

Running head: DEVELOPMENT OF FACE MEMORY

*In press in Developmental Science*

**Domain-specific development of face memory but not face perception**

Sarah Weigelt, Kami Koldewyn, Daniel D. Dilks, Benjamin Balas, Elinor McKone and

Nancy Kanwisher

Massachusetts Institute of Technology

Author Note

Sarah Weigelt, Kami Koldewyn, Daniel D. Dilks, Benjamin Balas and Nancy Kanwisher, Department of Brain and Cognitive Science, Massachusetts Institute of Technology

Benjamin Balas, Department of Psychology, North Dakota State University

Elinor McKone, Department of Psychology, Australian National University

This research was supported by Ellison Medical Foundation (NK), Simons Foundation (KK, SW) and Australian Research Council DP0984558 (EM). Authors thank Joshua B. Julian and Alex Kell for help with testing.

Correspondence concerning this article should be addressed to Sarah Weigelt, Ruhr-Universität Bochum, Fakultät für Psychologie, Universitätsstr. 150, 44801 Bochum Germany. E-mail: sarah.weigelt@gmail.com

**Abstract**

How does the remarkable human ability for face recognition arise over development? Competing theories have proposed either late maturity (beyond 10 years) or early maturity (before 5 years), but have not distinguished between perceptual and memory aspects of face recognition. Here, we demonstrate a perception-memory dissociation. We compare rate of development for (adult, human) faces versus other social stimuli (bodies), other discrete objects (cars), and other categories processed in discrete brain regions (scenes, bodies), from 5 years to adulthood. For perceptual discrimination, performance improved with age at the same rate for faces and all other categories, indicating no domain-specific development. In contrast, face memory increased more strongly than non-face memory, indicating domain-specific development. The results imply that each theory is partly true: the late maturity theory holds for face memory, and the early maturity theory for face perception.

(138 words)

### **Domain-specific development of face memory but not face perception**

Human observers can recognize a face from a brief glimpse, with great speed and accuracy. How does this remarkable ability develop? Two competing theories have been extensively debated. The *late maturity view* of face recognition initially proposed that children process faces in a qualitatively different way from adults, and the basic abilities of face perception (e.g., holistic processing, ability to encode a novel face) were not even present until 10 years of age (Carey & Diamond, 1977; 1994). Following evidence that such basic abilities are present much earlier than 10 years of age (for a review see McKone, Crookes, Jeffery & Dilks, 2012), this theory has been reformulated, now arguing only for late *quantitative* maturity in which skills are present, but do not reach full adult levels until late childhood or adolescence (e.g., Mondloch, LeGrand & Maurer, 2002). The alternative *early maturity* view holds that all face recognition skills are adultlike by age five both qualitatively (i.e., they are present) and quantitatively (i.e., they are at adults levels). This theory attributes the ongoing improvement in laboratory face task performance with age to improvements in other task-relevant cognitive abilities such as ability to concentrate and follow instruction (Crookes & McKone, 2009; McKone et al., 2012).

Here, we show that each theory captures only a partial truth. Specifically, both fail to distinguish conceptually between the development of face *memory*, and face *perception*. By face perception, we mean the ability to *discriminate* individual faces with little or no memory requirement, plus the perceptual coding skills that presumably drive this ability (e.g., face-space coding, and holistic processing, both of which contribute to face recognition ability in adults; Dennett, McKone, Edwards & Susilo, 2012; Wang, Li,

Fang, Tian & Liu, 2012). By face memory, we mean the ability to retain and individuate faces in long-term memory. This ability relies substantially on face perception, but additionally requires: retention; comparison of a currently-perceived test face to faces that are only stored not perceived; and, in the standard "explicit" memory tasks (e.g., old-new recognition), conscious awareness of whether the face was previously learned.

Evidence from outside the developmental literature suggests that face perception and face memory may engage partly dissociable mechanisms. The classic literature on acquired prosopagnosia distinguishes between "apperceptive" and "associative" prosopagnosia, pointing to a dissociation between face discrimination versus face memory (De Renzi, Faglioni, Grossi & Nichelli, 1991; see also McKone et al., 2011; Tippett, Miller & Farah, 2000). Similarly, in autism, deficits in face memory tend to be more severe than deficits in face perception (Weigelt, Koldewyn & Kanwisher, 2012). These dissociations raise the question of whether face memory and face perception develop differently.

Here we ask whether *domain-specific* processing for faces develops at different rates for memory and perception. By domain-specific development, we mean a finding in which the slope of development with age is steeper for faces than for other comparison categories. Although developmental improvements in face recognition performance throughout childhood and adolescence are well established on laboratory tasks testing both memory (Carey & Diamond, 1977; Crookes & McKone, 2009; Golarai et al., 2007; Lawrence et al., 2008; O'Hearn et al., 2010) and perceptual discrimination (e.g., Johnston, Kaufman, Bajic, Sercombe & Karayanidis, 2011; Mondloch et al., 2002), the developmental trends for *faces* on these tasks cannot be used, in isolation, to determine

the age at which face-specific processing is mature. The problem is that performance improvement with age does not reflect purely the development of face ability per se; some or all of the improvement in task performance with age will reflect the development of domain-general functions like attention and executive processing, which continue to improve during this period (e.g. Best & Miller, 2010). To isolate development of face-specific processing mechanisms, the developmental trajectory of performance on face tasks must be compared with that for other stimulus classes (Crookes & McKone, 2009; Want, Pascalis, Coleman & Blades, 2003).

Here, we test the prediction that, if the face recognition system undergoes domain-specific development then performance should improve with age more for faces than for other stimulus categories. Only a very few studies have addressed this prediction, and results are contradictory<sup>1</sup>. One study finds no developmental change for either faces or houses (Aylward et al., 2005). Two studies find greater development for faces than houses/scenes arguing for domain-specific development of face recognition (Carey & Diamond, 1977; Golarai et al., 2007). And two studies (Crookes & McKone, 2009; Johnston et al., 2011) find equal development of accuracy for faces and non-faces (dogs/butterflies), thus arguing for domain-general improvement and not face-specific development.

These studies vary in whether they test memory (Aylward et al., 2005; Carey & Diamond, 1977; Crookes & McKone, 2009; Golarai et al., 2007), or perceptual discrimination (Johnston et al., 2011). However, attempts to determine whether they support a memory-perception distinction in development is complicated by two other factors. First, different studies tested different age ranges. Here, therefore, we test a

broad range from 5 years to adulthood, and test each participant on both memory and perceptual discrimination.

