ and 8-year-olds’ accuracy in a same-view condition by adjusting the number of faces studied; the decrement in performance in their different-view condition was similar for the two age groups. However, they used highly controlled computer generated faces (minimizing variability) and their 2AFC task does not allow one to infer the strategy that children were using to complete the task (rejecting the foil vs. recognizing the target). Our results suggest that children’s invariant recognition (e.g., their ability to recognise a person despite variability in their appearance) does improve with age—at least when viewing photographic images of real faces that include natural variability in appearance.

Perception versus Memory

Weigelt et al. (2014) suggest a developmental dissociation for face perception and face memory, with no face-specific perceptual development after 5 years of age but continuing development of face-specific memory until age 10 (see also Crookes & Robbins, 2014; McKone et al., 2012). Our findings provide evidence that face-specific perceptual development might well continue throughout childhood; when tested with unfamiliar identities, performance improved between 6 and 12 years of age (see also Megreya & Bindemann, 2015).

Our task was designed to be perceptual in nature; memory demands were completely eliminated (although children for whom the target face was familiar were able to rely on previous experience with their teacher’s face). The house had a photograph of the teacher on the front,
allowing children to compare each photograph in the pile with the photograph on the house.

However, whereas previous studies required children to match nearly identical images (e.g., that differed only in viewpoint or emotional expression and were taken on the same day under similar lighting conditions), our task required them to recognize an identity across naturally varying images. Burton (2013) argues that identity recognition—the ability to recognize identity in natural, rather than tightly controlled, images—is a critical component of face recognition that has been largely ignored. As noted by Burton, “it is always easier to recognize a picture then to recognize the face” (p. 1469). Our findings suggest that when identity perception rather than image/picture perception is measured, development of identity perception of unfamiliar identities continues throughout childhood.

**Familiar versus Unfamiliar Face Recognition**

Our data indicate that the developmental trajectory for recognizing unfamiliar faces is different than that for familiar faces. The ability to recognize unfamiliar identities improved between 6 and 12 years age. In contrast, recognition of familiar identities continued to improve until 6 years of age, with no evidence for age-related improvement after that. This is consistent with Burton’s argument that there is an important difference between familiar and unfamiliar face recognition (Burton, 2013). In contrast to Weigelt et al. (2014) who found that memory for faces learned in the context of the experiment was close to chance performance in 5-year-olds, we found that familiar face recognition was good in 4- to 5-year-olds (a number of children performed without error), and (nearly) without error by 6 years of age. Unlike the children in Weigelt et al’s study, children tested with familiar identities in our study could base their decisions on preexisting knowledge of their teacher’s appearance in addition to the perceptual information available on each trial. It therefore seems that even very young children are able to build abstract representations of identity that are sufficient to
allow recognition even when viewing never-before-seen images, at least for identities with longstanding representations (Burton et al., 2005; Burton et al., 2011). An interesting direction for future research will be to find out whether children take longer than adults to acquire robust representations for faces. The fact that 4- and 5-year-olds in our study made errors even when sorting familiar faces suggests that they do.

**Domain-Specific vs. Domain-General Effects**

Performance on any perceptual task, including face recognition, is influenced by general cognitive factors (e.g., selective attention, concentration, visual development [e.g., Vernier acuity]) as well as domain-specific abilities. Crookes and Robbins (2014) take the strong view that “improvement with age observed on face tasks results entirely from general development rather than face-specific development” (page 104). We reasoned that asking children to sort 40 photographs, as in Jenkins et al. (2011), would be challenging as such a task requires that they keep track of multiple images in multiple piles. Thus we created a child-friendly version of the Jenkins task in which children made a binary decision on each trial. We do not deny that general cognitive development might influence performance on this task. However, several pieces of evidence suggest that its influence on our key findings is minimal. First, children performed without error when given an image that was an exact match to the image mounted on the house for both familiar and unfamiliar identities. They also performed without error when given an image of a highly dissimilar identity. These control stimuli (four each) were evenly distributed across blocks, suggesting no decline in attention or concentration—at least on these relatively easy trials. Second, failure to discriminate (the number of false alarms for the matched identity) did not vary as a function of age or across blocks, regardless of whether children were tested with a familiar or unfamiliar identity. Third, recognition did not decline across blocks and an analysis of accuracy during the first block of trials yielded the
same pattern of results as the analysis collapsed across all trials. We contend that our observed age-related improvements in identity recognition are, at least in part, attributable to face-specific effects.

