

Open Research Online

The Open University's repository of research publications and other research outputs

How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity

Journal Item

How to cite:

Baker, Kristen A.; Laurence, Sarah and Mondloch, Catherine J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, 161 pp. 19–30.

For guidance on citations see [FAQs](#).

© 2016 Elsevier B.V.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1016/j.cognition.2016.12.012>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity

Kristen Baker¹, Sarah Laurence^{1,2}, Catherine J. Mondloch¹

¹Department of Psychology, Brock University, St Catharines, Canada

²School of Psychology, Keele University, UK

Address Correspondence To:

Catherine J. Mondloch
Department of Psychology
Brock University
1812 Sir Isaac Brock Way
St. Catharines, Ontario
Canada L2S 3A1
E-mail: cmondloch@brocku.ca
Phone: (905) 688-5550 ext. 5111

Abstract

Adults and children aged 6 years and older easily recognize multiple images of a familiar face but often perceive two images of an unfamiliar face as belonging to different identities. Here we examined the process by which a newly encountered face becomes familiar, defined as accurate recognition of multiple images that capture natural within-person variability in appearance. In Experiment 1 we examined whether exposure to within-person variability in appearance helps children learn a new face. Children aged 6 to 13 years watched a 10-minute video of a woman reading a story; she was filmed on a single day (low variability) or over three days, across which her appearance and filming conditions (e.g., camera, lighting) varied (high variability). After familiarization, participants sorted a set of images comprising novel images of the target identity intermixed with distractors. Compared to participants who received no familiarization, children showed evidence of learning only in the high-variability condition, in contrast to adults who showed evidence of learning in both the low- and high-variability conditions. Experiment 2 highlighted the efficiency with which adults learn a new face; their accuracy was comparable across training conditions despite variability in duration (1 vs. 10 minutes) and type (video vs. static images) of training. Collectively, our findings show that exposure to variability leads to the formation of a robust representation of facial identity, consistent with perceptual learning in other domains (e.g., language), and that the development of face learning is protracted throughout childhood. We discuss possible underlying mechanisms.

Key Words: Face Recognition; Children's Face Learning; Perceptual Development; Within-Person Variability; Perceptual Learning

How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity

Two pictures of the same person can look very different and pictures of two different people can look very similar. Thus, accurate person recognition requires both discrimination (telling people apart) and identity matching (recognizing a person when his/her appearance changes). Despite over 30 years of psychological research aimed at understanding face recognition, it is only relatively recently that the challenge of recognizing identity despite within-person variability in appearance (resulting from changes in hairstyle, make-up, lighting, point of view, camera angle/distance) has been brought to the forefront of face recognition research (Burton, 2013). This has allowed parallels to be drawn between faces and other domains (e.g. language [see Watson, Robbins, & Best, 2014]) in which understanding within-exemplar variability has received attention. In addition to its broad theoretical implications, understanding face recognition across changes in appearance represents a challenge faced in daily interactions (e.g., for recognizing our colleague when he shaves his beard), in the security industry (i.e., for determining whether the identity in a photograph matches that of the person holding the passport) and in eyewitness testimony (e.g., for recognizing someone we saw commit a crime in a photo line-up).

Understanding the effects of within-person variability on identity perception is central to understanding the difference between familiar and unfamiliar face recognition. We can easily recognize hundreds of images of famous people or those with whom we are personally familiar. In contrast even a small change in appearance can impair our recognition of unfamiliar faces. For example, accuracy in a 1-in-10 task, in which sample

and target photos were taken with different cameras, was only 70% despite the photos being taken on the same day, from the same viewing angle, and with a neutral expression (Bruce, Henderson, Greenwood, Hancock, Burton, & Miller, 1999; Megreya & Bindemann, 2015).

A seminal paper by Jenkins and colleagues most clearly demonstrated how familiarity influences recognition of identity in ambient images (Jenkins, White, Van Montfort, & Burton, 2011). Participants were asked to sort a stack of 40 photographs (20 photos of two identities) into piles such that each pile contained all of the images of a single identity; participants were not informed about the number of identities present. When the identities were familiar, participants performed without error (i.e., they accurately perceived that only two identities were present). In contrast, when the identities were unfamiliar, participants perceived an average of six different identities. The impact of familiarity is even stronger for other-race faces: Although adults make twice as many piles when sorting unfamiliar other-race faces than when sorting unfamiliar own-race faces, the own-race advantage is eliminated (i.e., performance is perfect) when sorting familiar faces (Zhou & Mondloch, 2016). These results are attributable to familiar faces having a sufficiently robust representation to allow recognition across a range of inputs; recognition of unfamiliar faces relies more on lower level image properties and is heavily tied to a specific instance (see Burton, Jenkins, Hancock, & White, 2005; Burton, Jenkins, & Schweinberger, 2011; Hancock, Bruce, & Burton, 2000; Jenkins et al., 2011; Johnston & Edmonds, 2009).

A current hot topic in the field of face recognition, then, is how does recognition of a newly encountered face make the transition from image dependent (unfamiliar) to robust

(familiar)? Recent evidence from adult participants suggests that exposure to the way in which a particular face varies is key to the formation of a robust representation of that face (Andrews, Jenkins, Cursiter & Burton, 2015; Bindemann & Sandford, 2011; Dowsett, Sandford & Burton, 2016; Menon, White & Kemp, 2015b; Ritchie & Burton, 2016). Menon, White, & Kemp (2015a) showed participants a pair of images.

Participants were told that the two sample images either belonged to two different people or (correctly) to the same person. The task was to decide whether a third image matched the identity of one (2-person condition) or both (1-person condition) sample images.

Accuracy was higher in the 1-person condition, suggesting that knowing how a face can vary in appearance facilitates recognition of a new instance. Likewise, recognizing new images of learned identities is more accurate after studying 10 images with high variability in appearance than after studying 10 images with low variability in appearance (Ritchie & Burton) and finding a target identity in a 30-image lineup becomes easier if the to-be-matched sample comprises six images rather than a single image (Dowsett et al.). Collectively, these studies show that as new instances are encountered a robust representation develops. The more variability incorporated in a representation, the greater the likelihood that a novel instance will be recognized (Burton, Kramer, Ritchie, & Jenkins, 2016). That variability is a route to learning is consistent with variability leading to optimal training of perceptual expertise in other domains (for detecting dangerous items in luggage, Gonzalez & Madhavan, 2011; texture discrimination, Hussain, Bennett, & Sekuler, 2012).

The Development of Face Recognition

Numerous studies have investigated the development of expert face recognition and

its underlying mechanisms. These studies have greatly advanced our understanding of how children discriminate faces (tell people apart): They present children with identical (e.g., Baudouin, Gally, Durand & Robichon, 2010; Gilchrist & McKone, 2003; Macchi Cassia, Luo, Pisacane, Li & Lee, 2014; McKone & Boyer, 2006; Mondloch & Thomson, 2008; Mondloch, Le Grand & Maurer, 2002; Pellicano & Rhodes, 2003; Pellicano, Rhodes & Peters, 2006; Tanaka, Kay, Grinnell, Stansfield, & Szechter, 1998) or nearly identical (Bruce et al., 2000; Megreya & Bindemann, 2015; Mondloch, Geldart, Maurer, & Le Grand, 2003) images of unfamiliar faces at study and test. The same is true of the few studies that have examined children's ability to recognize personally familiar faces (Bonner & Burton, 2004; Ge et al., 2008; Mondloch & Thompson, 2008; Newcombe & Lie, 1995; Wilson, Blades, Coleman, & Pascalis, 2009). Very little is known about the development of the other central component of face recognition: Children's ability to recognize identity in images that capture natural variability in appearance and the process by which faces become familiar during childhood.

Laurence and Mondloch (2016) adapted Jenkins et al.'s (2011) protocol to provide the first examination of children's ability to recognize a face's identity across a set of images that incorporate natural variability. They presented children with a toy house on which a single photo of a target identity was mounted. That identity was either highly familiar (the child's own teacher) or wholly unfamiliar (a teacher from a different school). Children were provided with a stack of photographs that included nine novel images of the target identity and nine different images of a similar-looking distractor (plus control stimuli). These images were presented sequentially to children, who were asked to place all of the images of the target into the house but to keep everyone else out.