Second, different non-face comparison stimuli are used in each study. Moreover, these vary in the ways they relate theoretically to faces. Thus, here, we tested four different categories – faces, and three comparison classes of cars, (headless) bodies, and scenes. Key properties of faces are that they are discrete objects, social stimuli, and activate specialized brain areas in adults. Thus, we compared faces to: other discrete objects (cars); other social stimuli (bodies); and other stimuli that activate specialized brain regions in adults (here, we used scene-like layouts of internal views of houses that been previously established to selectively activate the parahippocampal place area). This enabled us to test several hypotheses for both perception and memory: (a) domain-specific development for faces (predicting greater developmental improvement for faces than for all other stimulus categories); (b) no domain-specific development (predicting equal developmental improvement across faces and all other stimulus categories); and (c) domain-specific development that is broader than just faces, either extending to all social stimuli (predicting greater developmental improvements for faces and bodies than cars or scenes), for discrete objects (greater for faces, bodies, and cars than scenes), or all categories that are processed in specialized brain regions in adults (greater for faces, bodies, and scenes than cars).

## Methods

### Participants

We tested 118 participants comprising 17 5-year olds ( $M = 5.5$  y,  $SD = 0.3$  y, 7 male), 13 6-y olds ( $M = 6.4$  y,  $SD = 0.3$  y, 12 male), 15 7-year olds ( $M = 7.4$  y,  $SD = 0.3$  y, 13 male), 11 8-year olds ( $M = 8.7$  y,  $SD = 0.2$  y, 10 male), 15 9-year olds ( $M = 9.6$  y,  $SD = 0.3$  y, 8 male), 15 10-year olds ( $M = 10.7$  y,  $SD = 0.3$  y, 9 male) and 32 adults ( $M = 33.7$  y,  $SD = 11.1$  y, 18 male). All were recruited from the Boston metropolitan area. Parental consent was obtained.

### Design

Each participant was tested on four stimulus categories (faces of human adults, cars, headless bodies of human adults, and scenes) in each of two tasks: memory and perceptual discrimination. We used the same stimuli in both tasks, so that any dissociation in developmental trajectories between tasks cannot be attributed to differences between stimulus sets. Since we did not want exposure to the stimuli in the perception task to taint memory performance for those stimuli, memory was always tested before perception.

This study was part of a larger study investigating individual differences in multiple cognitive tasks. To reduce between-subject variability we kept the presentation order of the categories for one set of participants the same to faces, cars, bodies, and scenes. Thus, participants performed face, car, body, and scene memory, and then face, car, body, and scene perception. For some participants we reversed the presentation order to scenes, bodies, cars, and faces. These participants thus did scene, body, car, and face

memory, and then scene, body, car, and face perception. Presentation order did not have an effect on our data (see supporting information available online). In the following we describe experimental procedures for the forward presentation order.

### **Procedure**

Stimuli were presented on a MacBook Pro or Elo touchscreen computer using Matlab and the Psychtoolbox extension (version 3.0.9, Brainard, 1997; Pelli, 2007). Responses were recorded via either buttons or touchscreen. Some children chose to give their answers by pointing towards the screen, and the experimenter pressed the buttons accordingly. Because of these different response devices, we are not considering reaction time as a dependent measure. Participants were tested individually.

### **The memory task**

**Segment 1: faces and cars.** In the first segment, participants studied ten face items followed by ten car items. During the study phase, participants were told to watch all items carefully and to remember them as best as they could. A fixation cross appeared until the experimenter judged the participant was concentrating, followed by the twenty study items sequentially, each presented at screen center for 3 s. Study was immediately followed by a 2AFC test phase (Figure 1), with 10 face pairs followed by 10 car pairs. Items in each test pair were shown simultaneously side-by-side until the participant responded. To approximately equate memory delay across items (about 1 minute between study and test), items in the test phase appeared in the same order as in the study phase. Participants were asked to indicate the “old” item (on the right 50% of the time), guessing



if necessary. There was no feedback. The dependent measure was accuracy. Chance is 50%.

**Segment 2: Bodies and scenes.** The procedure of the first segment was repeated for bodies and scenes (bodies first at both study and test).

### **The perceptual discrimination task**

This task measured discrimination threshold (stimulus difference necessary to perform 75% correct) in a 2AFC match-to-sample task (Figure 2A). Again, each participant was tested on all four stimulus categories, in the order faces, cars, bodies scenes. On each trial, a fixation cross appeared until the experimenter judged the participant was concentrating, followed by the sample item at screen center for 1 s, then a test pair simultaneously side-by-side until the participant responded. The test pair comprised the sample item and a distracter created by morphing the target item towards a different-identity exemplar of the same category (Figure 2B). Participants were told to report which item they had just seen. Note that because each test pair was presented immediately after the sample item, our perceptual discrimination task requires holding face information in memory for a few hundred milliseconds. We chose this task over a simultaneous same-different task (which still requires a shift in overt or covert attention, hence holding information in mind for at least one to two hundred milliseconds) because many five year olds are confused about the criteria for what counts as the same, and might simply say that all of our (very similar) face pairs are the same; this ambiguity does not exist with the task we used.

**Experimental trials.** The dependent measure was morph distance at which participants could discriminate between the study item and the morphed item with an accuracy of 75% correct. This was estimated using a QUEST staircase (Watson & Pelli, 1983) with parameters: number of trials = 30; Beta (slope of the estimated psychometric function) = 3.5; Delta (estimated probability of a failure well above threshold) = 0.01; Gamma (estimated probability of a correct response at zero intensity) = 0.5; Grain (intensity steps, i.e. minimum morph difference between two images) = 5. We checked the reliability of individual participants' threshold values via the standard deviation associated with their QUEST threshold; this reliability was as good in the youngest children as in adults (see supporting information available online).

Participants were told the task would get harder as they went along, until they might not be able to tell the difference between the two items, at which point they should guess. Children were told not to feel bad about not knowing which is the correct one. The correct item was on the right 50% of the time. There was no feedback.

**Practice trials.** Practice trials at the beginning of each category used the same stimuli as the main experiment and very easy test pairs (target item and an 80%-morph distracter). Child participants received feedback and encouragement. When children answered four practice items in a row correctly, the program advanced to the experimental trials. If a child was unable to complete any four consecutive trials of 12 trials total, a new practice session started; if unable to complete any of three practice sessions, this was counted as the child not being able to perform the discrimination task. This only occurred in four participants (one 5 y old for faces, two 5 y olds for bodies, one 8 y old for scenes). The respective data points were treated as missing.

**Lapse trials.** Six lapse trials per category were interspersed at regular intervals among the experimental trials. These contained very easy test pairs (study item and a 100%-morph distracter) and were included to keep children's attention on the task as high as feasible. Errors on lapse trials decreased with age ( $F(6, 104) = 7.375, p < .001$ ), as would be expected given younger children are more likely to suffer lapses in attention. However, even the five-year-olds made only 20% errors on lapse trials, indicating they understood instructions, and their errors were evenly spread across categories (faces = 20%, cars = 16%, bodies = 22%, scenes = 22%).