**Theoretical Implications**

Our data call for new ways of thinking about the development of face recognition that incorporate within-person variability. The ability to recognize identity despite within-person variability in appearance can be conceptualized in the light of recent extensions of Valentine's (1991) influential model, in which each individual face is represented as a unique point in a multidimensional face space. The location of each face is determined by its values on the dimensions underlying face space, along which faces vary. These characteristics might be features and their spacing (e.g., nose length, distance between the eyes,) or more abstract dimensions (e.g., eigenfaces; Hancock, Burton & Bruce, 1996). The dimensions of face space are refined through perceptual experience to represent the facial properties that are optimal for discriminating identities from highly familiar categories (see O'Toole & Natu, 2013 for a discussion).

Recent extensions of this model take into account natural variability in appearance by incorporating attractor fields/identity regions. Attractor fields around each point in face space (Tanaka & Corneille, 2007; Tanaka, Giles, Kremen & Simon, 1998; see also Lewis & Johnston, 1999) reflect the range of inputs that are perceived as belonging to a given identity, allowing recognition despite changes in appearance. For example, all recognizable images of our neighbour would be represented within her attractor field. The size of an identity’s attractor field is constrained by the density of nearby representations (i.e., by its location in face space). This was directly shown by Tanaka and colleagues (Tanaka et al., 1998; Tanaka & Corneille, 2007; Tanaka, Meixner & Kantner, 2011) who reported that 50/50 morphs between two “parent” faces are perceived as more
closely resembling the distinctive parent than the typical parent, consistent with distinctive faces having bigger attractor fields because they reside in a less populated area of face space (but also see Burton & Vokey, 1998).

Development has been characterized in terms of the refinement of face space (Short et al., 2014)—an increased sensitivity to differences along the dimensions of face space (e.g., Anzures et al., 2009; Jeffery et al., 2010) and an improved ability to rely on multiple dimensions simultaneously (Nishimura, Maurer & Gao, 2009; Tanaka et al., 2014). As noted by Johnston & Ellis (1995), adding exemplars to face space, as happens during childhood, likely increases inter-face distances; one obvious implication of this is improved discrimination. Our findings suggest that increased sensitivity to the dimensions along which faces vary (i.e., larger inter-face distances) also increases children’s ability to recognize identity when appearance varies. One way to conceptualize development, then, is in terms of increases in the size of attractor fields.

This explanation is consistent with evidence that adults’ face memory and their ability to recognize an unfamiliar identity across natural variation in appearance are both affected by experience. Adults who encountered fewer faces during childhood by virtue of growing up in a rural environment make more errors on a face memory task than adults who grew up in an urban environment (Balas & Seville, 2015) perhaps because they have a more compact face space with smaller attractor fields. Likewise, adults make more piles when sorting other-race faces than own-race faces (Laurence et al., 2015), suggesting that limited experience with other-race faces leads to those faces having smaller attractor fields. Just as variability in input facilitates language development, so too does it affect the refinement of face space (see Watson, Robbins, & Best, 2014).

Limitations & Future Directions
One possible limitation of this study was that only teachers’ faces were used as familiar stimuli. Although each child tested with a familiar teacher had known their teacher for a minimum of 3 months (with no difference in performance between children who had known their teacher for 3 versus 9 months), it is possible that children would show adult-like tolerance to variability for their caregiver’s face due to children’s early advantage with these faces. Newborn infants show a preference for their mother’s face (Pascalis, De Schonen, Morton, Deruelle, & Fabre-Grenet, 1995; Bushnell, 2001), and older infants show adult-like patterns of recognition for their mother’s face, such as recognition despite geometric distortions (Yamashita, Kanazawa & Yamaguchi, 2014) and masking of external features (Bartrip, Morton, & de Schonen, 2001), as well as inversion effects (Balas et al. 2009). In short, representations for caregivers might be more robust than for other familiar faces due to them being highly overlearned (Tong & Nakayama, 1999).