When tested with an unfamiliar identity, performance improved between 5 and 12 years of age. When tested with a familiar identity (i.e., their own teacher), children aged 6 years and older performed (nearly) without error; however, several 4- and 5 year-olds made multiple errors despite knowing their teacher for several months. These results suggest that by age 6 years children, like adults, are able to build robust representations of identity that allow recognition even when viewing never-before-seen images, at least for identities with long-standing representations (Burton et al., 2005, 2011). However, children knew their teacher for a minimum of 3 and as many as 9 months, and so what remains unknown is the process by which a face becomes familiar during childhood and whether the ability to use variability in appearance to form a representation changes between 6 and 12 years of age. Given that exposure to variability facilitates learning in other domains early in development (e.g. early word learning; Rost & McMurray, 2009; Singh, 2008), and exposure to within-person variability in appearance facilitates adults' face learning (see Burton et al., 2016 for a discussion), systematically varying the amount of variability to which children are exposed might influence their ability to build a robust representation for a newly encountered face.

Here we directly investigated the contribution of exposure to variability in appearance by familiarizing children aged 6 to 13 years with a target identity, an age range over which the ability to recognize an unfamiliar identity despite variability in appearance continues to improve (Laurence & Mondloch, 2016). We endeavoured to maximize children's opportunity to learn by presenting each child with a 10-minute video in which one of three target identities read a children's storybook. Each of the three models was filmed reading the identical story on three separate days and across days we

altered the target's appearance (hair, make-up), lighting, and the camera used for recording. Each child watched the video of one target. We manipulated variability by presenting the video as it was filmed on a single day (low-variability condition) or as it was filmed across the three days (high-variability condition), which we did by splicing each video into three segments and creating various combinations. Children in a no-training control group did not watch the video. After watching the video (or not) all children completed the sorting task designed by Laurence and Mondloch (2016). We hypothesized that performance would increase with age and that children in the training conditions would perform more accurately than children in the no-training control group. Most notably we predicted that children in the low-variability condition would show a reduced benefit of training, consistent with evidence from adults that exposure to higher variability enhances learning (see above). We also hypothesized that the effectiveness of training would vary with age, with older children benefitting more than young children. We based the latter hypothesis on two lines of evidence. Firstly, 4- and 5-year-olds are less able than older children to build a representation of a familiar teacher (many of these children made errors; see Laurence & Mondloch, 201). Secondly, a face starts out as unfamiliar before any learning takes place. Laurence and Mondloch (2016) found that unfamiliar face recognition was worse in younger children than in older children. As a result, for younger children to learn newly encountered face, they must overcome a greater transition (from unfamiliar to familiar) than older children.

Prior to testing children we tested a group of adult participants to ensure that our training was effective. We hypothesized that adults in the training groups would perform more accurately than adults in the no-training group and that adults in the high-variability

training group would perform more accurately than those in the low-variability training group (see Menon et al., 2015a; Ritchie & Burton, 2016).

The results of Experiment 1 suggest that children are less able than adults to use minimal variability in appearance to form a robust representation of a newly encountered identity. To further examine face learning in adults in Experiment 2 we tested MTurk workers under three different training conditions that, again, incorporated both high and low variability in appearance: 10-minute videos, 1-minute videos (to examine the effect of duration of exposure), or a 1-minute slide show comprised of 12 stills extracted from the videos (to examine the effect of motion). Collectively these two studies provide novel insights about the process by which a newly encountered face becomes familiar.

Adults and children performed a child-friendly sorting task after no training, low-variability training (watching a video as it was filmed on a single day) or high-variability training (watching a video as it was filmed across three days). We examined the effectiveness of training by measuring both hits (number of images of the model that were recognized) and false alarms (number of images of the similar-looking distractor that were falsely recognized). Prior to administering the task we familiarized participants with the protocol by giving them a pile of photographs comprising five different images of Buzz Lightyear and four images of Noddy. Participants were asked to place all of the images of Buzz into his spaceship while keeping everyone else out.

Experiment 1

Methods

Participants. 108 children (40 male) aged 6 to 13 years ($M = 110.90$ months, range = 72 to 162; see Table 1 for participant ages in each condition) and 108 adults (20 male; $M = 20.23$ years, range = 17 to 29) from southern Ontario were tested. All participants were Caucasian and were unfamiliar with the identities used in this experiment. Before the start of the experiment, written consent was obtained from adult participants and the parents or guardians of each child; verbal assent was obtained from each child. Following the completion of the experiment all child participants received a small toy and a certificate and adult participants were compensated for their time with psychology course participation credit or \$5. An additional 11 children and 14 adults were excluded from the analyses because they failed control trials ($n = 9$ children and 6 adults) or misidentified the model as someone they knew ($n = 2$ children and 2 adults) or because of experimenter error ($n = 6$ adults).

Table 1.

Age (in years) of children tested in each condition ($n = 36$ /condition)

	HV	LV	NT
<i>M_{age}</i>	8.89	8.69	8.86
<i>SD</i>	2.01	2.01	2.18
<i>range</i>	6-13	6-13	6-13

Note. Mean age for children in each condition. HV = High-Variability training; LV = Low-Variability training; NT = No training.

Materials. Seven Caucasian women donated their photographs to be used as stimuli. Three of them agreed to be videotaped reading a story on three separate days and thus served as target identities. Each participant was tested with one of the three target identities; participants in the control condition were unfamiliar with the identity whereas participants in the training conditions were familiarized with the identity prior to the test phase.

Training stimuli. We made three separate videos of each model reading Chapter 1 from the children's story "About Teddy Robinson" (Robinson, 1975). The videos for each model were made on three different days and across days we varied the setting (in a building containing some natural light; in a basement with very limited natural light; outdoors), models were instructed to alter their hairstyle, outfit, and makeup, and we used a different video camera (Sony Cyber-shot camera, a Rebel E03 T3i Cannon and a Blackberry Q10). Each video was 10 minutes long.

Participants in the low-variability (LV) training condition watched a video that featured one of the three models as it was recorded on one day (e.g. Day 1, Day 2, or Day 3). The LV condition exposed participants to some variability in the model's appearance (i.e., expression changes, head movements, etc.) but lighting, camera and the model's general appearance (e.g., hairstyle, make-up, clothing) was held constant. Participants in the high-variability (HV) training condition also saw a video that featured one of the three models; however, to increase variability in appearance we spliced the original videos into three separate subsections (i.e., the beginning, middle and end) that were approximately 3.5 minutes in length. The HV videos included one section from each of the three days (e.g., the beginning section from Day 1, the middle from Day 2, and the

end from Day 3). Thus the HV condition exposed participants to variability in the model's appearance associated with intrinsic factors (e.g., make-up), camera, and lighting. We created three HV videos for each model such that each segment was included in one high-variability video.

In short, we created three LV and three HV videos for each model (18 10-minute videos in total); their use was counterbalanced across participants in each condition. The video was shown on a LG monitor that was 21 in. by 12 in. in size. Participants sat 55cm in front of the monitor and the size of the model's head in the video varied across recordings with height of face ranging from 7.5 to 12.7cm.