### **Stimuli**

For all stimulus categories, stimuli were grayscale, static, not displaying emotion (relevant to faces and bodies), of similar age (again relevant to faces and bodies), displayed in 'canonical' viewpoint, and always presented in the same images at study and test in the memory task. This matching of basic format across categories ensured that any difference in developmental trend for face versus nonfaces could not be attributed to differences in these factors (color, movement, emotion, viewpoint, etc.). The 20 stimuli of each category used for the memory task all differed in individual identity. The specific item set for each category was selected so as to match task difficulty across the categories (using 10 y olds as the matching age group; see Results for rationale). Importantly, all items were within the normal range of natural items from the category (i.e., none were bizarre or highly atypical); under these circumstances, we would not expect the matching to induce unusual processing strategies (i.e. participants of all ages should be using the processing strategy that is natural for each category in their age group).

Faces were natural, real world faces of Caucasian men from the Harvard Face Database, with neutral expression, in front view, sized  $11.137^\circ$  vertical x  $11.137^\circ$  horizontal including the black background, with no facial hair or glasses, and black hats to hide hair and ears; hair was hidden to ensure that this could not operate as a nonface cue to memory/discrimination. Cars were photographs of typical cars, in side view, sized  $7.153^\circ$  vertical x  $16.63^\circ$  horizontal, on a white background. Car stimuli were freely available images from the Internet and were chosen to allow morphing between relatively similar pairs. The headless bodies were images of human adults (50% female) generated with Poser 6 software (Smith Micro Software Inc., Watsonville, CA, USA), in front view, in constant pose (i.e., arms out), not expressing emotions, on white background. The size of the bodies varied between  $13.544^\circ$  and  $15.658^\circ$  vertically,  $4.295^\circ$  and  $7.153^\circ$  horizontally with respect to the torso width and  $17.061^\circ$  and  $19.852^\circ$  with respect to the arm width. Scenes were perspective views of houses missing one wall, sized  $9.25^\circ$  vertical x  $14.955^\circ$  horizontal including a shaded background, and generated with Google SketchUp.

The 20 individual stimuli of each category were paired to make 10 morph continua, by morphing one endpoint exemplar into its paired exemplar (e.g. one face into its paired face, see Figure 2B) in steps of 5%. Morphing was realized within FantaMorph Software (Abrosoft) for faces and cars, Poser 6 for bodies (only between stimuli of the same gender with same clothing), and Google SketchUp for scenes.

## Results

### **Preliminary analysis: matching performance in the 10-y olds**

In our type of design, where a very large age range is tested and all age groups are tested on identical tasks, it is very difficult to avoid performance approaching either ceiling or floor in one or more stimulus categories. Under these circumstances, valid comparison of the slope of development between different stimulus categories requires that the stimuli be matched for difficulty by choosing items such that, at some age group, performance is matched across all stimulus categories (face, car, body, scene). If this is not done, then a false category x age interaction can occur (see supporting information [Figure S1] available on-line for illustration of how this can come about; for example, in the presence of a ceiling effect for one category and not another in older age groups, results could simply reflect the category with the easiest stimulus set reaching adult levels earlier in development than categories with more difficult stimulus sets).

As the age group to match on, we selected 10 year olds. Our rationale for this choice (as illustrated in Figure S1B-D), was that in the presence of a broad age range, matching on a middle age group minimizes the chance of a false *absence* of a category x age interaction and, at the same time, cannot produce a false *presence* of a category x age interaction even if performance unavoidably hits floor or ceiling at one extreme of the age range in one stimulus category (e.g. faces, if development is strongest for faces).

Results showed that, as desired, performance for 10-year-olds was roughly in the middle of the scales (see Figures 3 and 4), thus allowing room to observe both lower performance in the younger participants and higher performance in adults. Performance in 10-year olds was also matched across all stimulus categories (face, car, body, scene),

with no main effect of category (memory:  $F(3,42) = 1.064, p = .375, \eta_p^2 = .071$ ;  
 perception:  $F(3, 42) = 2.285, p = .093, \eta_p^2 = .140$ ); confirming that this matching was  
 genuine (not merely due to lack of room to show differences), it took place with  
 performance levels significantly different performance from both floor and ceiling (all  $ps$   
 $< .005$ ).

### **Memory: Face-specific development**

Results for memory (Figure 3) indicated domain-specific development of memory  
 for faces, and did not support any of our other hypotheses. ANOVA including all age  
 groups (5y, 6y, 7y, 8y, 9y, 10y, adults) and within-group factor category (face, car, body,  
 scene) revealed: no main effect of category,  $F(3, 333) = 1.712, p = .164, \eta_p^2 = .015$ ,  
 indicating good matching of overall stimulus difficulty; a main effect of age,  $F(5, 80) =$   
 $3.047, p = .044, \eta_p^2 = .130$ , indicating, as expected, overall improvement in memory  
 performance with age; and, crucially, a significant age x category interaction,  $F(18, 333)$   
 $= 2.187, p = .004, \eta_p^2 = .106$ , reflecting steeper developmental slopes for faces than the  
 other categories (see Figure 3). Follow-up ANOVAs contrasted rate of development for  
 faces with that for each other category in turn. These demonstrated that memory  
 increased significantly faster with age for faces than for bodies (category x age  
 interaction,  $F(6, 111) = 2.892, p = .012, \eta_p^2 = .135$ ), cars ( $F(6, 111) = 3.305, p = .005, \eta_p^2$   
 $= .152$ ), and scenes ( $F(6, 111) = 3.099, p = .008, \eta_p^2 = .143$ ). Lack of any broader  
 domain-specificity was indicated by equal rates of memory development with age for the  
 three non-face categories (body, car, scene), which showed no interaction with age,  $F(12,$   
 $222) = 1.202, p = .283, \eta_p^2 = .061$ .

A second observation is that the face-specificity in memory development occurred primarily before 10 years of age. After 10 years of age, there is no suggestion in Figure 3 that face memory developed faster than other memory, and correspondingly a category (face, body, car, scene) x age group (10 y, adults) ANOVA revealed no evidence of any interaction,  $F(3, 135) = 1.102, p = .351, \eta_p^2 = .024$ . In contrast, before 10 years of age, linear trend analyses (more powerful than ANOVA because the order and spacing of age groups is coded) targeting the children only (5y, 6y, 7y, 8y, 9y, 10y) found that face memory improved significantly more with age than did car memory,  $t(170) = 2.576, p = .012$ , or scene memory,  $t(170) = 1.988, p = .048$ , although this contrast did not quite reach significance for body memory,  $t(170) = 1.832, p = .069$ . Further, the three nonface categories did not differ (trend analysis on body vs. car:  $t(170) = .791, p = .430$ ; scene vs. car:  $t(170) = .661, p = .509$ ; body vs. scene:  $t(170) = .141, p = .888$ ). To give a numerical value for the rate of development with age, across the 5y to 10y age range we calculated regression slopes: regressing memory accuracy on individuals' age gave a slope for faces that was steeper than the slopes of the other stimuli ( $B$  for faces = 0.049, bodies = 0.022 cars = 0.011, scenes = 0.020).

Given that the developmental trend was strongest for faces, and we needed to set task difficulty suitable to a very large age range (5 y olds to adults), performance in the youngest group (5 y olds) was close to chance for faces (54.71%, see Figure 3). Crucially, this result cannot explain our key category x age interaction. Indeed, if anything, it would have caused us to *underestimate* the strength of the age x category interaction: if a floor effect was present such that the 'true' 5 y old face score was lower than the measured

score, then the developmental trend for faces would simply have been stronger, thus making an already significant interaction more so).

