



Cognitive Science 45 (2021) e13021

© 2021 Cognitive Science Society LLC

ISSN: 1551-6709 online

DOI: 10.1111/cogs.13021

Distorting Face Representations in Newborn Brains

Samantha M. W. Wood,^a  Justin N. Wood^{a,b,c,d}

^a*Informatics Department, Indiana University*

^b*Center for the Integrated Study of Animal Behavior, Indiana University*

^c*Cognitive Science Program, Indiana University*

^d*Department of Neuroscience, Indiana University*

Received 16 September 2020; received in revised form 8 June 2021; accepted 24 June 2021

Abstract

What role does experience play in the development of face recognition? A growing body of evidence indicates that newborn brains need slowly changing visual experiences to develop accurate visual recognition abilities. All of the work supporting this “slowness constraint” on visual development comes from studies testing basic-level object recognition. Here, we present the results of controlled-rearing experiments that provide evidence for a slowness constraint on the development of face recognition, a prototypical subordinate-level object recognition task. We found that (1) newborn chicks can rapidly develop view-invariant face recognition and (2) the development of this ability relies on experience with slowly moving faces. When chicks were reared with quickly moving faces, they built distorted face representations that largely lacked invariance to viewpoint changes, effectively “breaking” their face recognition abilities. These results provide causal evidence that slowly changing visual experiences play a critical role in the development of face recognition, akin to basic-level object recognition. Thus, face recognition is not a hardwired property of vision but is learned rapidly as the visual system adapts to the temporal structure of the animal’s visual environment.

Keywords: Face recognition; Object recognition; Controlled rearing; Chicks; Slowness constraint; Unsupervised temporal learning

1. Introduction

To navigate complex social worlds, humans and animals depend heavily on their ability to recognize faces. Numerous studies have examined face recognition (i.e., the ability to encode and recognize specific faces)¹ after individuals have acquired months to years of visual

Correspondence should be sent to Samantha M. W. Wood or Justin N. Wood, Informatics Department, Indiana University, Bloomington, Indiana 47408, USA. E-mail: sw113@indiana.edu; woodjn@indiana.edu

experience (e.g., Carey & Diamond, 1977; de Haan, Johnson, Maurer, & Perrett, 2001; Frank, Vul, & Johnson, 2009; Kelly et al., 2007; Pascalis, Deschonen, Morton, Deruelle, & Fabregrenet, 1995). Despite this progress, fundamental questions remain about the origins and development of face recognition. Can newborn brains begin building high-level (view-invariant) face representations at the onset of vision? If so, what role does early visual experience play in the development of face recognition? Here, we address these questions with controlled-rearing experiments, using newborn chicks as a model system.

1.1. *Controlled-rearing studies of face recognition*

Studying the origins of face recognition requires studying newborn organisms. A small number of studies have explored the origins of face recognition in newborn humans. These studies have started to reveal the initial capacities and limitations of face recognition. For example, newborn infants can recognize their mother within the first few days of life (Bushnell, 2001; Pascalis et al., 1995). This early emerging ability appears to be driven largely by recognition of hair features (Pascalis et al., 1995; Turati, Macchi Cassia, Simion, & Leo, 2006), although there is evidence that newborn infants can recognize faces without hair cues by extracting low spatial frequency information (De Heering et al., 2008) and motion cues (Bulf & Turati, 2010).

One limitation of studies with newborn humans is that their visual environment cannot be strictly controlled and manipulated from birth. As a result, newborn infants have typically acquired several days of natural visual experience prior to testing, which obscures the contribution of early visual experience on the development of face recognition. To overcome this limitation, we can turn to controlled-rearing studies of newborn animals. In controlled-rearing experiments, researchers can manipulate the visual experiences provided to newborn subjects and measure the effects of those experiences on behavioral and neural development (Arcaro, Schade, Vincent, Ponce, & Livingstone, 2017; Held & Hein, 1963; Hubel & Wiesel, 1962; Matteucci & Zoccolan, 2020; Walk, Gibson, & Tighe, 1957; Wood, 2014). With controlled rearing, it is possible to determine how specific visual experiences shape early visual development.

A few controlled-rearing studies have explored the role of visual experience in the development of face recognition. In one seminal study, Sugita (2008) reared infant monkeys with no exposure to faces for 6–24 months, then tested the monkeys' face-processing abilities. After a month of face experience, the monkeys showed a preference for faces and were able to distinguish familiar faces from novel faces. Unfortunately, this result has been erroneously interpreted as evidence that face recognition can develop *without* face experience (see Arcaro et al., 2017). During the face deprivation period, the monkeys did not show a statistically significant preference for faces (p. 395, lines 26–30), nor did the monkeys show statistically significant evidence for face discrimination (p. 395, lines 52–56). Thus, this study suggests that face experience is required for the development of face recognition.

More recently, Arcaro et al. (2017) reared infant monkeys without exposure to faces. As in Sugita (2008), the face-deprived monkeys did not preferentially look at faces. The monkeys also did not develop face-selective cortical regions. Conversely, when monkeys were reared

with faces, they developed both a preference for faces and face-selective cortical regions. Together, these two controlled-rearing studies suggest that experience with faces is necessary to develop face recognition.

These studies raise additional questions about the role of experience in the development of face recognition. First, how much face experience is needed to develop face recognition? Can newborn brains begin building high-level (view-invariant) face representations from experience with a single face, or does the development of this ability require experience with multiple faces?

Second, do newborn brains need a particular *type* of face experience to develop face recognition? In natural visual environments, objects (including faces) tend to move slowly over time, and newborn brains might leverage this temporally structured training data to build accurate face representations (Masquelier & Thorpe, 2007; Wallis & Rolls, 1997; Wiskott & Sejnowski, 2002; Wyss, Konig, & Verschure, 2006). The logic of this learning process is as follows: during natural visual experience, the environment changes slowly (i.e., faces are typically present for seconds or longer; Jayaraman & Smith, 2019), whereas the primary sensory signal changes rapidly (i.e., photoreceptors change on a timescale of milliseconds). Thus, neural mechanisms that extract slowly varying features from quickly varying input signals should create stable representations of the external causes of the sensory input. While there is evidence that *mature* brains leverage temporally structured visual experiences to build face representations (Wallis & Bühlhoff, 2001), it is unknown whether *newborn* brains rely on this same strategy.

In prior work, we discovered that the development of object recognition requires visual experience with slowly moving objects (Wood & Wood, 2016). When newborn chicks were reared with slowly moving objects, the chicks built invariant object representations that generalized across new viewpoints and rotation speeds. Conversely, when chicks were reared with quickly moving objects, the chicks built inaccurate object representations that failed to generalize across new viewpoints and rotation speeds. Thus, in the case of basic-level object recognition, invariant object recognition is not a hardwired property of vision; rather, invariant recognition is learned rapidly as newborn animals adapt to their slowly changing visual environment. At this point, however, it is unknown whether this slowness constraint also shapes the development of face recognition. Indeed, researchers have long noted that object recognition and face recognition differ in many respects. For example, unlike most object recognition tasks, face recognition requires distinguishing between two faces that share a common configuration of features, making face recognition a subordinate-level recognition task (Gauthier et al., 2000). There is also evidence that object recognition and face recognition depend on at least partially separable neural substrates (Kanwisher, 2010; Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997). Given these differences, it is an open question whether face recognition and object recognition are subject to common developmental constraints.