Sorting stimuli. To measure facial recognition despite within-person variability in appearance, participants performed a child-friendly version of the sorting task (Laurence & Mondloch, 2016), in which 32 images in total were used. All photographs had a roughly frontal view of the model's face and each image was edited in Photoshop to be 50 mm by 70 mm in size and formatted in grey-scale (as in Jenkins et al., 2011). Each participant was shown a toy house on which we mounted one image of the target identity. The sorting task contained 18 test stimuli—nine novel images of the target identity and nine images of a similar looking distractor (similar age, hair color, face shape). These images were provided by the models and their distractors and comprised naturalistic photographs taken on different days and different occasions; hairstyle, make-up, expressions, camera, and lighting all varied. The same test images were shown to all participants. The sorting task also included 14 control stimuli. To ensure that participants understood the task we included four control stimuli comprising images of the target identity identical to that mounted on the toy house and four images of a dissimilar

looking distractor (different age and hair style). Participants were required to include all control images of the target identity and to exclude all images of the dissimilar distractor to be included in our analyses; four adults and three children failed to do so. To verify that participants had attended during the learning phase we included six images extracted from the video (two from each of the three segments). To be included in our analyses, adults in the training conditions were required to include five of these images and children were required to include four. Two adults and six children failed to do so and were excluded from analyses; 69 adults and 61 children included all six of these extracted images and four children included five of them¹. Only seven children put in four of the extracted images, all of whom were in the HV condition. Children in NT condition for whom these extracted images were unfamiliar put in a mean of 0.36 images (Mode = 0; Median = 0).

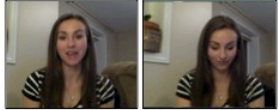
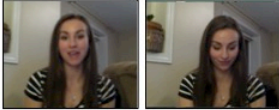

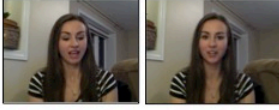
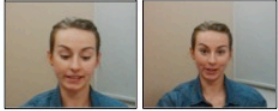
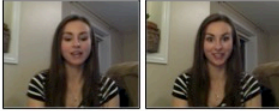
Buzz Lightyear training task stimuli. Five different images of Buzz Lightyear and four images of Noddy were selected from an Internet search, with one image of Buzz selected to go on the front of a spaceship (see Laurence & Mondloch, 2016).

Procedure. Prior to starting the task participants were randomly assigned to one of three conditions (HV training, LV training or no-training [NT], $n = 36$ per condition) and, within each condition, to one of the three target identities. Participants in the HV and LV training conditions were informed that they would be watching a video of a woman named Alice who was going to read a story (see Figure 1). After watching the video, participants were asked whether they had ever seen Alice or heard the story from the

¹ We excluded the two adults who only included four images extracted from the video because they appeared to be outliers. We lowered the criterion for children because so many children only included four images.

video before.

Similar to the design developed by Laurence and Mondloch (2016), participants were then introduced to the sorting task by performing a training task involving Buzz Lightyear. The participants were shown Buzz Lightyear's spaceship (on which there was an image of Buzz Lightyear) and a stack of images including Buzz ($n = 5$, one identical to that on the spaceship and four novel images) and Noddy ($n = 4$). The researcher explained that Buzz's photographs were mixed together with photos of another person and asked participants to put all of the photos of Buzz into the spaceship, but to be sure to exclude photographs of anyone else.

A.	High Variability	Low Variability
Beginning	<i>Day 1</i> 	<i>Day 1</i> 
Middle	<i>Day 2</i> 	<i>Day 1</i> 
End	<i>Day 3</i> 	<i>Day 1</i> 



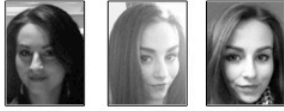

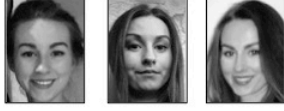

B. Test Stimuli	
Target	Similar Distractor
	
	
	

Figure 1. Panel A: A depiction of the variability provided in our high- (left) and low- (right) variability conditions; two stills are shown from each video. Panel B: Nine novel images of the model from the video (left) and nine novel images of the similar distractor.

After completing the Buzz Lightyear task, participants were presented with a toy house (on which there was an image of the target identity) and the stack of 32 images (containing images of the identity from the video and the distractors). The researcher explained that it was Alice's (the woman from the video) house, that there was another image of Alice on the front, and that Alice looked somewhat different from day to day. The researcher then asked the participants to put all of the photos of Alice into the house while keeping everyone else out. Participants were then sequentially shown each photo and asked to decide whether it belonged to Alice.

Analyses. Preliminary analyses revealed that the effect of condition did not vary across models; thus, we collapsed across models in all analyses. Participant accuracy was analyzed using signal detection theory (d'). A hit was defined as correctly putting an image of the target identity into the house and a false alarm was defined as putting an image of the similar distractor model in the house. Analyses were based only on the 18 test stimuli. Control stimuli were used as exclusion criterion only. We analyzed d' , hits and false alarms separately. Perfect performance would be reflected in nine hits, zero false alarms and $d' = 3.19$.

We first analyzed adults' performance to verify that our training protocol was effective by conducting a one-way ANOVA with three levels (NT, LV training, HV training). The significant effect was examined with contrast analyses because of our a priori predictions. A comparable analysis was conducted for children as a group. We analyzed children's data separately from that of adults because they were our primary focus and doing so allowed us to take advantage of our wide age range to examine whether the effect of training was moderated by children's age. To examine this question

we conducted a regression analysis in which variability was dummy coded in order to make two separate contrasts: LV vs. NT and HV vs. NT.

Results

Adults. The one-way ANOVA showed a significant effect of condition for d' , $F(2, 105) = 4.314, p = 0.02, \eta^2 = 0.076$. Contrast analyses revealed that participants in both the HV condition ($M_{d'} = 2.16, SD = 0.64$) and the LV condition ($M_{d'} = 2.01, SD = 0.89$) were more accurate than those in the NT condition ($M_{d'} = 1.64, SD = 0.75$), $t(105) = 2.84, p = 0.005$; $t(105) = 2.05, p = 0.04$, respectively. Accuracy did not differ between the HV and LV conditions, $p = 0.43$ (see Figure 2a).

Separate analyses of hits and false alarms revealed that the number of hits varied across conditions, $F(2, 105) = 5.54, p = 0.005, \eta^2 = 0.096$ (see Figure 2b). Contrast analyses revealed that participants in both the HV condition ($M_{hits} = 6.78, SD = 1.71$) and the LV condition ($M_{hits} = 6.53, SD = 2.08$) made more hits than those in the NT ($M_{hits} = 5.39, SD = 1.85$) condition, $t(105) = 3.12, p = 0.002$; $t(105) = 2.56, p = 0.01$, respectively. The number of hits did not differ between the HV and LV conditions, $p = 0.75$. Thus training improved adults' ability to recognize new images of the target identity. False alarms were rare across all conditions and did not vary among them, $F(2, 105) = 0.18, p = 0.83, \eta^2 = 0.004$ (Figure 2c).

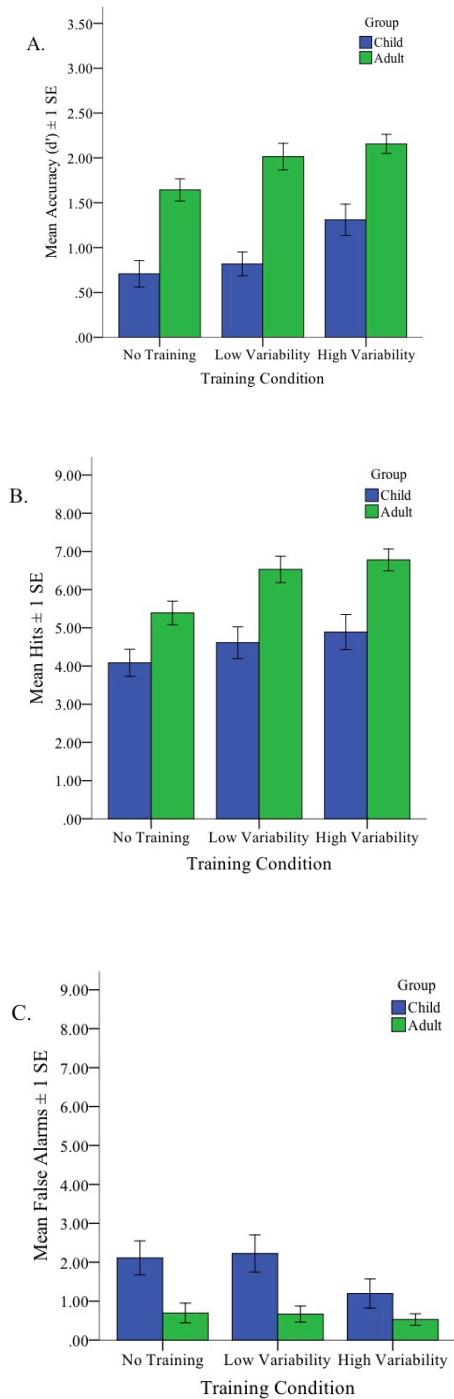


Figure 2. Mean accuracy for child and adult participants in the three experimental conditions of Experiment 1: No-training control, low-variability training, and high-variability training. Three dependent variables are shown: d' (A), hits (B), and false alarms (C). Error bars show \pm SEM.