### **Perceptual discrimination: No face-specific development**

Figure 4a shows discrimination threshold (morph level difference necessary to perform 75% correct) as a function of age. Higher scores denote worse performance.

Results were strikingly different than for memory, with no evidence of specific development for faces. In an ANOVA with between-group factor age (5y, 6y, 7y, 8y, 9y, 10y, adults) and within-group factor category (faces, cars, bodies, scenes), the key age x category interaction was not significant,  $F(18, 321) = .551, p = .932, \eta_p^2 = .030$ , indicating no difference in rates of development across all stimulus types. Main effects of age,  $F(6, 111) = 9.503, p < .001, \eta_p^2 = .348$ , and category,  $F(3, 321) = 9.998, p < .001, \eta_p^2 = .085$ , were significant.

Note that the lack of any face-specific development is a strong null result rather than a weak one. First, it was obtained over a very broad age range (5y to adult) rather than a small range. Second, we tested a large sample (86 children), providing substantial power to detect an age x category interaction if it existed. Third, the age x category interaction did not even approach significance, and faces did not even weakly trend towards showing steeper development than the other categories. This is illustrated across the full age range in Figure 4B-D. Across the 5y to 10y age range we calculated regression slopes to give a numerical value for the rate of development with age. Regressing discrimination threshold on individuals' age gave a slope for faces that fell in



the middle of the other stimuli ( $B$  for faces = -4.91, bodies = -4.95 cars = -4.24, scenes = -5.10) rather than being numerically steeper for faces.

### **Controlling for race and gender effects**

Additional analyses (see supporting information available on-line) found no significant effects of participant race or gender on memory or perceptual discrimination, so neither other-race effects nor possible other-age effects can explain our findings.

### **Lack of face-specific development in perception: not due to low-level strategies or carryover**

We considered, and rejected, two uninteresting explanations of the lack of face-specific development in face perception. First, a supplemental study provides evidence that this was not attributable to the perceptual discrimination task encouraging low-level strategies (e.g., matching on a local image feature), rather than face-level processing. If older age groups were applying low-level strategies, then this would have reduced the amount by which their face perception was better than their object perception, thus reducing the amount of face-specific development. We thus tested 22 adults on the discrimination task in both upright and inverted versions, comparing the size of inversion effect for faces with that for cars. Consistent with the use of face-level processing of the face stimuli (e.g., Yin, 1969, and many subsequent studies), results showed a significant interaction of orientation by stimulus category,  $F(1, 84) = 6.631, p = .018, \eta_p^2 = .240$ , reflecting a large significant inversion effect for faces,  $M(\text{upright}) = 25.4, M(\text{inverted}) =$

42.5,  $F(1, 42) = 19.446$ ,  $p < .001$ ,  $\eta_p^2 = .481$ , but essentially no inversion effect for cars,  $M(\text{upright}) = 34.0$ ,  $M(\text{inverted}) = 34.7$ ,  $F(1, 42) = .027$ ,  $p = .872$ ,  $\eta_p^2 = .001$ .

Second, our perception results cannot be attributed to carryover from the memory task that used the same items. The previous exposure to the item set might well increase overall perceptual discrimination performance across all age groups. However, it cannot explain the particular pattern of perception findings. The memory results show that face-specific memory is stronger in older participants than in younger ones. Thus, any memory carried over from the memory task to the perception task should have produced a similar face-specific developmental increase in face perception. However, this was not the observed result. To explain our *lack* of face-specific development via carryover would require an implausible assumption that either carryover or face perception *decrease* with increasing age.

### Discussion

Our results demonstrate for the first time a developmental dissociation between face memory and face perception. Face memory follows the predictions of the "late maturity" theory, with domain-specific development for faces continuing until 10 years of age (i.e., steeper developmental slope for faces than all nonface classes). Perceptual discrimination follows the predictions of the "early maturity" theory, with no domain-specific development for faces (i.e., equal developmental slopes for faces and nonfaces). This developmental dissociation between face perception and face memory is particularly compelling because it was shown within the same subject group and because the two

tasks used not only the same comparison categories but even the same exemplars from those categories. These results have important implications for the functional architecture of the face processing system, and for the mechanisms by which it develops. First we consider the relationship between our findings and the prior literature.

### **Resolving the conflict in previous studies**

As discussed in the introduction, the five previous studies on whether recognition develops differently for faces versus other stimulus classes revealed considerable internal disagreement. Our results showing a dissociation between perception and memory for faces resolve that disagreement in four out of five cases.

Regarding perceptual discrimination, Johnston et al. (2011) found equal rates of development in  $d'$  accuracy, for same-different decisions for faces and butterflies (see plot in Fig. 3 of McKone et al., 2012; note that the use of the  $d'$  measure was required because response bias changed with age<sup>2</sup>). Thus, both our study and Johnston et al. find no face-specific development for *perception*. Indeed, our result strengthens their conclusion by extending it to a wider range of comparison categories.

Regarding *memory*, our results are consistent with three of the four previous studies. The continuing development of face-specific memory up to 10 years of age in our study agrees with two prior studies that tested children with a similar lower age range (5-6 years). Both of these studies found face-specific memory development compared to houses/scenes (Carey & Diamond, 1977; Golarai et al., 2007). Our results also explain why one study did not find face-specific memory development for houses: Aylward et al (2005) tested older participants (mean of youngest group was 9.3 years) in an age range

where we also see little or no face-specific development. Again, our study also extends these previous conclusions to a wider range of comparison categories, including social stimuli (bodies) and discrete objects (cars).

One discrepancy remains, concerning a memory study. Crookes and McKone (2009) found equal development for faces and Labrador dogs, despite testing memory, despite using a similar delay to here (approximately 1 min on average between study and test presentations of the same item), and despite including the 5 to 10 year age range. One possible account of this discrepancy is that the comparison stimuli in the Crookes & McKone (2009) study (dogs) were processed like faces because they had heads (unlike the body stimuli used here). This idea may not be correct, in that other results argue these same dog stimuli are not in fact processed like faces (cf. the lack of an inversion effect; Robbins & McKone, 2007; Crookes & McKone, 2009), but it could be tested in future by more extensive comparison of development of memory for human faces versus the faces of other animals. Another idea that could, perhaps, explain Crookes and McKone's (2009) finding of similar rates of development of memory for human faces and dogs is that their human faces were in front view while their dog faces were in profile. In babies, ability to discriminate human faces in profile emerges later in development than the ability to discriminate human faces in front or 3/4 view (Fagan, 1979; Turati, Bulf & Simion, 2008). This fact raises the possibility that, to catch up to adult levels of performance, memory development across the childhood age range might have to be steeper for profiles than for front view faces. If correct, this interpretation would suggest that the Crookes and McKone procedure might reveal faster memory development for faces than dogs if the faces of the dogs were matched in viewpoint to the human faces (i.e., both

front view, or both profile), a possibility that could be tested in future research. More generally, it might be useful to test for face-specific memory development using a different type of methodology to that used here (e.g., assessing memory at each age via learning-to-criterion, as in Freire and Lee, 2003, rather than by performance levels for a fixed number of learning trials as used here and in Crookes and McKone, 2009).