Third, what role does object experience play in the development of face recognition? Since the monkeys in Sugita (2008) and Arcaro et al. (2017) were raised in environments containing a variety of objects, it is possible that the monkeys' experiences with objects influenced the development of face recognition. To isolate the role that face experience

plays in the development of face recognition, it is necessary to raise newborn animals in environments that control both face *and* object experience.

To address these questions, we developed an automated controlled-rearing method for studying the origins of face recognition, using newborn chicks as a model system. Unlike monkeys, newborn chicks can be reared in strictly controlled environments that lack faces, objects, and caregivers. In a prior study (Wood & Wood, 2015b), we reared newborn chicks in environments containing no objects other than a single virtual face, then tested whether the chicks could distinguish that familiar face from novel faces. The chicks successfully distinguished the familiar face from most of the unfamiliar faces. We found that the chicks were sensitive to changes in the face's age, gender, and orientation (upright vs. inverted). Thus, in the absence of extensive face and object experience, newborn brains can build reasonably accurate face representations.

Here, we extend our prior study in two important ways. First, in Wood and Wood (2015b), we only tested face recognition across familiar viewpoints. Thus, the chicks could have succeeded in the task by using a simple “image-matching” strategy (i.e., matching visible patterns of retinal activity to patterns of retinal activity stored in memory). In contrast, face recognition in the real world is far more challenging, requiring the use of invariant face representations that can generalize across novel viewing situations (e.g., changes in viewpoint, size, lighting, and background features). Indeed, from an artificial intelligence perspective, building visual representations that support image matching is trivial, whereas building representations that are selective for particular faces and tolerant to identity-preserving image transformations requires a more complex computational system (DiCarlo & Cox, 2007). Can newborn chicks solve this complex computational task, converting raw visual inputs into invariant face representations that generalize across novel viewpoints?

Second, while we *controlled* the visual input in the chicks' environment in Wood and Wood (2015b), we did not *manipulate* that input. Thus, we could not distinguish whether the visual features that supported face recognition were innate (hardwired into the visual system) or whether the features emerged from the particular visual experiences that the chicks acquired during postnatal development. As discussed above, a large body of work in the neurosciences suggests that the development of visual recognition requires slowly changing visual experiences. According to these unsupervised temporal learning (UTL) models, the brain leverages the temporal structure of natural visual environments to develop invariance (tolerance) to identity-preserving image changes (Masquelier & Thorpe, 2007; Wallis & Rolls, 1997; Wiskott & Sejnowski, 2002; Wyss et al., 2006). If UTL governs the development of face recognition, then it should be possible to distort face representations in newborn brains simply by manipulating the speed of a face when it is being encoded into memory. Thus, by rearing chicks with faces that rotate at different speeds, we can explore whether UTL mechanisms underlie the development of face recognition.

To preview our findings, Experiment 1 tested whether newborn chicks can perform view-invariant face recognition. We found that chicks can distinguish familiar faces from unfamiliar faces across novel viewpoints. In Experiments 2 and 3, we found that the development of this ability relies on experience with slowly moving faces, as predicted by UTL models. Moreover, we discovered that it is possible to distort a newborn chick's first face representation simply

by manipulating the face's rotation speed during encoding. Overall, our findings indicate that (1) newborn brains are capable of "one-shot" view-invariant face recognition, and (2) the development of face recognition is subject to a slowness constraint, akin to basic-level object recognition.

2. Experiment 1

We first tested whether newborn chicks can perform view-invariant face recognition. We adapted the view-invariant recognition task used previously to study basic-level object recognition (Wood, 2013, 2015; Wood & Wood, 2015a). In this prior work, we reared chicks with one object seen from a limited viewpoint range (e.g., 60° of rotation), then tested whether the chicks could recognize their imprinted object from novel viewpoints. The chicks successfully recognized their imprinted object across novel viewpoints, indicating that newborn visual systems are capable of view-invariant recognition when objects differ on the basic level (i.e., different configurations of features). Can newborn chicks also succeed in a view-invariant recognition task when the objects share the same configuration of features, such as face recognition?

2.1. Methods

2.1.1. Subjects

Ten Rhode Island Red chicks of unknown sex were tested. No subjects were excluded from the analyses. The sample size was based on previous automated controlled-rearing experiments with newborn chicks (Wood & Wood, 2015b; Wood & Wood, 2016). We also note that our sample size was highly powered for our expected effect size. Specifically, in our prior face recognition paper (Wood & Wood, 2015b), we obtained a group effect size (Cohen's *d*) of 1.191 across the conditions corresponding to the face stimuli used here (all gender and age conditions). A sample size of eight subjects is required to achieve 80% power for an effect size of 1.191 (R package *pwr*; R Core Team, 2015, Version 3.1.3). Because we tested 10 chicks in Experiment 1, our experiment was powered to 91.658% to detect an effect size of 1.191.

The eggs were obtained from a local distributor and incubated in darkness in an OVA-Easy incubator (Brinsea Products Inc., Titusville, FL). The temperature was maintained at 99.6°F and the humidity was maintained at 45% for the first 19 days of incubation. On day 19, we increased the humidity to 60%. After hatching, the chicks were moved from the incubator room to the controlled-rearing chambers (Figure 1a) in darkness with the aid of night vision goggles. Each chick was housed singly in its own chamber. This research was approved by the University of Southern California Institutional Animal Care and Use Committee.

2.1.2. Controlled-rearing chambers

The chicks were reared for 2 weeks in controlled-rearing chambers (66 cm length × 42 cm width × 69 cm height). The chambers were constructed from white, high-density plastic. Two