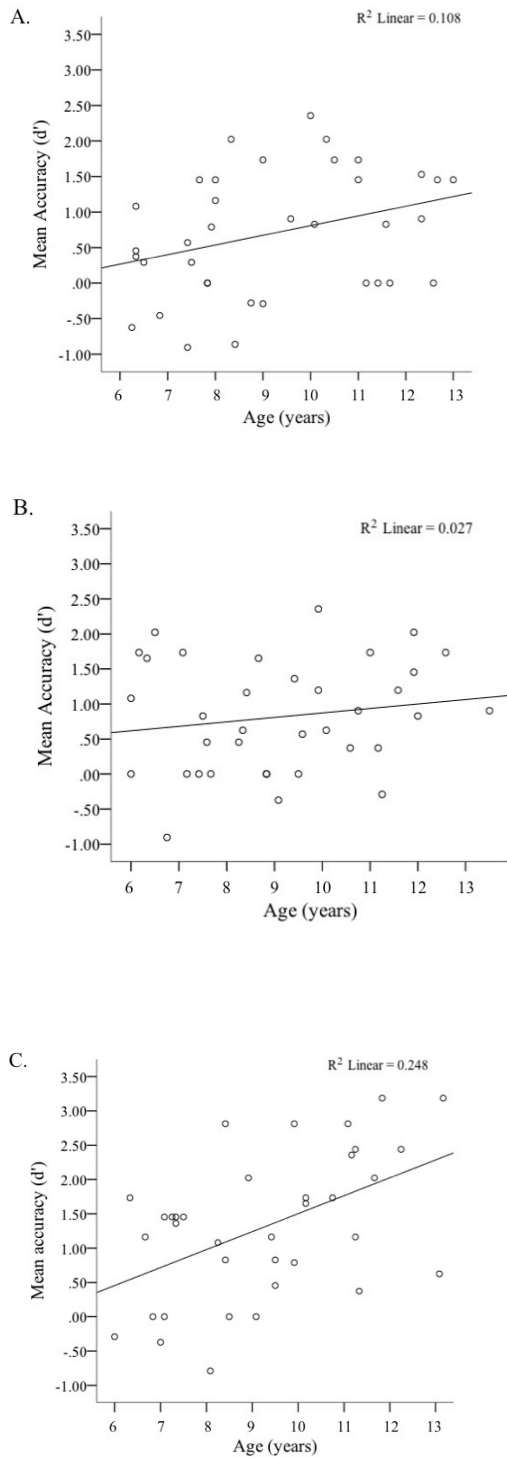


Figure 3. d' as a function of age. Each dot represents a single child participant in No-training (A), Low-variability training (B), and High-variability training (C).

Table 2

	<i>B</i>	<i>B SE</i>	β	<i>p</i>	<i>R</i> ²
Step 1					0.185
Constant	0.703	0.144	-	<i>p</i> < 0.001	
Age (months)	0.013	0.003	0.328	<i>p</i> < 0.001	
LV vs. NT	0.127	0.204	0.063	<i>p</i> = 0.537	
HV vs. NT	0.600	0.204	0.301	<i>p</i> = 0.004	
Step 2					0.215
Constant	0.703	0.143	-	<i>p</i> < 0.001	
Age (months)	0.011	0.006	0.295	<i>p</i> = 0.05	
LV vs. NT	0.120	0.202	0.060	<i>p</i> =0.555	
HV vs. NT	0.594	0.202	0.298	<i>p</i> =0.004	
Age x LV vs. NT	-0.006	0.008	-0.091	<i>p</i> =0.461	
Age x HV vs. NT	0.010	0.008	0.152	<i>p</i> =0.211	

Note. Step one depicts age and variability conditions as separate predictors. Step two depicts the interaction terms for Age and variability conditions.

Children. A one-way ANOVA revealed a significant effect of condition for children's accuracy, $F(2, 105) = 4.41, p = 0.015, \eta^2 = 0.077$. Contrast analyses revealed that children in the HV condition were more accurate ($m_{d'} = 1.35; SD = 1.04$) than children in both the LV ($m_{d'} = 0.82; SD = 0.798$), $t(105) = 2.275, p = 0.025$, and NT ($m_{d'} = 0.71; SD = 0.88$) conditions, $t(105) = 2.789, p = 0.006$. Accuracy of children in the LV condition did not differ from that of children in the NT condition, $p = 0.61$. Separate analyses of hits and false alarms revealed no effect of condition on either variable (hits: $F(2, 105) = 0.985, p = 0.377, \eta^2 = 0.018$; false alarms: $F(2, 105) = 1.705, p = 0.187, \eta^2 = 0.031$), suggesting that a combination of these two indexes of recognition contributed to the overall effect on accuracy.

Three separate Pearson r correlations were conducted to look at the relationship between age and accuracy within each training condition (see Figure 3). Accuracy was positively correlated with age in the NT ($r = 0.328, p = 0.051$) and HV ($r = 0.498, p = 0.002$) conditions, but not in the LV condition ($r = 0.165, p = 0.337$).

To investigate whether the effect of training was moderated by age we conducted linear regression analyses. Age and the two variability contrasts (LV vs. NT; HV vs. NT) were entered in the first step, and two interaction terms (LV vs. NT x age; HV vs. NT x age) were entered in the second step. The first step was significant, $F(3, 104) = 7.851, p < 0.001$. Age predicted d' ($t(102) = 3.698, p < 0.001, 95\% \text{ CI } [0.006, 0.019]$) and, consistent with the ANOVA, the HV vs. NT contrast was significant ($t(102) = 2.941, p = 0.004, 95\% \text{ CI } [0.195, 1.004]; pr = 0.111$), whereas the LV vs. NT contrast was not ($t(102) = 0.620, p = 0.537$). Critically, the second step was not found to account for any additional variance, $\Delta F(2, 102) = 1.969, p = 0.145; \Delta R^2 = 0.030$, and neither of the

interaction terms were significant predictors of child performance ($p > 0.05$; see Table 2). Thus, we found no evidence that the effect of training was moderated by age.

Discussion

Laurence and Mondloch (2016) reported that by age 6 years children are capable of forming a robust representation of a familiar face that allows for accurate recognition despite variability in appearance. However, because children in that study were tested with a highly familiar person (their own teacher) with whom they had interacted for several months it remained unknown whether children's face learning is comparable to that of adults. Here we provided the first examination of the process by which children build robust representations for faces.

Our results showed that children, like adults, are able to form a robust representation of a face by being exposed to variability in its appearance; children in the HV condition were more accurate than children in the NT condition. Previous research showed that adults profit from high variability; they recognized new instances of a face after viewing images collected across different days than after viewing the same number of images collected on a single day (Ritchie & Burton, 2016; see also Murphy, Ipser, Gaigg & Cook, 2015). The benefit of high variability was even more apparent in our child participants. Unlike adults, children in the LV condition showed no evidence of being able to generalize to novel instances; they performed less accurately than children in the HV condition and did not differ from children in the NT condition. The failure of children to learn a new face in the LV condition cannot be attributed to a lack of attention; our control trials show that children of all ages had attended to the video as they were able to recognize images of the target that were extracted from the video.