Nevertheless, however the conflict is ultimately resolved, the present results show clear domain-specific development of memory for faces.

### **Early versus late maturity of domain-specific face recognition**

For memory, all studies that have attempted to test for domain-specific development for faces have done so using the method we employ here, comparing developmental slopes for faces versus other categories. As noted above, our results agree with three of the four previous such studies, which together lead to the conclusion that there is domain-specific development in memory for faces up until 10 years of age. In terms of current theoretical debate, this is considered "late" maturity.

For perception, we found that perceptual discrimination improved after age 5 but, crucially, this development was *the same for all four stimulus categories tested*. These data argue against domain-specific development of face perception, instead implicating development of domain-general abilities such as perceptual attention. Importantly, this result dovetails not only with Johnston et al (2011)'s previous study of developmental slopes for face versus butterflies, but also with the quantitative maturity found in young children on many other face perception tasks. As reviewed by McKone et al. (2012) this evidence includes early maturity of face-space coding (e.g., as assessed by size of

adaptation aftereffects in children and adults), holistic processing (e.g., as assessed by size of inversion effects<sup>3</sup>, part-whole effects, and composite effects), and the ability to perceptually encode a novel face (as assessed by repetition priming). Taken together, the results now available overturn the classical view that face perception resembles other domains of expertise in requiring ten or more years of experience to reach adultlike levels (Carey, 1992): instead, the evidence for "early" maturity for face-specific *perception* is now very strong.

Overall, our results taken in conjunction with the previous literature make a strong case for different ages of maturity for domain-specific face *perception* (at or before 5 years), and domain-specific face *memory* (10 years).

### **Theoretical origin of the memory/perception dissociation in development**

What might explain this developmental dissociation? That is, how might face-specific memory abilities continue to develop after face-specific perceptual abilities have reached adult levels?

Several obvious proposals that can potentially explain the memory results (stronger development for faces than other categories) in isolation, but do not easily explain the perception results. These ideas are that: experience in individual-level discrimination increases more for faces than other categories; or that verbal rehearsal, which is typically absent in 5 year olds and emerges only in older children is most efficient with faces; or that face-specific scanning strategies in eye movements are optimized between age 5 and 10; or that social development increases attention specifically to faces. It could also be the case that face processing improves with the total

number of faces stored in memory, which is bound to increase with development.

However, none of these ideas straightforwardly explains why parallel face-specific development is not also found for perceptual discrimination.

Thus, we suggest the following speculations. It could be that, of the various face-selective cortical regions present in adults (Kanwisher & Barton, 2011), the more posterior regions (e.g., OFA and FFA) are involved in perception and mature early, while other perhaps more anterior face-selective regions are involved in face-specific memory and mature late (but see Golarai et al., 2007, Scherf et al., 2007). Or, it could be that all face-processing regions have well established function early in life, but white-matter connections supporting face memory (Thomas et al., 2008; 2009) mature late: these could be either connections between different face-selective areas, or connections from face areas to domain-general memory regions (e.g., hippocampal regions). Current data do not allow us to choose between these possibilities, as the neuronal mechanisms underlying any potential behavioral development in face recognition are currently a matter of debate. While a number of empirical studies have found a several-fold increase in the size of the right face-selective fusiform face area (FFA) between children (7-12 years) and adults (Golarai et al., 2007; Scherf, Behrmann, Humphreys & Luna, 2007), another study found little or no difference in selectivity of the FFA between children (4-6 years) and adults (Cantlon, Pinel, Dehaene & Pelphrey, 2011). Further, the face-selective N170 response is also mature by age 5 (Kuefner, de Heering, Jacques, Palmero-Soler & Rossion, 2010). Thus, the time course of development of face-specific neural mechanisms and their connections are not yet clear.

**Face memory and face perception need to be understood separately**

Our finding of a functional dissociation between face perception and face memory dovetails with several prior lines of evidence.

First, in neuropsychology, the distinction between perceptual (“apperceptive”) and memory-based (“associative”) deficits has long been recognized (Lissauer, 1890) for the general case of disorders of visual recognition (agnosia), and occasionally this distinction has been applied (De Renzi et al., 1991) to the case of specific deficits in face recognition (prosopagnosia). Although many prosopagnosic patients have deficits in both face perception and face memory (e.g. Susilo et al., 2010), some have stronger deficits in face memory (McKone et al., 2011; Stollhoff et al., 2011). In the extreme, two neuropsychological patients have been reported (Tippett et al., 2000; Williams et al., 2007), who are impaired at remembering faces, but not other object categories, while at the same time showing no deficit in the perceptual discrimination of faces (“prosopamnesia”). These findings highlight the possibility of separate cognitive and neural systems subserving face memory and face perception.

Second, in autism, deficits in face memory are more robust and widely replicated than deficits in face perception (for a review, see Weigelt et al., 2012). Only a few studies so far have tested whether the deficit in memory in autism is specific to faces (Hauck et al., 1998; McPartland et al., 2011; Wolf et al., 2008), but evidence to date suggests that it may be. Of particular interest is a study by O’Hearn and colleagues (2010), who found that deficits in face memory were maximal in adults, trending in adolescents and only small in children with autism in comparison to their age- and IQ-matched controls. These findings raise the intriguing possibility that the process responsible for normal



development of face memory is impaired in autism. Taken together, these two lines of evidence from prosopagnosia and autism reinforce our conclusion, obtained from typical development, of a functional dissociation between face perception and face memory.

### **Limitations and possible alternative interpretations**

As with any study, we did not test every possible kind of face (and nonface) stimuli, and it is worth considering whether the particular stimuli we chose could have affected our core findings. First, we tested only adult face stimuli, so child participants were farther in age from the stimulus faces than were adult participants. Memory is sometimes better when faces are 'own-age' as the participant (although not always, for a recent review see Macchi Cassia, 2011). Might this "other-age effect" have reduced performance for children in our study, spuriously increasing developmental change? We think not, for several reasons: (a) the adult face stimuli were still other-age for our 10 year old children yet our results showed no face-specific development of memory after age 10; (b) the body stimuli in the memory task were also all adults and yet development of body memory tracked development of car and scene memory not face memory; and (c) any other-age effect that contributes to face *memory* would also likely contribute to face *perception*, yet only face memory showed the face-specific development.

Second, we measured memory and discrimination only for single-image conditions (i.e. same viewpoint, lighting, and emotional expression). It is possible that rates of development in any one category might change if generalization to novel images (e.g., memory for the same face in new viewpoint) were tested. However, this possibility seems unlikely to explain our core findings, because our use of same rather than different

images applied to all four stimulus categories and to both memory and perception tasks. Similar arguments apply to other specific aspects of our stimuli (e.g., static rather than moving, etc.; see Method).