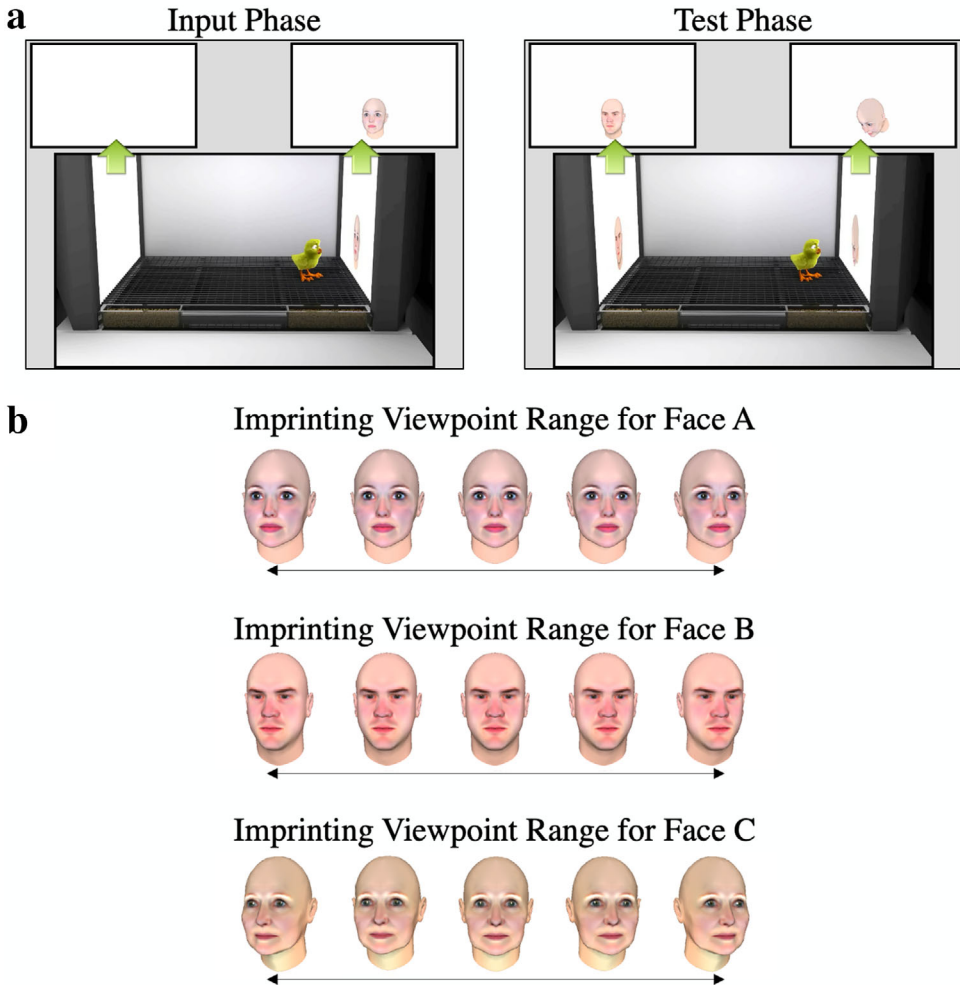


Fig. 1. Experiment 1 methods. (a) During the input phase, we reared newborn chicks with a single virtual face rotating through a 20° viewpoint range. During the test phase, we presented the chicks with two-alternative forced-choice test trials. One display wall showed the imprinted face from an unfamiliar viewpoint, while the other display wall showed an unfamiliar face from the familiar (imprinted) viewpoint. (b) Each chick was imprinted to either face A or face B. The nonimprinted face was used as an unfamiliar face during test trials. Face C also served as an unfamiliar face during test trials for both groups of chicks.

of the chamber walls were display walls (19" liquid crystal display monitors with 1440×900 pixel resolution). The display walls were situated on opposite sides of the chamber and were used to present the face stimuli to the chicks.

The chambers contained no rigid, bounded objects other than the virtual faces presented on the display walls. All care of the chicks was performed in darkness with the aid of night vision goggles. Food and water were provided *ad libitum* in transparent, rectangular troughs in the ground (66 cm length \times 2.5 cm width \times 2.7 cm height). We used grain as food because

grain does not behave like an object (i.e., grain does not maintain a rigid, bounded shape). The floors were black wire mesh supported over a black surface by thin, transparent beams.

We recorded all of the chicks' behavior continuously (24/7) across the duration of the 2-week experiment. Microcameras (1.5 cm) embedded in the ceilings of the chambers recorded all of the chicks' behavior. Ethovision XT software (Noldus Information Technology) used the microcamera footage to track the chicks' behavior. The software calculated the amount of time each chick spent within zones (22 cm × 42 cm) next to the left and right display walls. This automated data collection approach made it possible to measure each chick's performance with high precision (Wood & Wood, 2019). In total, 3360 h of video footage (14 days × 24 h per day × 10 subjects) were collected for Experiment 1.

2.1.3. *General logic of design*

As discussed in Wood and Wood (2021), our controlled-rearing method is designed to mimic the experimental design from machine learning studies of artificial neural network (ANN) models. ANN models are provided with a set of training data for learning during a "training phase." Subsequently, learning is turned off, and the ANN model is tested on a variety of novel stimuli. Results from this "test phase" allow researchers to measure what the model learned from the specific set of data presented during the training phase. Likewise, in our study, we presented chicks with a set of training data during the input phase (a single face moving through a limited viewpoint range). To "turn off" learning, we relied on the naturally occurring sensitive period in filial imprinting. The imprinting window for visual learning ends ~3 days after hatching, so by rearing the chicks with their imprinted face for 1 week, we could generally ensure that visual learning of the imprinted face had turned off. Then, during the test phase, we measured what the chicks had learned from the training data provided in the input phase. Our study uses the same experimental design as machine learning studies to encourage researchers to make direct comparisons between the learning abilities of newborn chicks and ANN models.

2.1.4. *Input phase*

During the input phase (first week of life), the chicks were reared in an environment that contained a single virtual face (ear-to-ear width = 6.2–6.5 cm; height = 10 cm; distance above flooring unit = 1 cm; Figure 1b). The virtual face moved continuously, rotating through a 20° viewpoint range about a vertical axis passing through its centroid. The individual frames were created using FaceGen software (Singular Inversion, Inc.) and concatenated into an animation using QTCoffee (3AM Coffee Software). The face was displayed on a uniform white background. During the input phase, the imprinted face switched display walls every 2 h, following a 1-min period of darkness. The face appeared for an equal amount of time on each display wall. Video S1 shows a sample animation presented during the input phase.

As in our prior work (Wood & Wood, 2015b), we used human faces (rather than chicken faces) because the face images could be precisely manipulated using the FaceGen software. Moreover, using human faces allows for a more direct comparison with studies of face recognition in humans and other avian species (e.g., Bogale, Aoyama, & Sugita, 2011; Gibson, Wasserman, Gosselin, & Schyns, 2005; Troje, Huber, Loidolt, Aust, & Fieder, 1999).

Unlike traditional controlled-rearing studies of newborn chicks, in which the chick typically receives a short imprinting phase (e.g., 2 h), our design included a long input phase (7 days). By extending the input phase to 7 days, we could ensure that the imprinting period had ended and that the chicks would not imprint to the test stimuli (Wood & Wood, 2015a). Using a long input phase also improved the ecological validity of our task. In the real world, chicks have multiple days to imprint (i.e., chicks are not restricted to a short 2-h imprinting window).

2.1.5. Test phase

During the test phase (second week of life), we used a two-alternative forced-choice task to measure whether the chicks could recognize their imprinted face across familiar and novel viewpoints. On each test trial, the imprinted face was presented on one display wall and an unfamiliar face was presented on the other display wall. The faces rotated through a 20° viewpoint range. The imprinted face was shown from one of 20 viewpoint ranges (the imprinted viewpoint range plus 19 novel viewpoint ranges) on each test trial. The unfamiliar faces had the same size, motion speed, and viewpoint range as the imprinted face from the input phase. Consequently, on most of the test trials, the unfamiliar face was more similar to the imprinting stimulus than the imprinted face was to the imprinting stimulus (from a pixel-level perspective). Thus, to recognize their imprinted face, the chicks needed to generalize across large changes in the face's appearance. If newborn chicks can recognize faces across novel viewpoints, then they should have spent a greater proportion of time in proximity to the imprinted face than the unfamiliar face during the test trials. Video S2 shows a sample animation presented during the test phase.

Six of the chicks were imprinted to face A (with face B serving as an unfamiliar face), and four of the chicks were imprinted to face B (with face A serving as an unfamiliar face). Both groups were also tested with a third face (face C) that was unfamiliar to both groups of chicks (Figure 1b).