Although seven children recognized only four of the six video stills, they were all in the HV condition where face learning was evident. Thus only after exposure to variability in intrinsic appearance (e.g., make-up, hair), lighting, and camera were children able to learn a face well enough to recognize new instances more accurately.

We tested children across a wide age range (6 to 13 years). Our findings suggested that the development of face learning is protracted. Accuracy in both the NT and HV conditions increased with age, but accuracy in the LV condition did not. The difference between accuracy in the NT and LV conditions (or lack thereof) was not moderated by age. In short, any potential effectiveness of LV training was not masked by the performance of the younger children. However, our sample size precludes drawing strong conclusions about age-related changes in face learning during childhood; future research should include a larger sample and a wider age range (e.g., children between 13 and 17 years of age).

One strength of our design is the inclusion of three types of control stimuli. Including images identical to that mounted on the front of the toy house and images of a dissimilar distractor allowed us to exclude participants who either did not understand the task or were inattentive. Including six stills extracted from the video allowed us to assess participants' face memory and confirm attention during the familiarization period. Only two of 72 adults failed to place at least five of the six stills into the house, confirming that adults attended to the video. Because the two adults who included four images were outliers we replaced them, as our primary goal in testing adults was to confirm the effectiveness of our protocol. When testing children, we set an inclusion criterion to placing four video stills into the house. We note that seven children placed four (rather

than five or six) of the video stills into the house. These children ranged in age from 7 to 13 years and all of them were in the HV condition, the condition in which performance on the test stimuli was most accurate. Collectively, accuracy for our control stimuli suggest that children's performance was limited by their ability to incorporate the novel test images into their representation of the learned identity, rather than their comprehension of the task or ability to remember instances they viewed during the familiarization period.

Although our goal was not to compare the groups directly, children and adults showed different patterns of results. Whereas children only showed evidence of learning in the HV condition, adults showed evidence of learning in both the LV and HV conditions, with no difference between them. Although on the surface the lack of difference between adults in the LV vs. HV condition is surprising given evidence from previous studies that high variability is beneficial (e.g., Murphy et al., 2015), one possibility is that even our LV video incorporated more variability than multiple static images. To further address how readily adults are able to exploit variability to form a representation of a newly encountered face, we tested a large sample of adults in Experiment 2 and varied both the duration and type of exposure that they had to the target identity.

Experiment 2

MTurk workers completed the sorting task from Experiment 1. As in Experiment 1, some did so after watching a 10-minute video in either the HV or LV condition and others did so without any familiarization (NT). To push the limits of adults' learning we introduced two additional conditions. First, we manipulated the duration of exposure by

introducing a 1-minute video condition in which only the first minute of the story was presented, either as it was filmed on a single day (LV) or as it was filmed across three days (HV). Second, we manipulated the type of exposure by extracting a still image from each 5s epoch of the 1-minute video. Using these still images we created a 12-image slide show that was presented while the first minute of the story was read.

We hypothesized that, like adults tested in our lab in Experiment 1, adults in the 10-minute video conditions would learn the target identity, with little or no difference between the HV and LV conditions. Our central question was whether reducing the duration of exposure and/or replacing the video with a series of stills would impair learning and, if so, whether learning under these conditions would be enhanced in the HV condition relative to the LV condition.

Methods

Participants. 863 Caucasian adult MTurk workers from North America were tested and included in the final analysis (356 Male; $m_{age} = 37.65$, $SD_{age} = 11.57$, $range_{age} = 19-74$); 108 participants completed each condition (e.g., HV, 10-minute video) with the exception of the HV-no training condition that was completed by 107 participants. Participants were prescreened such that all participants reported being Caucasian and used a laptop or a desktop. An additional 414 participants were excluded from the final analysis for failing to pass criterion trials ($n=232$) or for failing an attention check while watching the video (see procedure; $n = 182$). Informed consent was obtained from each participant.

Materials.

Training stimuli. Training stimuli were formatted to be 642 pixels wide x 311 pixels

high. Participants in the training conditions watched either the low- or high-variability videos from Experiment 1 (10-minute groups) or revised videos that were 1 minute in length. The 1-minute video in the LV condition included the first minute from one day of filming (counterbalanced across participants) and in the HV condition included 20s from each of the three days (three versions, counterbalanced across participants). To create a 1-minute slide show we extracted one still image from each 5s epoch of each 1-minute video. Thus the slide show comprised a total of 12 images of the target identity, each of which was presented for 5 seconds.

Sorting stimuli. The same images that were used in Experiment 1 were also used in this experiment. Images were formatted and presented on the screen (W: 152 pixels/2.111 in, H: 199 pixels/ 2.674 inches) in size. As in Experiment 1, participants were required to include all four control images of the target identity and to exclude all four images of the dissimilar distractor to be included in the final analyses. 116 Mturk workers failed to do so and were excluded from analyses. To verify that participants had attended to the video we included six images extracted from the video (two from each of the three segments). Mturk workers were required to include five of these images; 116 failed to do so and were excluded from analyses. 827 of the Mturk workers in the training conditions included all of these images.

Procedure. As in Experiment 1, participants were randomly assigned into one of four training conditions (10-min, 1-min, stills, or NT), one of two types of variability (HV, LV) and to one of the three target identities. (To allow for a 4 x 2 design, two groups were in the NT condition; one was assigned to the HV condition and one to the LV condition. Thus, a difference in the effectiveness of HV and LV training might be

reflected in a main effect of training condition or a training condition x type of variability interaction.) For each of these combinations we posted three different videos (e.g., LV: Day 1, Day 2, Day 3) on MTurk and allowed 36 participants to complete each condition. Participants in the training conditions were asked to watch the assigned video throughout the familiarization period.

All participants performed a computerized version of the sorting task from Experiment 1. A photo of the target identity was presented in the top left hand corner (106 pixels wide x 147 pixels high) and remained visible while participants performed the sorting task. Test stimuli were presented one at a time in the center of the screen ($n = 32$); for each image participants were asked to press a button to indicate whether or not the image was a photo of the target identity. Following the sorting task, participants completed a short questionnaire confirming that they were unfamiliar with all of the identities and the story used in the experiment.

Results

Adults in the 10-minute video conditions performed nearly identically to the adults in Experiment 1. In congruence with Experiment 1, analyses of accuracy were conducted using signal detection theory (d'); only responses to the 18 test stimuli were included. To analyze the effect of training we conducted a 2 (Variability: high/low) x 4 (Training Type: 10-minute video, 1-minute video, 1-minute Slide Show, NT) ANOVA. The ANOVA revealed a small, but significant main effect of variability, $F(1, 855) = 4.997, p = 0.026, \eta^2 = 0.006$; accuracy was higher in the HV ($m_{d'} = 2.04, SD = 0.86463$) than the LV ($m_{d'} = 1.91, SD = 0.92$) condition. As shown in Figure 4a, there was also a main effect of training condition, $F(3, 855) = 11.842, p < 0.001, \eta^2 = 0.040$. A Bonferroni

Post-Hoc analysis revealed that participants in the NT ($m_d' = 1.67$, $SD = 0.84575$) condition were less accurate than participants in the 10-minute video condition ($m_d' = 2.08$, $SD = 0.87$, $p < 0.001$, 95% CI [-0.6345, -0.1870]), the 1-minute video condition ($m_d' = 2.12$, $SD = 0.92$, $p < 0.001$, 95% CI [-0.6707, -0.2232]), and the stills condition ($m_d' = 2.03$, $SD = 0.87521$, $p < 0.001$, 95% CI [-0.5841, -0.1365]). There were no significant differences between the three training conditions ($ps = 1.0$) and no variability x training type interaction, $F(3, 855) = 0.274$, $p = 0.844$, $\eta^2 = 0.001$.

As in Experiment 1, false alarms were rare across all conditions. A 2 (Variability) x 4 (Training Type) ANOVA with hits as the dependent variable revealed a main effect of variability, $F(1, 855) = 9.626$, $p = 0.002$, $\eta^2 = 0.011$, and a main effect of training type, $F(3, 855) = 18.464$, $p < 0.001$, $\eta^2 = 0.061$ (see Figure 4b). A comparable analysis of false alarms revealed no significant effects, $ps > .08$ (see Figure 4c). Thus, as in Experiment 1, the effect of training on adults' performance was driven by hits.