### **Conclusion**

In conclusion, the present results resolve the long-standing debate over whether face recognition develops early or late: late maturity (e.g., Diamond & Carey, 1977) applies to face memory, while early maturity (e.g., Crookes & McKone, 2009) applies to face perception. Beyond their implications for development, our results highlight the importance of distinguishing between face perception and face memory, which has previously played only a minor role in theories of face processing. Important questions remain for future research, concerning the neural basis of the dissociation between face memory and face perception, and the question of whether the developmental change we find after age five is purely maturational, or depends on experience.

### **References**

- Aylward, E. H., Park, J. E., Field, K. M., Parsons, A. C., Richards, T. L., Cramer, S. C., & Meltzoff, A. N. (2005). Brain activation during face perception: evidence of a developmental change. *Journal of Cognitive Neuroscience*, *17*(2), 308-319.  
doi:10.1162/0898929053124884
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function.

- Child Development*, 81(6), 1641-1660. doi: 10.1111/j.1467-8624.2010.01499.x.
- Blaney, R. L., & Winograd, E. (1978). Developmental differences in childrens' recognition memory for faces. *Developmental Psychology*, 14(4), 441-442. doi:10.1037/0012-1649.14.4.441
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433-436. doi:10.1163/156856897X00357
- Cantlon, J. F., Pined, P., Dehaene, S., & Pelphrey, K. A. (2011). Cortical representations of symbols, objects, and faces are pruned back during early childhood. *Cerebral Cortex*, 21(1), 191:199. doi:10.1093/cercor/bhq078
- Carey, S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society London B Biological Science*, 335(1273), 95-102; discussion 102-103. doi:10.1098/rstb.1992.0012
- Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, 195(4275), 312-314. doi:10.1126/science.831281
- Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition*, 2(3), 253-274. doi:10.1080/13506289408402302
- 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. doi:10.1016/j.cognition.2009.02.004
- Dennett, H.W., McKone, E., Edwards, M., & Susilo, T. (2012). Face aftereffects predict individual differences in face recognition ability. *Psychological Science*, 23(11), 1279-1287. doi:10.1177/0956797612446350

- 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
- De Renzi, E., Faglioni, P., Grossi, D., & Nichelli, P. (1991). Apperceptive and associative forms of prosopagnosia. *Cortex, 27*(2), 213-221.
- Fagan, J. (1979). The origins of facial pattern recognition. In M.H. Bornstein & W. Keesen (Eds), *Psychological development from infancy: Image to intention*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Freire, A. & Lee, K. (2001). Face recognition in 4- to 7-year-olds: processing of configural, featural, and paraphernalia information. *Journal of Experimental Child Psychology, 80*(4), 347-71. doi:10.1006/jecp.2001.2639
- Gepner, B., de Gelder, B., & de Schonen, S. (1996). Face processing in autistics: evidence for a generalized deficit? *Child Neuropsychology, 2*(2), 123-139. doi:10.1080/09297049608401357
- Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition, 118*(2), 201-210. doi:10.1016/j.cognition.2010.11.002
- Golarai, G., Ghahremani, D. G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J. L., Gabrieli, J. D., & Grill-Spector, K. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nature Neuroscience, 10*(4), 512-522. doi:10.1038/nn1865
- Hauck, M., Fein, D., Maltby, N., Waterhouse, L., & Feinstein, C. (1998). Memory for faces in children with autism. *Child Neuropsychology, 4*(3), 187-198.

- doi:10.1076/chin.4.3.187.3174
- Johnston, P. J., Kaufman, J., Bajic, J., Sercombe, A., P.T., M., & Karayanidis, F. (2011). Facial emotion and identity processing development in 5- to 15-year-old children. *Frontiers in Psychology, 2*(24), 1-9. doi:10.3389/fpsyg.2011.00026
- Kanwisher, N., & Barton, J. (2011). The functional architecture of the face system: integrating evidence from fMRI and patient studies. In A. Calder, G. Rhodes, M. Johnson, & J. Haxby (Eds), *Oxford Handbook of Face Perception*, (pp. 111 – 128), Oxford: Oxford University Press.
- Kuefner, D., de Heering, A., Jacques, C., Palmero-Soler, E., & Rossion, B. (2010). Early visually evoked electrophysiological responses over the human brain (P1, N170) show stable patterns of face-sensitivity from 4 years to adulthood. *Frontiers in Human Neuroscience, 6*(3), 67. doi:10.3389/neuro.09.067.2009
- Lawrence, K., Bernstein, D., Pearson, R., Mandy, W., Campbell, R., & Skuse, D. (2008). Changing abilities in recognition of unfamiliar face photographs through childhood and adolescence: performance on a test of non-verbal immediate memory (Warrington RMF) from 6 to 16 years. *Journal of Neuropsychology, 2*(Pt 1), 27-45. doi:10.1348/174866407X231074
- Lissauer, H. (1890). Ein Fall von Seelenblindheit nebst einem Beitrage zur Theorie derselben. *European Archives of Psychiatry and Clinical Neuroscience, 21*(2), 222-270. doi:10.1007/BF02226765
- Macchi Cassia, V. (2011). Age biases in face processing: the effects of experience across development. *British Journal of Psychology, 102*(4), 816-829. doi:10.1111/j.2044-8295.2011.02046.x

- McKone, E., Crookes, K., Jeffery, L., & Dilks, D. D. (2012). A critical review of the development of face recognition: experience is far less important than previously believed. *Cognitive Neuropsychology*. doi:10.1080/02643294.2012.660138.
- McKone, E., Hall, A., Pidcock, M., Palermo, R., Wilkinson, R. B., Rivolta, D., Yovel, G., Davis, J.M., & O'Connor, K. B. (2011). Face ethnicity and measurement reliability affect face recognition performance in developmental prosopagnosia: evidence from the Cambridge Face Memory Test-Australian. *Cognitive Neuropsychology*, 28(2), 109-146. doi:10.1080/02643294.2011.616880
- McPartland, J. C., Webb, S. J., Keehn, B., & Dawson, G. (2011). Patterns of visual attention to faces and objects in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 41(2), 148-157. doi:10.1007/s10803-010-1033-8.
- 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
- O'Hearn, K., Schroer, E., Minshew, N., & Luna, B. (2010). Lack of developmental improvement on a face memory task during adolescence in autism. *Neuropsychologia*, 48(13), 3955-3960. doi:10.1016/j.neuropsychologia.2010.08.024
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10(4), 437-442. doi:10.1163/156856897X00366
- Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, 103(1), 34-79.