Unlike traditional controlled-rearing studies of newborn chicks, in which the chick is typically tested for a short period of time (e.g., a single 6-min test trial), our design included a long test phase (7 days). We used a long test phase because we wanted to collect precise measurements of performance from each chick. In previous work, we found that increasing the amount of data collected from each chick massively *decreases* measurement error and *increases* effect sizes (Wood & Wood, 2019). Thus, collecting more data per chick improved the precision of the data and the power of the experiment. Using a long test phase also improved the ecological validity of our task. In the real world, after chicks imprint, they spend several weeks in proximity to their imprinted objects. During this time, chicks repeatedly separate and reunite with their imprinted objects (such as when foraging for food). Accordingly, our test phase was more similar to the repeated-recognition task that chicks face in nature.

Each test trial was 12 min in duration followed by a 24-min rest period. During the rest periods, the imprinted face appeared on one display wall and a white screen appeared on the other display wall. We included rest periods in the design for three reasons. First, the rest periods allowed the chicks to reunite with their imprinted face without needing to make a choice between two faces. Second, the rest periods provided a baseline measure of the amount of time the chicks generally preferred to spend in proximity to their imprinted face. Third, the

rest periods allowed us to check whether the chicks stayed attracted to their imprinted face across the test phase. Each subject received 40 test trials per day.

2.2. Results

The results are shown in Figure 2. To measure the chicks' performance, we computed the percent of time the chicks spent with the imprinted face compared to the unfamiliar face (1) on the trials in which the imprinted face switched display walls from the preceding rest period and (2) on the trials in which the imprinted face did not switch display walls from the preceding rest period. Then, we computed the average of these two values to obtain a single recognition performance score for each chick.² Chance performance was 50%. Across all of the test trials, the chicks spent significantly more time with their imprinted face than the unfamiliar face ($M = 67\%$, $SD = 5\%$; one-sample t -test, $t(9) = 10.76$, $p = .000002$, Cohen's $d = 3.40$). Recognition performance was also high when we removed the test trials in which the imprinted face was shown from the familiar viewpoint range ($M = 67\%$, $SD = 5\%$; one-sample t -test, $t(9) = 11.13$, $p = .000001$, Cohen's $d = 3.52$).

To examine whether the chicks showed impaired recognition performance across larger viewpoint changes, we grouped the test trials into three categories: 0° viewpoint change (the middle frame of the test animation was the same as the middle frame of the imprinted animation), $\pm 25^\circ$ viewpoint change (the middle frame of the test animation differed $\pm 25^\circ$ from the middle frame of the imprinted animation), and $\pm 50^\circ$ viewpoint change (the middle frame of the test animation differed $\pm 50^\circ$ from the middle frame of the imprinted animation). Then, we performed a repeated-measures ANOVA with the within-subjects main effect of viewpoint change. We found a significant effect of viewpoint change ($F(2, 18) = 6.89$, $p = .006$, $\eta^2 = .43$). Post-hoc paired-sample t -tests revealed that performance in the 0° test trials was significantly higher than both the $\pm 25^\circ$ test trials (Mean Difference = 4%, $SD = 5\%$, $t(9) = 2.66$, $p = .026$, Cohen's $d = .84$) and the $\pm 50^\circ$ test trials (Mean Difference = 6%, $SD = 6\%$, $t(9) = 3.26$, $p = .010$, Cohen's $d = 1.03$). Importantly, however, performance was still well above chance level in all three viewpoint change conditions (one-sample t -tests, 0° viewpoint change: $M = 72\%$, $SD = 7\%$, $t(9) = 9.76$, $p = .000004$, Cohen's $d = 3.09$; $\pm 25^\circ$ viewpoint change: $M = 67\%$, $SD = 5\%$, $t(9) = 10.35$, $p = .000003$, Cohen's $d = 3.27$; $\pm 50^\circ$ viewpoint change: $M = 65\%$, $SD = 6\%$, $t(9) = 8.20$, $p = .00002$, Cohen's $d = 2.59$).

To test whether performance varied as a function of test day, we performed a repeated-measures ANOVA with the within-subjects factor of test day. The ANOVA did not reveal a significant main effect of test day ($F(6, 54) = .83$, $p = .55$). Performance was also above chance level on all test days (one-sample t -tests, Holm–Bonferroni corrected for multiple comparisons, all $ps < .001$). The chicks immediately achieved maximal levels of performance and did not improve significantly across the test phase.

2.3. Discussion

These results suggest that newborn chicks can perform view-invariant face recognition. As with basic-level object recognition (Wood, 2013), newborn chicks can recognize an imprinted face across large, novel, and complex changes in the face's appearance.