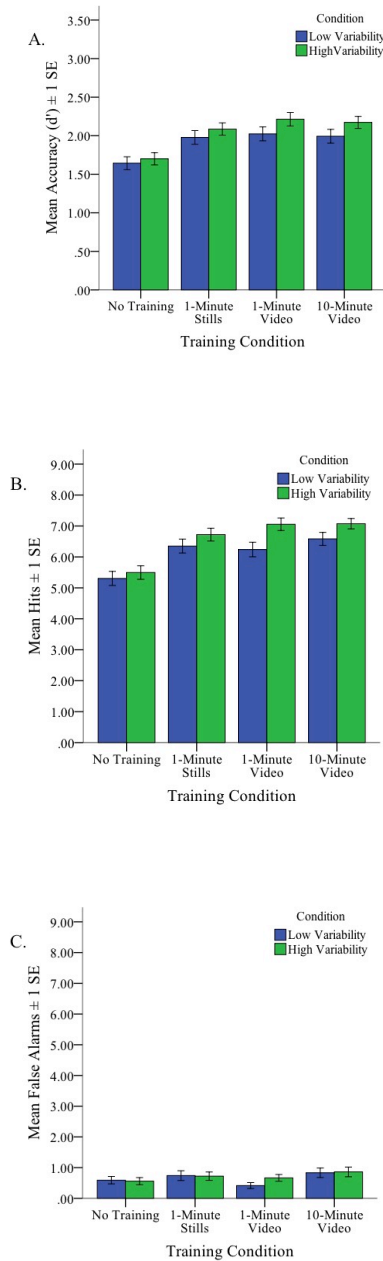


Figure 4. Mean accuracy of adult participants in the low- and high-variability training conditions of Experiment 2. Accuracy is shown separately for four groups within each condition: No-training control, 1-minute video stills, 1-minute video and 10-minute video. Three dependent variables are shown: d' (A), hits (B), and false alarms (C). Error bars show \pm SEM.

Discussion

In summary, the results of Experiment 2 show that adults learn a new identity with remarkable efficiency. Performance in all training conditions was more accurate than that of participants in the no-training condition. Unlike Experiment 1, there was some evidence that adults benefitted from high compared to low variability. Although the effect size was very small and the significant effect likely attributable to our large sample size, this finding is consistent with past evidence that exposure to high variability in appearance facilitates face learning (e.g., Menon et al., 2015a; Ritchie & Burton, 2016). Most notably, performance was comparable across training conditions despite variability in both duration (1 vs. 10 minutes) and type of exposure (Video vs. Stills). In short, among adult participants, viewing 12 images over a 1-minute period lead to the same level of familiarity as did viewing a 10-minute video. This stands in contrast to the children tested in Experiment 1 who showed evidence of learning from a 10-minute video only in the HV condition. We address the developmental implications of Experiment 2 in the General Discussion.

General Discussion

A fundamental difference between familiar vs. unfamiliar faces is the accuracy with which two or more instances of the same identity are recognized as belonging to the same person. Whereas multiple images of a familiar face are easily recognized as belonging together, those same images are perceived as belonging to two or more identities when the face is unfamiliar (e.g., Jenkins et al., 2011). This contrast is attributable to familiar faces having an abstract representation that allows recognition across a range of inputs, whereas recognition of unfamiliar faces relies more on lower level image properties (e.g.,

pictorial cues; see Burton et al., 2011). Here we provide the first examination of the process by which a face becomes familiar in children aged 6 to 13 years and we contrast children's performance to that of adults.

We report two key findings. First, after watching a 10-minute video filmed across three days (HV condition) both adults and children showed evidence of becoming familiar with a new facial identity; they were able to recognize more novel instances of the model than participants in the no-training control condition. The face did not become highly familiar to either age group; unlike adults (Jenkins et al., 2011) and 6- to 12-year-old children (Laurence & Mondloch, 2016) tested with a personally familiar face, our participants made errors after familiarization. Nonetheless, exposure to within-person variability in appearance resulted in a newly encountered face becoming partially familiar. Future research should examine the process by which a face becomes completely, rather than partially, familiar; we predict that exposure to more variability than was captured in our HV video (e.g., more extensive variability in appearance and environments) would further expand the number of recognizable images (see Burton et al., 2016 for a discussion).

Second, adult's performance showed minimal benefits of high compared to low variability in appearance (no effect in Experiment 1 and a small effect in Experiment 2). Their performance was comparable across exposure times (1 vs. 10 minutes) and type (video vs. stills), further highlighting adults' remarkable ability to build a new representation based on minimal variability. In contrast, children showed evidence of learning only when the video incorporated variability in the model's appearance across days (lighting, camera, make-up, hairstyle). Whether children would show evidence of

learning after viewing the 1-minute HV video or 12 slides incorporating high variability in appearance is a matter for future research, but exposure to lots of variability appears key to children's face learning.

The current study extends Laurence and Mondloch's (2016) examination of children's ability to recognize multiple images of familiar and unfamiliar faces. In that study children aged 5 to 12 years were tested with an unfamiliar identity and children aged 4 to 12 years were tested with a highly familiar face—that of their own teacher. Laurence and Mondloch reported age-related improvements when children were tested with an unfamiliar face; here we observed a similar pattern in the NT and HV conditions, but found no age-related improvement in the LV condition. Collectively, these results suggest that the ability to recognize multiple images of an unfamiliar or partially learned face improves with age. The lack of age-related improvement in the LV condition is a bit surprising and should be replicated.

Laurence and Mondloch also reported (nearly) perfect performance in children aged 6 years and older who were tested with their own teacher's face, evidence that by 6 years of age children are able to build an abstract representation of facial identity that allows them to tolerate variability in appearance and recognize new instances. Because children had known their teacher for several months the time course of their learning remained unknown. One finding in Laurence and Mondloch's study suggested that face learning in early childhood is not adult-like: several 4- and 5-year-old children made errors when tested with their teacher's face—despite knowing her for between 3 and 9 months. The current study confirmed that hypothesis; children did show evidence of learning after watching a 10-minute video, but only in the HV condition—a condition in which they

observed natural changes across days (i.e., changes our models normally display across settings such as work, home, and evenings out). Despite experiencing a great deal of variability in appearance (e.g., head orientation, expression), children in the LV condition did not experience day-to-day variability in hairstyle, make-up and lighting.

Burton et al. (2016) argue that the type of variability encountered for any identity constrains the range of inputs that will be attributed to a particular person; celebrities' families will experience and tolerate more variability in appearance than those who know the celebrity only from television and the movies. Our data suggest that children require more variability than adults to recognize novel images of a learned facial identity, offering a possible explanation for 4- and 5-year-olds' poor performance when tested with their own teacher's face (Laurence & Mondloch, 2016). Many teachers and daycare workers keep their appearance fairly constant across workdays and it might be that 4- and 5-year-olds' failure to recognize their teacher's face reflects constraints on the type of variability they experienced. Very young children might be less able than older children and adults to extrapolate beyond experienced within-person variability. Future studies should examine whether very young children tolerate more variability for faces they encounter in a wider range of settings (e.g., relatives, neighbours).

What develops? Underlying mechanisms. By 6 years of age children can form a robust representation of a highly familiar face (Laurence & Mondloch, 2016), but they are less tolerant than adults of within-person variability in unfamiliar faces (current study; Laurence & Mondloch) and require exposure to more variability than adults to show evidence of face learning, as evidenced from children showing evidence of learning only in the HV condition, in contrast to adults who showed learning even in the LV video stills

condition (current study). Two processes likely contribute to a face becoming familiar (i.e., to an increased tolerance of within-person variability): the formation of an average that includes constants in a person's appearance but excludes pictorial cues (Burton et al., 2005; Kramer et al., 2015) and a representation of how an individual face varies (Burton et al., 2016). Several potential mechanisms are likely candidates for the protracted development of one or both of these processes.