- doi:10.1016/j.cognition.2006.02.008
- Scherf, K. S., Behrmann, M., Humphreys, K., & Luna, B. (2007). Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Developmental Science*, *10*(4), F15-30. doi: 10.1111/j.1467-7687.2007.00595.x
- Stollhoff, R., Jost, J., Elze, T., & Kennerknecht, I. (2011). Deficits in Long-Term Recognition Memory Reveal Dissociated Subtypes in Congenital Prosopagnosia. *PLoS ONE* *6*(1), e15702. doi:10.1371/journal.pone.0015702
- Susilo, T., McKone, E., Dennett, H., Darke, H., Palermo, R., Hall, A., Pidcock, M., Dawel, A., Jeffery, L., & Rhodes, G. (2010). Face recognition impairments despite normal holistic processing and face space coding: evidence from a case of developmental prosopagnosia. *Cognitive Neuropsychology*, *27*(8), 636-664. doi:10.1080/02643294.2011.613372
- Thomas, C., Avidan, G., Humphreys, K., Jung, K. J., Gao, F., & Behrmann, M. (2009). Reduced structural connectivity in ventral visual cortex in congenital prosopagnosia. *Nature Neuroscience*, *12*(1), 29-31. doi:10.1038/nn.2224
- Thomas, C., Moya, L., Avidan, G., Humphreys, K., Jung, K. J., Peterson, M. A., & Behrmann, M. (2008). Reduction in white matter connectivity, revealed by diffusion tensor imaging, may account for age-related changes in face perception. *Journal of Cognitive Neuroscience*, *20*(2), 268-284. doi:10.1162/jocn.2008.20025
- Tippett, L. J., Miller, L. A., & Farah, M. J. (2000). Prosopamnesia: a selective impairment in face learning. *Cognitive Neuropsychology*, *17*(1), 241-255. doi:10.1080/026432900380599

- Turati, C., Bulf, H., & Simion, F. (2008). Newborns' face recognition over changes in viewpoint. *Cognition*, *106*(3), 1300-1321. doi:10.1016/j.cognition.2007.06.005
- Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science*, *23*(2), 169-177. doi:10.1177/0956797611420575
- Want, S. C., Pascalis, O., Coleman, M., & Blades, M. (2003). Face facts: Is the development of face recognition in early and middle childhood really so special? In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood* (pp. 207-221). New York: Nova Science Publishers.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113-120. doi: 10.3758/BF03202828
- Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience & Biobehavioral Reviews*, *36*(3), 1060-1084. doi:10.1016/j.neubiorev.2011.12.008
- Williams, M. A., Berberovic, N., & Mattingley, J. B. (2007). Abnormal fMRI adaptation to unfamiliar faces in a case of developmental prosopamnesia. *Current Biology*, *17*(14), 1259-1264. doi:10.1016/j.cub.2007.06.042
- Wolf, J. M., Tanaka, J. W., Klaiman, C., Cockburn, J., Herlihy, L., Brown, C., South, M., McPartland, J., Kaiser, M. D., Phillips, R., & Schultz, R. T. (2008). Specific impairments of face-processing abilities in children with autism spectrum disorders using the Let's Face It! skills battery. *Autism Research*, *1*(6), 329-340. doi:10.1002/aur.56



Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1), 141-145. doi:10.1037/h0027474

### Footnotes

<sup>1</sup> We review only studies where the evaluation of the age x category interaction is methodologically sound, in that (a) performance was matched between categories in one age group, and (b) results are not marred by floor/ceiling effects. (Additional studies not meeting these criteria are discussed in Crookes & McKone, 2009).

<sup>2</sup> Accuracy ( $d'$ ) was well matched in the youngest age group across face identity and butterflies, making the developmental comparison valid for this measure. For reaction time, RTs for face identity discrimination improved at a steeper rate with age than RTs for butterfly discrimination. However, the tasks were not matched for RT in any age group. The face identity task RTs were consistently slower than the butterfly task RTs, meaning that (a) the effects of age are being evaluated out of different baselines, and (b) the results are open to the interpretation that RTs simply reach adult levels earlier on the task that is fastest already in the youngest children (rather than having anything to do with the stimulus category per se).

<sup>3</sup> The conclusion that the face inversion effect does not increase with age comes from studies, which have avoided restriction of range problems in the extreme age groups (see review in McKone et al., 2012). There are other studies that have been argued to show smaller inversion effects in children; notably for perceptual matching in the Benton task (de Heering, Rossion & Maurer, 2012). However, as argued by McKone et al (2012), such findings can commonly be accounted for by restriction of range in the youngest children. If the adult inversion effect on face perception is approximately 15% as in de Heering et al. (2012), then accuracy in younger age groups of anything less than 65% (as occurred in de Heering's 6-8 yr old children) restricts the range to see this size of effect.

This is the case even if upright is significantly greater than chance (as in de Heering et al., 2012).

### Figure legends

*Figure 1.* The memory paradigm.

The study phase consisted of 10 items per category (faces and cars, or bodies and scenes), each presented at screen center for 3 s. A 2AFC test phase consisting of 10 pairs of stimuli per category (one studied, one new) followed immediately after.

*Figure 2.* The perception paradigm.

2A. Perceptual discrimination was tested via an immediate 2AFC match-to-sample task, separately for the four categories. A target item was presented at screen center for 1 s, and immediately followed by a test pair of stimuli. Each test pair included the target item and a distractor item created by morphing the target item towards a different-identity exemplar of the same category. 2B shows examples of morph continua, one example from each stimulus category. There were ten morph continua per category.

*Figure 3.* The memory data.

Memory accuracy in percent correct as a function of age for the four categories (faces, cars, bodies, scenes) indicating domain-specific memory development only for faces.

Chance was 50% indicated by the horizontal line. Error bars denote SEM.

*Figure 4.* The perception data.

A. Perceptual discrimination threshold (morph-level difference necessary to perform at 75% correct) as a function of age for the four categories (faces, cars, bodies, scenes).

B-D. Perceptual discrimination threshold (morph-level difference necessary to perform at 75% correct) as a function of age contrasting faces with each other category in turn.

Results indicate domain-general improvements for the four categories under study. Error bars denote SEM.