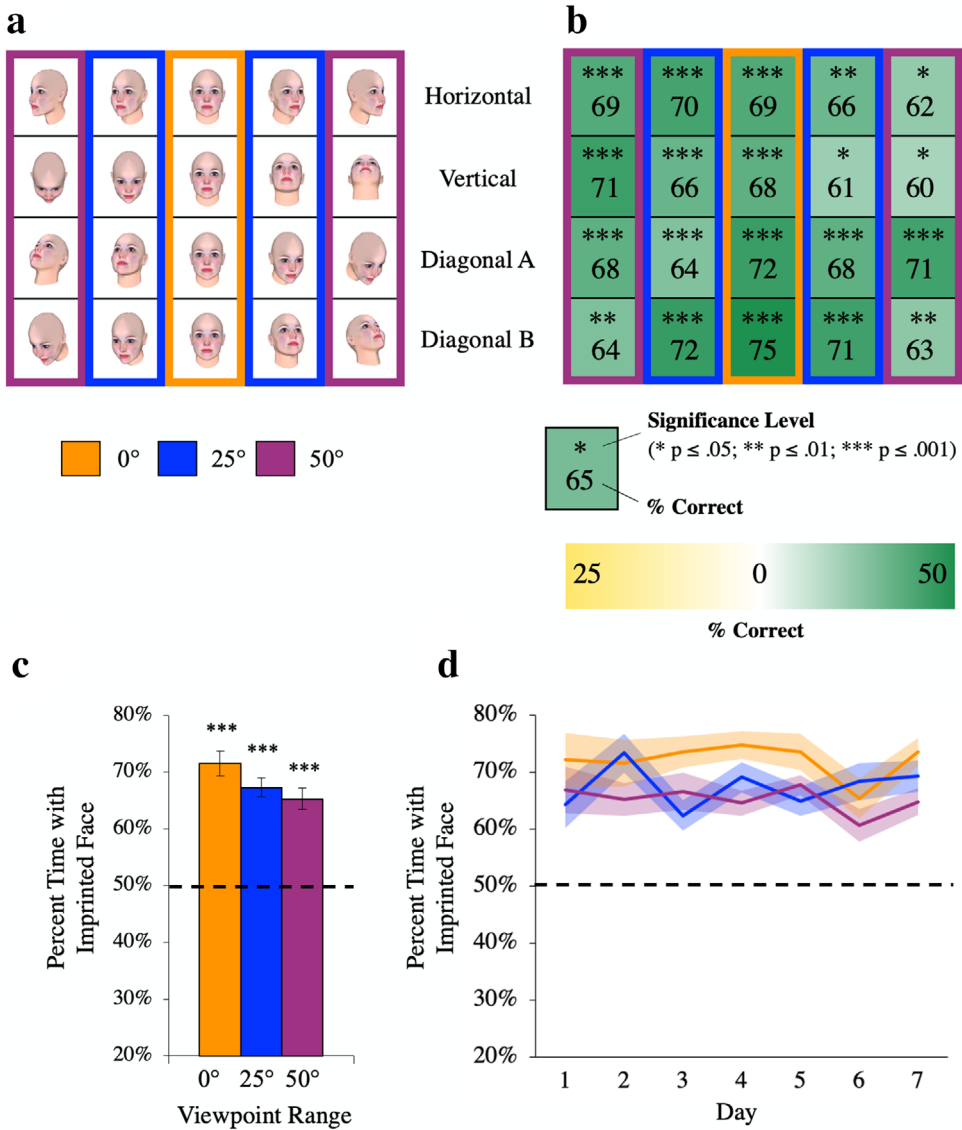


Fig. 2. Stimuli and results for Experiment 1. (a) The viewpoints presented during the test phase. The table shows the middle frame of each viewpoint range. (b) The chicks' face recognition performance for each viewpoint range. Performance (percent correct) is both color-coded and reported as a numeric value, together with its significance according to a two-tailed *t*-test (see key for significance values). Performance was above chance level (50%) for each viewpoint. (c) The bar graphs show recognition performance as a function of the three degrees of viewpoint change. The chicks' recognition performance was above chance level for all three of the viewpoint ranges (0°, 25°, and 50°). (d) Overall performance by test day for each viewpoint range. Performance did not improve over the course of the test phase. The shaded ribbon around each line shows ± 1 SE.

We next examined whether the development of view-invariant face recognition requires experience with slowly changing faces. In prior studies, we discovered that experience with slowly changing objects is necessary for the development of view-invariant object recognition (Wood, 2016; Wood & Wood, 2016, 2018; Wood, Prasad, Goldman, & Wood, 2016). Specifically, when chicks were reared with objects rotating slowly, the chicks built view-invariant object representations that generalized across novel viewpoints and rotation speeds. Conversely, when chicks were reared with objects rotating quickly, the chicks built view-specific representations that failed to generalize across novel viewpoints and rotation speeds. Thus, with objects, it is possible to systematically manipulate the amount of view-invariance in a newborn chick's first object representation simply by changing the object's rotation speed during the learning period. Can we also systematically manipulate the amount of view-invariance in a newborn chick's first face representation, as predicted by UTL models?

3. Experiment 2

During the input phase, newborn chicks were reared with a virtual face that rotated at a slow, medium, or fast speed. During the test phase, we measured the chicks' sensitivity to view-specific features and identity features. We measured sensitivity to these two feature types because building an invariant face representation requires converting retinal input (view-specific features) into a high-level representation that is tolerant to retinal image variation and selective for a particular face (identity features). Accordingly, measuring sensitivity to view-specific features and identity features reveals whether the chick successfully built a view-invariant face representation.

3.1. Methods

3.1.1. Subjects

Thirty-five Rhode Island Red chicks of unknown sex were tested; we tested 11-12 chicks per condition (11 slow, 12 medium, and 12 fast). No subjects were excluded from the analyses. The incubation procedures and test chambers were identical to those used in Experiment 1. We recorded the chicks' behavior 24/7 across the duration of the 2-week experiment. In total, 11,760 h of video footage (14 days \times 24 h per day \times 35 subjects) were collected for Experiment 2.

3.1.2. Input phase

During the input phase (first week of life), the chicks were reared in an environment that contained no objects other than a single virtual face. The face moved continuously, rotating through a 120° viewpoint range about a vertical axis passing through its centroid. Each chick was randomly assigned to either the slow, medium, or fast condition. In the slow condition, the face rotated back and forth in 20 s. In the medium condition, the face rotated back and forth in 5 s. In the fast condition, the face rotated back and forth in 1 s.

We chose these speed parameters based on prior controlled-rearing experiments with newborn chicks (Wood & Wood, 2016). In experiments exploring the development of view-invariant object recognition, we found that chicks created *view-invariant* representations when their imprinted object rotated slowly (one rotation every 15 s) and *view-specific* representations when their imprinted object rotated quickly (one rotation every second). When their imprinted object rotated at a medium speed (one rotation every 5 s), the chicks created mixed representations that contained both view-invariant and view-specific features. We used the same speed parameters here, except that we used an even slower rotation speed (20 s) because the slowest rotation speed (15 s) from Wood and Wood (2016) still produced some sensitivity to view-specific features. If our hypothesis is correct (i.e., that slowly changing visual experiences are necessary for the development of view-invariant recognition), then using a slower 20 s rotation speed should allow chicks to build view-invariant representations that do not contain any view-specific information.

We used the same faces as in Experiment 1. Seventeen chicks were imprinted to face A and 18 chicks were imprinted to face B (Figure 1b). Videos S3–S5 show sample animations presented during the input phase for the slow, medium, and fast conditions, respectively.

3.1.3. Test phase

During the test phase (second week of life), we probed the face representation built by each chick. The chicks were presented with two types of test trials (Figure 3).

On the viewpoint trials, one display wall showed the imprinted face rotating around the imprinted axis, which presented familiar views of the face (see Video S6). The other display wall showed the imprinted face rotating around a novel axis, which presented unfamiliar views of the face. For the novel axes, the imprinted face rotated around an axis tilted 45° or 90° from the imprinted axis. If the chicks built face representations that contained view-specific features, then they should have preferred the face rotating around the familiar axis compared to the unfamiliar axes.

On the identity trials, one display wall showed the imprinted face rotating around a novel axis (an axis tilted 45° or 90° from the imprinted axis, see Video S7). The other display wall showed an unfamiliar face rotating around the imprinted axis. This contrast increased the within-face image difference (i.e., the image-level difference between the test animation of the imprinted face and the input animation of the imprinted face) and decreased the between-face image difference (i.e., the image-level difference between the test animation of the unfamiliar face and the input animation of the imprinted face). In other words, the test animation of the unfamiliar face was *more similar* to the imprinting stimulus than the test animation of the imprinted face was to the imprinting stimulus. Because the unfamiliar face and the original imprinting stimulus both showed a face moving through the same viewpoint range, the facial features (e.g., eyes, nose, and mouth) in the animations moved through the same spatiotemporal trajectory. Thus, to recognize their imprinted face, the chicks needed to build invariant representations that were selective for the identity of the face and tolerant to viewpoint changes.