First, children might be less sensitive than adults to the dimensions (e.g., principal components) of idiosyncratic within-person variability such as an individual's smile or facial hair/makeup (see Burton et al., 2016). This hypothesis is consistent with evidence that children are less sensitive than adults to dimensions along which different facial identities vary (e.g., Anzures, Mondloch, & Lackner, 2009; Crookes & McKone, 2009; Jeffery et al., 2010) and that they utilize fewer dimension simultaneously than do adults when discriminating between identities (Nishimura, Maurer, & Gao, 2009).

Second, children might be less efficient than adults in extracting an average representation of a set of images (or from a video). Kramer, Ritchie and Burton (2015) provided strong evidence of ensemble encoding of facial identity: After briefly viewing four images of an identity, adults were as likely to report that the average of those images had been in the set as they were to report having seen one of the original images. Furthermore, this ensemble encoding was unaffected by whether the four images were presented simultaneously or sequentially. Kramer et al. suggested that ensemble encoding is a means by which rapid learning occurs. Little is known about ensemble encoding in children but one study (Rhodes et al., in press) showed that ensemble encoding of four different identities continues to improve until 18 years of age even when four images are

presented simultaneously. What remains unknown is children's ability to extract an average from images of the same person, whether that ability is impaired when images are presented sequentially, and whether their memory for the original images (memory that would constrain a representation of variability) is less reliable.

Third, both extracting an average and storing several instances of an identity require that an instance be successfully encoded in visual working memory (VWM). Thus developmental changes in face learning might be attributable to limitations in the capacity and/or precision of VWM. It is known that experience affects the precision of adults' representations, with impairments for inverted relative to upright faces (Lorenc, Pratte, Agneloni, & Tong, 2014) and for other-race relative to own-race faces (Zhou, Mondloch, & Emrich, submitted). Future studies should directly compare the contribution of age-related changes in VWM for faces to improved face learning.

Face specific or domain general? A hotly contested debate among researchers investigating the development of face recognition during childhood is the extent to which age-related changes are domain specific vs. domain general. Some researchers (e.g., Crookes & Robbins, 2014; McKone, Crookes, Jeffrey, & Dilks, 2012; Weigelt et al., 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 for prolonged development of face-specific mechanisms. Performance on any task is influenced by domain-general cognitive development (e.g., age-related changes in attention); thus we included control trials to ensure that children understood our task and remained attentive throughout. Given children's accuracy on control trials, we contend

that the age-related improvement we observed is not solely attributable to general cognitive development.

What is novel about our method and differentiated adults and children is that we measured participants' ability to utilize within-person variability in appearance to form a representation of identity that allows them to recognize novel instances. This is an important aspect of face recognition (Burton, 2013) that has been largely ignored in both the adult and developmental literatures. The prolonged development of this ability, as reflected in the lack of learning in the LV condition of the current study and the poor performance of 4- and 5-year-olds when tested with their teacher's face (Laurence & Mondloch, 2016), might well reflect the development of domain-general mechanisms (e.g., VWM, ensemble encoding; see above). Nonetheless adults' remarkable ability to recognize familiar faces despite natural variability in appearance and to form representations of newly encountered faces likely is face-specific. There is no comparable object in the visual domain requiring sensitivity to both within-class discriminations and within-exemplar variability. Even the faces of our closest phylogenetic relative, Chimpanzees, do not vary as much as human faces. (Chimpanzees do not use make up or vary their hairstyle; they have fewer Action Units and less perceptible eye gaze; see Vick, Waller, Parr, Pasqualini & Bard, 2007 for a discussion.) At least in the visual domain, recognizing identity is likely face specific.

Although the proficiency with which adults recognize facial identity is unique in the visual domain, variability seems to foster learning in several domains. It facilitates perceptual expertise in adults (e.g. texture recognition; Hussain et al., 2012; detecting dangerous items in x-rays, Gonzalez & Madhavan, 2011) and language development in

children (see Watson et al., 2014 for a discussion). Furthermore the ability to form representations robust to within-exemplar variability is evident in the auditory domain. Adults are able to recognize individual words despite variability in environment (e.g., acoustics, sentence context) and speaker (e.g., age, sex, accent); likewise they recognize melodies despite changes in the environment (car radio vs. concert hall), source (e.g., instrument, musical key), and tempo. Comparisons of how adults build robust representations for faces, language and music and the development of these skills across domains would be a rich avenue for future studies. These categories are found universally and share common developmental mechanisms (e.g. show perceptual narrowing during the first year of life; Hannon & Trehub, 2005; Kelly et al., 2007; Werker & Tees, 1984).

Summary. An important challenge in daily life is to recognize familiar faces—an ability that requires discriminating between identities and recognition despite variability in appearance. Adults show partial familiarization after viewing 12 photos taken on a single day, with no increased benefit from either motion or a longer exposure time. This ability for faces is unique in the visual domain, but might be comparable to the ability observed in language and music perception. The development of the ability—at least in the domain of face perception—is prolonged; children make more errors than adults when a face is unfamiliar and require more within-person variability to form a robust representation.

Acknowledgements

This work was supported by an NSERC Discovery Grant (327520) and an NSERC Discovery Accelerator Supplement Award (412323) given to C.J.M. We thank Michael Busseri for statistical advice, our models for reading the story, as well as the children, parents, and their teachers for participating.

References

- Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *Quarterly Journal of Experimental Psychology*, *68*(10), 2041-2050. doi: 10.1080/17470218.2014.1003949.
- Anzures, G., Mondloch, C. J., & Lackner, C. (2009). Face adaptation and attractiveness aftereffects in 8-year-olds and adults. *Child Development*, *80*(1), 178-191. doi:10.1111/j.1467-8624.2008.01253.x.
- Baudouin, J. Y., Gallay, M., Durand, K., & Robichon, F. (2010). The development of perceptual sensitivity to second-order facial relations in children. *Journal of Experimental Child Psychology*, *107*(3), 195-206. doi:10.1016/j.jecp.2010.05.008
- Bindemann, M., & Sandford, A. (2011). Me, myself, and I: different recognition rates for three photo-IDs of the same person. *Perception*, *40*(5), 625–7. <http://doi.org/10.1068/P7008>
- 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 Section A*, *57*(6), 1019-1029. doi: 10.1080/02724980343000657
- Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Langton, S., McAuley, S., & Wright, R. (2000). Testing face processing skills in children. *British Journal of Developmental Psychology*, *18*(3), 319-333. doi:10.1348/026151000165715
- 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, 339–360. doi:10.1037/1076-898X.5.4.339
- Burton, M. A. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, 66(8), 1467–1485. <http://doi.org/10.1080/17470218.2013.800125>
- Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, 51(3), 256–284. doi:10.1016/j.cogpsych.2005.06.003
- Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, 102, 943–958. doi: 10.1111/j.2044-8295.2011.02039.x
- Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: representations of faces derived from multiple instances. *Cognitive Science*, 40, 202–223. doi: 10.1111/cogs.12231
- Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition*, 111(2), 219–247. <http://dx.doi.org/10.1016/j.cognition.2009.02.004>
- Crookes, K., & Robbins, R. A. (2014). No childhood development of viewpoint invariant face recognition: Evidence from 8-year-olds and adults. *Journal of Experimental Child Psychology*, 126, 103–111. doi:10.1016/j.jecp.2014.03.010
- de Heering, A., Rossion, B., & Maurer, D. (2012). Developmental changes in face recognition during childhood: Evidence from upright and inverted faces. *Cognitive Development*, 27(1), 17–27. doi:10.1016/j.cogdev.2011.07.001

Dowsett, A. J., Sandford, A., & Burton, A. M. (2016). Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *The Quarterly Journal of Experimental Psychology*, *69*(1), 1–10.

<http://doi.org/10.1080/17470218.2015.1017513>

Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., ... Lee, K. (2009).