Figure 1

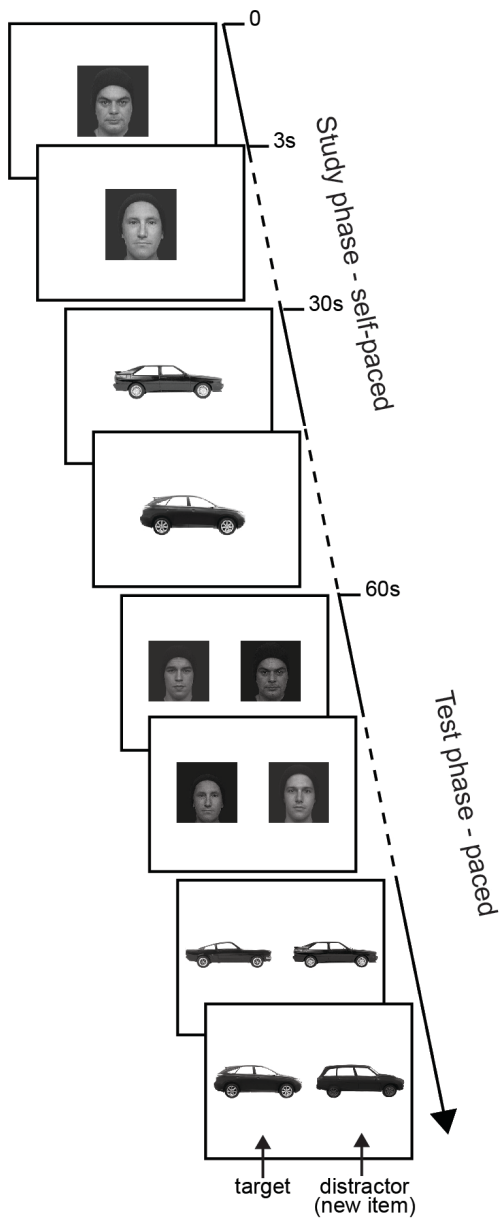


Figure 2

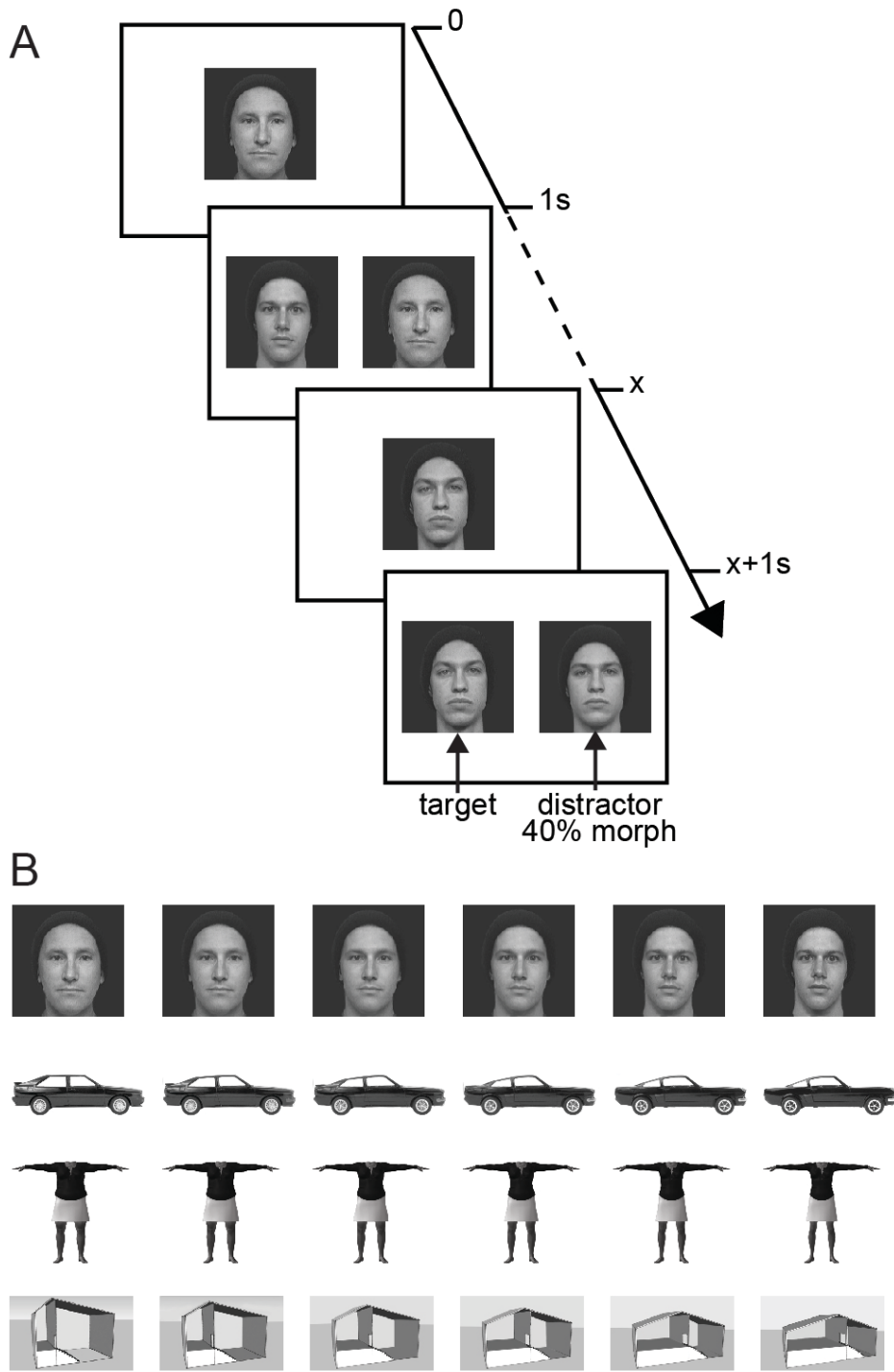


Figure 3

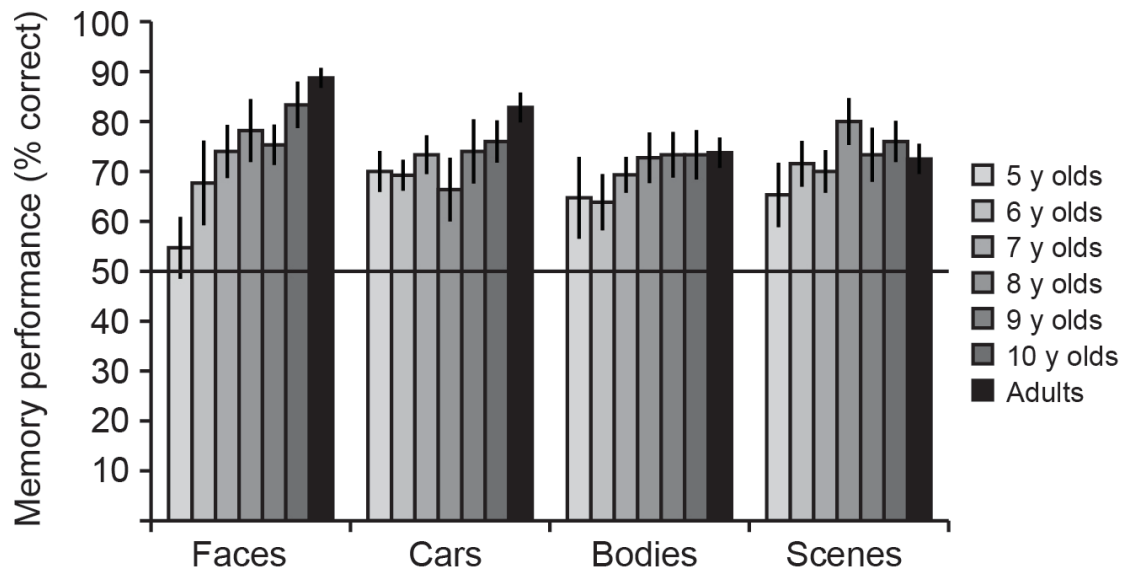




Figure 4