The test trials lasted 20 min and were separated from one another by 40-min rest periods. During the rest periods, the imprinted face appeared on one display wall and a white screen

Imprinting Viewpoint Range of Face A

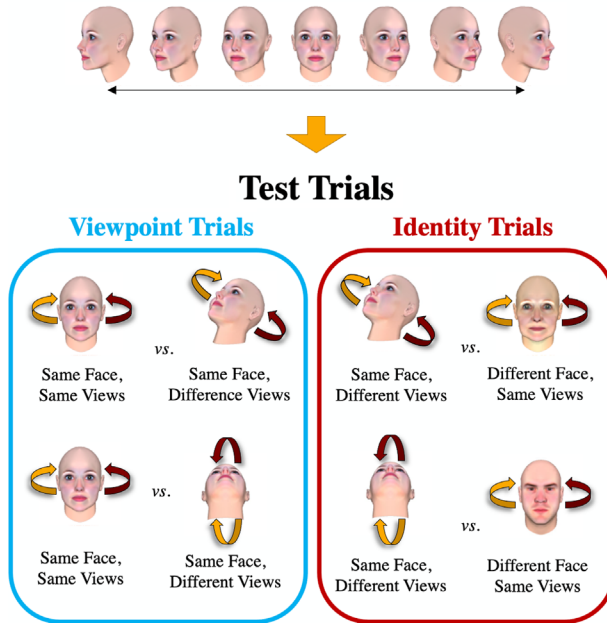


Fig. 3. Illustration of an imprinting animation and the associated test trials. Each chick was imprinted to a single virtual face, then presented with two types of test trials: viewpoint trials and identity trials. In the viewpoint trials (blue box), one display wall showed the imprinted face from the familiar viewpoint range and the other display wall showed the imprinted face from an unfamiliar viewpoint range. In the identity trials (red box), one display wall showed the imprinted face from an unfamiliar viewpoint range and the other display wall showed an unfamiliar face from the familiar viewpoint range. A view-invariant face representation should be selective for identity features but not view-specific features.

appeared on the other display wall. Each subject received 24 test trials per day. The conditions were presented in randomized blocks throughout the test phase.

3.2. Results

The results are shown in Figure 4a. Our main hypothesis was that the speed of the imprinted face would affect the information content (identity vs. viewpoint-specific features) of the face representation. To test this hypothesis, we performed a repeated-measures ANOVA with the within-subjects main effect of trial type (identity vs. viewpoint trials) and the between-subjects main effect of speed (slow, medium, or fast). The ANOVA revealed a significant main effect of trial type ($F(1, 32) = 24.85, p = .00002, \eta^2 = .44$) and a significant interaction between trial type and speed ($F(2, 32) = 22.72, p = .00000007, \eta^2 = .59$). The main effect of speed was not significant. The significant interaction was driven by higher performance on the identity trials in the slow speed condition and higher performance on the viewpoint trials in the fast speed condition (Figure 4a).

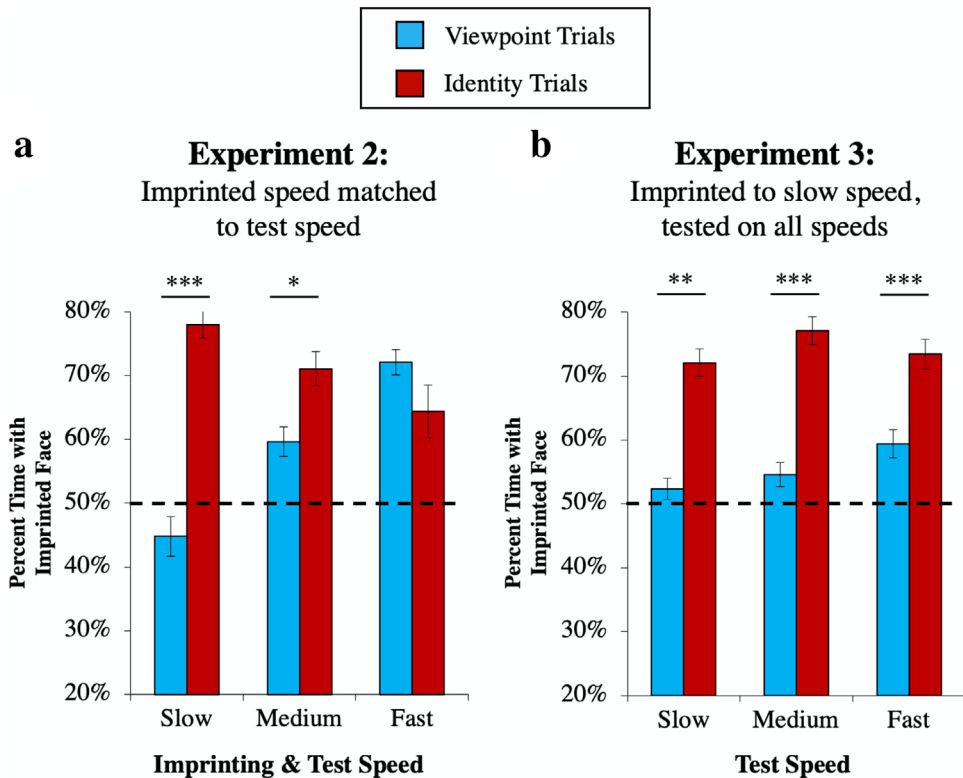


Fig. 4. (a) Results from Experiment 2. Newborn chicks were imprinted to a face that rotated at a slow, medium, or fast speed. Recognition performance varied as a function of the speed of the imprinted face. When the imprinted face moved slowly, the chicks built view-invariant face representations that were selective for identity features but not view-specific features. Conversely, as the imprinted face moved more quickly, the chicks built face representations that were less selective for identity features and more selective for view-specific features. (b) Results from Experiment 3. All of the chicks were imprinted to a slowly moving face, then tested with faces that moved at slow, medium, and fast speeds. Recognition performance did not vary as a function of the speed of the test faces, indicating that chicks can recognize quickly moving faces provided that the face moved slowly when being encoded into memory.

Performance on the identity trials was highest for the chicks reared with the slowly moving face. The chicks reared with a slowly moving face performed significantly better on the identity trials than the chicks reared with a medium speed face (independent samples t -test, $t(21) = 2.35$, $p = .03$, Cohen's $d = .99$) and the chicks reared with a fast moving face (independent samples t -test, $t(16.08) = 2.95$, $p = .009$, Cohen's $d = 1.21$). Conversely, the chicks reared with a fast moving face performed significantly better on the viewpoint trials than the chicks reared with a medium speed face (independent samples t -test, $t(22) = 4.40$, $p = .0002$, Cohen's $d = 1.80$) and the chicks reared with a slowly moving face (independent samples t -test, $t(21) = 7.52$, $p < 10^{-6}$, Cohen's $d = 3.11$).

When reared with a slowly moving face, the chicks built view-invariant face representations that were more sensitive to identity features than view-specific features (paired samples *t*-test, Mean Difference = 33%, SD = 16%, $t(10) = 7.00$, $p < 10^{-4}$, Cohen's $d = 2.11$). The chicks reared with a medium speed face also built representations that were more sensitive to identity features than view-specific features (paired samples *t*-tests, Mean Difference = 11%, SD = 14%, $t(11) = 2.72$, $p = .02$, Cohen's $d = .82$). However, the chicks reared with a medium speed face built representations that were significantly less sensitive to identity features (relative to viewpoint features) than the chicks reared with a slowly moving face (independent-samples *t*-test, $t(21) = 3.62$, $p = .002$, Cohen's $d = 1.51$).

Conversely, when reared with a fast moving face, the chicks built view-specific face representations that were not significantly more sensitive to identity features than viewpoint features (paired samples *t*-test, Mean Difference = -8%, SD = 15%, $t(11) = 1.82$, $p = .096$, Cohen's $d = -0.53$). The chicks reared with a fast moving face also built representations that were significantly less sensitive to identity features (relative to viewpoint features) than the chicks reared with a medium speed face (independent samples *t*-test, $t(22) = 3.33$, $p = .003$, Cohen's $d = 1.36$) and the chicks reared with a slowly moving face (independent samples *t*-test, $t(21) = 6.46$, $p < 10^{-5}$, Cohen's $d = 2.69$).