Two faces of the other-race effect : Recognition and categorisation of Caucasian and Chinese faces. *Perception*, *38*, 1199–1211. <http://doi.org/10.1068/p6136>

Gilchrist, A., & McKone, E. (2003). Early maturity of face processing in children: Local and relational distinctiveness effects in 7-year-olds. *Visual Cognition*, *10*(7), 769–793. doi:10.1080/13506280344000022

Gonzalez, C., & Madhavan, P. (2011). Diversity during training enhances detection of novel stimuli. *Journal of Cognitive Psychology*, *23*(3), 342–350.

<http://dx.doi.org/10.1080/20445911.2011.507187>

Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces.

Trends in Cognitive Sciences, *4*, 330–337. [http://dx.doi.org/10.1016/S1364-6613\(00\)01519-9](http://dx.doi.org/10.1016/S1364-6613(00)01519-9)

Hannon, E. E., & Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences USA*, *102*, 12639–12643. doi: 10.1073/pnas.0504254102

Hussain, Z., Bennett, P. J., & Sekuler, A. B. (2012). Versatile perceptual learning of textures after variable exposures. *Vision Research*, *61*, 89–94.

<http://dx.doi.org/10.1016/j.visres.2012.01.005>

- Jeffery, L., McKone, E., Haynes, R., Firth, E., Pellicano, E., & Rhodes, G. (2010). Four to-six-year-old children use norm-based coding in face-space. *Journal of Vision, 10*(5), 1-19. doi:10.1167/10.5.18
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313–23.
<http://doi.org/10.1016/j.cognition.2011.08.001>
- Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory, 17*, 577–596. <http://dx.doi.org/10.1080/09658210902976969>
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science, 18*(12), 1084–1089. doi: 10.1111/j.1467-9280.2007.02029.x
- Kramer, R. S. S., Ritchie, K. L., & Burton, A. M. (2015). Viewers extract the mean from images of the same person. A route to face learning. *Journal of Vision, 15*(4):1, 1-9. doi: 10.1167/15.4.1
- Laurence, S., & Mondloch, C. J. (2016). That’s my teacher! Children’s ability to recognize personally familiar and unfamiliar faces improves with age. *Journal of Experimental Child Psychology, 143*, 123-138. doi.org/10.1016/j.jecp.2015.09.030
- Lorenc, E.S., Pratte, M.S., Angeloni, C.F., & Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Attention, Perception, & Psychophysics, 76*, 1975-1984. doi:10.3758/s13414-014-0653-z
- Macchi Cassia, V., Luo, L., Pisacane, A., Li, H., & Lee, K. (2014). How race and age experiences shape young children’s face processing ability. *Journal of Experimental Child Psychology, 120*, 87–101. <http://dx.doi.org/10/1016/j.jecp/2013.11.016>

- McKone, E., & Boyer, B. L. (2006). Sensitivity of 4-year-olds to featural and second order relational changes in face distinctiveness. *Journal of Experimental Child Psychology, 94*, 134–162. <http://dx.doi.org/10.1016/j.jecp.2006.01.001>
- McKone, E., Crookes, K., Jeffery, L., & Dilks, D. D. (2012). A critical review of the development of face recognition: Experience is less important than previously believed. *Cognitive Neuropsychology, 29*(1-2), 174-212.
doi:10.1080/02643294.2012.660138
- Megreya, A. M., & Bindemann, M. (2015). Developmental improvement and age related decline in unfamiliar face matching. *Perception, 44*, 5-22. doi:10.1068/p7825
- Menon, N., White, D., & Kemp, R. I. (2015a). Identity-level representations affect unfamiliar face matching performance in sequential but not simultaneous tasks. *The Quarterly Journal of Experimental Psychology, 68*(9), 1777–1793.
<http://doi.org/10.1080/17470218.2014.990468>
- Menon, N., White, D., & Kemp, R. I. (2015b). Variation in photos of the same face drives improvements in identity verification. *Perception, 44*(11), 1332–1341.
<http://doi.org/10.1177/0301006615599902>
- Mondloch, C. J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology, 86*(1), 67-84.
doi:10.1016/S0022-0965(03)00102-4
- Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31*(5), 553-566.
doi:10.1068/p3339

Mondloch, C. J., & Thomson, K. (2008). Limitations in 4-year-old children's sensitivity to the spacing among facial features. *Child Development, 79*(5), 1513-1523.

doi:10.1111/j.1467-8624.2008.01202.x

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

Newcombe, N., & Lie, E. (1995). Overt and covert recognition of faces in children and adults. *Psychological Science, 6*(4), 241-245.

doi:10.1111/j.14679280.1995.tb00599.x

Nishimura, M., Maurer, D., & Gao, X. (2009). Exploring children's face-space: A multidimensional scaling analysis of the mental representation of facial identity. *Journal of Experimental Child Psychology, 103*(3), 355-375.

doi:10.1016/j.jecp.2009.02.005

Pellicano, E., & Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychological Science, 14*(6), 618-622.

doi:10.1046/j.09567976.2003.psci_1474.x

Pellicano, E., Rhodes, G., & Peters, M. (2006). Are preschoolers sensitive to configural information in faces? *Developmental Science, 9*(3), 270-277. doi:10.1111/j.1467-7687.2006.00489.x

Rhodes, G., Neumann, M., Ewing, L., Bank, S., Read, A., Engfors, L. M., Emiechel, R., & Palermo, R. (in press). Ensemble coding of faces occurs in children and develops dissociably from coding of individual faces. *Developmental Science*.

Ritchie, K. L., & Burton, A. M. (2016). Learning faces from variability. *The Quarterly*

Journal of Experimental Psychology, 1–9.

<http://doi.org/10.1080/17470218.2015.1136656>

Robinson, J. G. (1975). *About Teddy Robinson*. London: Puffin

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental science*, *12*(2), 339-349.

Short, L. A., Lee, K., Fu, G., & Mondloch, C. J. (2014). Category-specific face prototypes are emerging, but not yet mature, in 5-year-old children. *Journal of Experimental Child Psychology*, *126*, 161-177. doi:10.1016/j.jecp.2014.04.004

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*(2), 833-870.

Tanaka, J. W., Kay, J. B., Grinnell, E., Stansfield, B., & Szechter, L. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition*, *5*(4), 479-496. doi:10.1080/713756795

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) on the development of face perception. *Journal of Experimental Child Psychology*, *124*, 36-49. doi: 10.1016/j.jecp.2014.01.001

Vick, S. J., Waller, B. M., Parr, L. A., Pasqualini, M. C. S., & Bard, K. A. (2007). A cross-species comparison of facial morphology and movement in humans and chimpanzees using the facial action coding system (FACS). *Journal of Nonverbal Behavior*, *31*(1), 1-20. doi: 10.1007/s10919-006-0017-z

- Watson, T., Robbins, R., & Best, C. (2014). Infant perceptual development for faces and spoken words: An integrated approach. *Developmental Psychobiology*, *56*(7), 1454-1481. doi:10.1002/dev.21243
- Weigelt, S., Koldewyn, K., Dilks, D. D., Balas, B., McKone, E., & Kanwisher, N. (2014). Domain-specific development of face memory but not face perception. *Developmental Science*, *17*(1), 47-58. doi: 10.1111/desc.12089.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7* (1), 49–63. doi:10.1016/S0163-6383(84)80022-3
- Wilson, R. R., Blades, M., Coleman, M., & Pascalis, O. (2009). Unfamiliar face recognition in children with autistic spectrum disorders. *Infant and Child Development*, *18*(6), 545-555. doi:10.1002/icd.638
- Zhou, X. & Mondloch, C.J. (2016). Recognizing ‘Bella Swan’ and ‘Hermione Granger’: No own-race advantage in recognizing photos of famous faces. *Perception*.
Published online before print. 10.1177/0301006616662046
- Zhou, X., Mondloch, C.J., & Emrich, S.M. (submitted). Perceptual experience affects the rapid encoding of own- and other-race face representations in visual working memory.