A second possible limitation is that we did not present non-face stimuli to our participants. Strictly speaking, this means that we cannot conclude whether the age-related improvements we observed are face specific. One of the challenges in addressing this question is to find appropriate control stimuli. Weigelt et al. (2014) compared perception of and memory for faces to three other stimulus categories that shared critical characteristics with faces: bodies (another social stimulus), cars (another object class with reasonably consistent first-order relations), and scenes (another stimulus class associated with discrete brain regions). Cars and scenes both differ from faces in a critical way: Neither varies in appearance in a way that is analogous to faces. Whereas the appearance of an individual face varies from moment to moment due to both internal factors (e.g., emotional expressions, direction of eye gaze, makeup) and external factors (e.g., camera angle, lighting conditions) the appearance of both cars and scenes is more stable. Although bodies can change in appearance from moment to moment (e.g., as a person moves) there is no evidence of
expertise in individual body identity to our knowledge. The special challenge associated with identity recognition is the ability to recognize identity in natural images in which appearance varies; we know of no other stimulus class for which this is true. Whether experts with other biological categories (e.g., dog experts) show this extraordinary ability to recognize individual exemplars across variation in appearance remains to be seen.

To date, only one previous study has attempted to examine developmental changes in the size of attractor fields; however, this study investigated face recognition in infants rather than across childhood. Humphreys and Johnson (2007) habituated 4- and 7-month-old infants to one identity and then presented them with a morphed continuum in which that identity was morphed with another face. Younger infants were less sensitive to differences between morphed faces than older infants who, in turn, were less sensitive than adults; Humphreys and Johnson concluded that the size of attractor fields decreases during infancy. However two characteristics of their study warrant caution. First, each identity was represented by a single image; therefore, the range of variability represented in their stimuli was not representative of real world variability. Second, studies in which a single stimulus is presented during habituation are best suited to measure discrimination rather than recognition of an object/category across variability in appearance. To measure infants’ ability to recognize identity across variability in appearance (i.e., their ability to form a category ‘Alice’), a method akin to that used to study categorization is needed whereby infants are habituated to multiple exemplars of one identity and then tested with a novel image of the familiar identity versus a novel identity (see Quinn, 2002 for an example). An interesting direction for future research will be to find out if infants, like children, will show increasing tolerance for natural variability in familiar and unfamiliar faces (see Watson et al., 2014 for a discussion).
Finally our task had a different structure than that of Jenkins et al. (2011). An advantage of our task is that it is likely less cognitively demanding because children did not need to keep track of 40 photographs simultaneously; rather, they made a yes/no decision for each image. Nonetheless, an advantage of Jenkins’ sorting task is that participants are free to make as many or as few piles as they like. Future research should examine whether these two tasks yield similar outcomes.

Summary

In summary, the current study used a child-friendly face-sorting task to investigate children’s ability to recognise an identity in images containing within-person variability. Overall, our findings show that children’s recognition of both familiar and unfamiliar faces improves with age, with adult-like performance achieved at a younger age for familiar faces. Their improvement was driven both by their ability to discriminate between two identities and their ability to recognise a single identity in images containing within-person variability, an ability that has been ignored in the vast majority of previous work. These findings support a framework where face recognition is not yet adult-like during childhood.
Acknowledgements

Removed for the purpose of blind review.
References


CHILDREN’S FACE RECOGNITION


CHILDREN’S FACE RECOGNITION


CHILDREN’S FACE RECOGNITION


CHILDREN’S FACE RECOGNITION


CHILDREN’S FACE RECOGNITION

*Journal of Experimental Child Psychology, 108*(2), 338-357.
doi:10.1016/j.jecp.2010.07.007


Tanaka, J. W., Quinn, P. C., Xu, B., Maynard, K., Huxtable, N., Lee, K., & Pascalis, O. (2014). The effects of information type (features vs. configuration) and location (eyes vs. mouth)


Highlights

- We investigated age-related changes in the ability to recognize faces across natural variation in appearance
- 4- to 12-year-olds were asked to find all the images of a familiar or unfamiliar identity
- Performance was (nearly) perfect by 6 years for familiar identities, but improved across the entire age range for unfamiliar identities
- Findings have implications for models of the development of face perception