We also performed follow-up analyses to determine the conditions in which the chicks performed above chance level. The chicks performed significantly above chance level on the identity trials in the slow condition ($t(10) = 13.56$, $p = .0000001$, Cohen's $d = 4.09$), the medium condition ($t(11) = 7.15$, $p = .00002$, Cohen's $d = 2.06$), and the fast condition ($t(11) = 3.48$, $p = .005$, Cohen's $d = 1.01$). Conversely, the chicks performed significantly above chance level on the viewpoint trials in the fast condition ($t(11) = 11.19$, $p = .0000002$, Cohen's $d = 3.23$) and the medium condition ($t(11) = 3.58$, $p = .004$, Cohen's $d = 1.03$), but not in the slow condition ($t(10) = 1.66$, $p = .13$, Cohen's $d = -.50$).

Finally, we examined whether the chicks' face recognition performance changed across the 7 days of the test phase. Performance did not significantly increase or decrease across the test phase in any condition (slow, medium, and fast) on the viewpoint trials or identity trials (Pearson correlations between performance and test day; no analyses survive Holm-Bonferroni correction for six analyses, uncorrected *p*-values reported here; slow condition, viewpoint trials: $r = .015$, $p = .974$; slow condition, identity trials: $r = -.072$, $p = .878$; medium condition, viewpoint trials: $r = -.786$, $p = .036$; medium condition, identity trials: $r = -.087$, $p = .853$; fast condition, viewpoint trials: $r = -.602$, $p = .153$; fast condition, identity trials: $r = .346$, $p = .447$) As illustrated in Figure 5, performance was stable and robust: the chicks immediately achieved maximal levels of performance, and did not improve significantly across the test phase.

3.3. Discussion

In Experiment 2, we tested whether the development of view-invariant face recognition requires experience with a slowly changing face. We observed a strong relationship between the speed of face input and the resulting face recognition output. When the face rotated slowly, the chicks built view-invariant face representations that were selective for identity features and

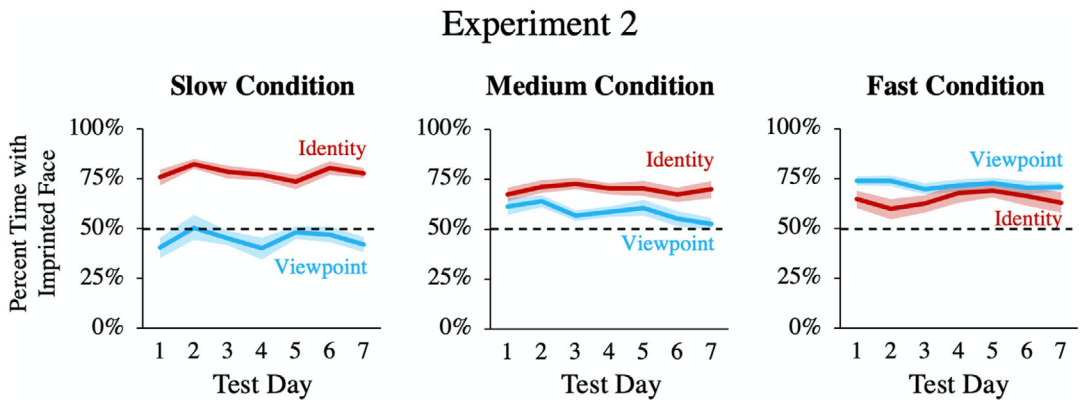


Fig. 5. Face recognition performance across the test phase in the slow, medium, and fast conditions (Experiment 2). Red lines show average performance on the identity trials. Blue lines show average performance on the viewpoint trials. The ribbons show ± 1 SE. Performance did not significantly increase across the test phase in any condition or trial type.

tolerant to changes in viewpoint. Conversely, when the face rotated quickly, the chicks built representations that were less selective for identity features and more selective for familiar viewpoints. Thus, there is a slowness constraint on the development of view-invariant face recognition.

Because the faces moved at the same speed during the input phase and test phase, it is possible that the impairment in recognizing quickly moving faces was caused by limitations in the chicks' ability to attend to or perceive quickly moving stimuli, rather than limitations in the chicks' ability to *encode* face information. We previously found that this alternative explanation cannot account for the slowness constraint in the development of basic-level object recognition (Wood & Wood, 2016, 2017). Newborn chicks can successfully recognize a quickly moving object, provided that the object moved slowly when being encoded into memory. In Experiment 3, we tested whether a similar pattern occurs in the development of view-invariant face recognition.

4. Experiment 3

4.1. Methods

Experiment 3 was identical to Experiment 2, with three key changes. First, a new group of 12 chicks were tested. Second, all of the chicks were imprinted to a slowly moving face in the input phase, with six chicks imprinted to face A and six chicks imprinted to face B. Third, rather than presenting test faces that rotated at the same speed as the input face, we presented test faces that rotated at slow, medium, and fast speeds on different test trials. Thus, each chick received six types of test trials: slow identity, slow viewpoint, medium identity, medium viewpoint, fast identity, and fast viewpoint. If experience with quickly moving faces

impairs newborn chicks' ability to *perceive* faces, then face recognition performance on the identity trials should be lower when the faces rotate quickly, and higher when the faces rotate slowly. Conversely, if experience with quickly moving faces impairs newborn chicks' ability to *encode* faces, then performance on the identity trials should be high regardless of whether the test faces rotate at slow, medium, or fast speeds (because the input face moved slowly when being encoded into memory).

4.2. Results

Results are shown in Figure 4b. An ANOVA with the within-subject factors of rotation speed (slow, medium, and fast) and trial type (identity and viewpoint) revealed significant main effects of trial type ($F(1, 11) = 90.85, p = .000001, \eta^2 = .89$) and rotation speed ($F(2, 22) = 4.07, p = .03, \eta^2 = .27$). The interaction of trial type and rotation speed was not significant ($F(2, 22) = 2.60, p = .10, \eta^2 = .19$).

For the main effect of trial type, performance was significantly higher on the identity trials than the viewpoint trials (paired t -test, $t(11) = 9.27, p = .000002$, Cohen's $d = 2.68$). This pattern was observed for all three rotation speeds (paired t -tests, all p s Holm–Bonferroni corrected; slow speed: $t(11) = 7.67, p = .00002$, Cohen's $d = 2.21$; medium speed: $t(11) = 8.02, p = .00002$, Cohen's $d = 2.31$; fast speed: $t(11) = 4.14, p = .002$, Cohen's $d = 1.20$). These results replicate the results from the slow condition in Experiment 2.

For the main effect of rotation speed, paired t -tests revealed that overall recognition performance was not impaired when the test faces rotated at faster speeds. In fact, overall recognition performance was significantly higher for the quickly rotating face than the slowly rotating face (paired t -test, $t(11) = 2.91, p = .01$, Cohen's $d = .84$).

4.3. Discussion

When newborn chicks were reared with a slowly moving face, they built view-invariant face representations that were highly sensitive to identity features, regardless of whether the test faces rotated at slow, medium, or fast speeds. Thus, experience with slowly moving faces improves chicks' ability to build face representations that are more sensitive to identity features than view-specific features. However, once a face representation has been built, the chicks' recognition performance is the same whether the face moves at slow, medium, or fast speeds. Therefore, the effect of rotation speed on face recognition performance observed in Experiment 2 cannot be explained by limitations in perception or attention. Rather, experience with slowly moving faces improves chicks' ability to encode view-invariant face representations.

5. General discussion

We used an automated controlled-rearing method to examine (1) whether newborn visual systems can perform view-invariant face recognition and (2) whether the development of this

ability requires experience with slowly moving faces. We found that newborn chicks can build view-invariant face representations, successfully recognizing their imprinted face across novel viewpoints and novel rotation speeds. Newborn chicks succeeded in this task despite never having seen any other faces or objects. We also found that the development of view-invariant face recognition is enhanced by experience with slowly moving faces. When chicks were reared with quickly moving faces, the chicks built distorted face representations that showed reduced selectivity for identity features and enhanced selectivity for view-specific features. Face recognition is, therefore, subject to the same slowness constraint that limits basic-level object recognition (Wood & Wood, 2016).

Our results indicate that the visual features that support invariant face recognition are learned when newborn brains experience slowly moving faces. In fact, we show that it is possible to causally manipulate the invariance (tolerance) of these features simply by changing the face's rotation speed when the face is being encoded into memory. Our study was not designed to isolate the specific visual features (e.g., color, shape, and configural cues) that drive face recognition in newborn chicks. Regardless of the specific nature of these features, our results support UTL theories, which propose that invariant features are learned by leveraging the spatiotemporal structure of natural (slowly changing) visual environments.

It is worth emphasizing that the chicks reared with the quickly moving faces still acquired visual experience with slowly moving features. For instance, the chicks acquired slowly changing visual experiences from the extended surfaces of the chamber. The chicks also acquired slowly changing visual experience from the grain and water during feeding and from their own body parts during preening and other forms of visual self-inspection. Nevertheless, when the *faces* in the chicks' visual environment did not move slowly over time, the chicks developed impaired face recognition. This finding suggests that the development of face recognition is facilitated by experience with slowly moving faces, rather than experience with slowly moving features more generally.

Ultimately, additional research is needed to fully characterize how motion speed impacts the development of face recognition during the first few months of life. While we found that experience with slowly moving faces facilitates the development of face recognition in the first few weeks of life, it is possible that increased exposure to quickly moving faces could lead to successful face recognition. Future experiments could test this possibility by rearing newborn chicks for several months with quickly moving faces and measuring whether face recognition improves over development.

5.1. *Why do fast visual experiences impair object and face recognition?*

Why do UTL models predict that fast visual experiences disrupt both face *and* object recognition in newborn animals? According to UTL models, invariant representations are learned from unsupervised adaptation to the spatiotemporal statistics of the natural visual world. Specifically, UTL models build invariant representations by associating retinal images that appear within the same temporal window, under the assumption that temporal adjacency is generally a good cue that two images are of the same face or object (Foldiak, 1991; Isik, Leibo, & Poggio, 2012; Li & DiCarlo, 2008; Mobahi, Collobert, & Weston, 2009; Rolls,

2012). Under this temporal association rule, the cells activated by one image will be linked with the cells activated by subsequent images. This linkage can be used to signal that the images represent the same object—despite its change in appearance. Temporal association is thought to occur through spike-timing dependent plasticity and the persistent firing of neurons. For example, following the brief (16 ms) presentation of a visual stimulus, neurons continue to fire for 100–400 ms (Rolls & Tovee, 1994; Rolls, Tovee, & Panzeri, 1999; Rolls, Tovee, Purcell, Stewart, & Azzopardi, 1994). There is also evidence that the prolonged (100 ms) binding period of N-methyl-D-aspartate channels may provide a temporal window for visual learning (Hestrin, Sah, & Nicoll, 1990).

Changing the speed of an object will change which images appear on the retina during the same temporal window, such that quickly moving objects activate larger swaths of photoreceptor cells than slowly moving objects. Thus, when UTL mechanisms encode quickly moving objects, larger swaths of cells become associated together in memory (compared to slowly moving objects). As a result, when newborn animals encode quickly moving objects, they build distorted representations with low-precision features that fail to generalize across new viewing conditions.

In addition to our *behavioral* findings, recent controlled-rearing experiments with rats provide convergent evidence that quickly changing visual experiences produce low-precision representations on the *neuronal* level (Matteucci & Zoccolan, 2020). When newborn rats were reared with frame-scrambled versions of natural movies—which preserved the natural spatial statistics but resulted in quickly changing, temporally unstructured input—the rats developed fewer complex cells (the cells which produce invariance in neural circuits) in primary visual cortex. These cells also showed abnormally fast response dynamics and were less likely to support stable decoding of stimulus orientation. Thus, depriving newborn brains of slowly changing visual experiences disrupts the normal development of the visual system.

5.2. *Common developmental constraints on face and object recognition*

To what extent do face recognition and object recognition share common computational foundations? Many researchers have argued that face and object recognition depend on separate, domain-specific systems from birth (Duchaine, Yovel, Butterworth, & Nakayama, 2004; Farah, 2004; Farah, Rabinowitz, Quinn, & Liu, 2000; McKone, Crookes, Jeffery, & Dilks, 2012). More recently, however, a growing body of evidence indicates that face and object recognition develop from common domain-general learning mechanisms, with domain-specific neural populations emerging in the cortex on the basis of visual experience. Support for this domain-general account comes from studies showing that (1) face recognition undergoes domain-specific development during the first decade of human life (Weigelt et al., 2014), (2) newborns' early emerging face preferences can be explained by domain-general computations operating over binocular input (Wilkinson, Paikan, Gredeback, Rea, & Metta, 2014), and (3) category-selective regions (including face regions) emerge in the cortex on the basis of early visual experiences (Arcaro et al., 2017; Roder, Ley, Shenoy, Kekunnaya, & Bottari, 2013; Srihasam, Mandeville, Morocz, Sullivan, & Livingstone, 2012). Our results

provide additional evidence that common computations underlie face and object recognition in newborn brains because both object and face recognition are subject to the same slowness constraint during early visual development.

Ultimately, a deep mechanistic understanding of the origins of face recognition will require (1) a formal characterization of the visual processing machinery built during prenatal development and (2) a formal characterization of how that machinery is shaped by experience. The present results should be valuable for this endeavor because they provide detailed descriptions of how specific visual inputs relate to specific face recognition outputs in a newborn model system. These input–output patterns can serve as benchmarks for evaluating computational models. For example, in order to successfully account for the development of face recognition in newborn chicks, a computational model should produce two patterns. First, the model should be capable of building view-invariant face representations from minimal training data (e.g., experience with a single face seen from a limited viewpoint range) using self-supervised learning rules. Second, the model should successfully develop view-invariant face representations when trained with slowly moving faces, but develop view-specific representations when trained with quickly moving faces.

5.3. Conclusion

We used an automated controlled-rearing method to explore the origins of face recognition. Our study supports two conclusions. First, view-invariant face recognition can emerge rapidly in newborn brains from minimal experience with faces. Second, the development of this ability is subject to a slowness constraint. Experience with slowly moving faces facilitates the development of invariant face recognition, consistent with UTL models in computational neuroscience. More generally, this study reveals how experience shapes the development of a foundational ability during the earliest stages of visual learning.

Notes

1. We distinguish face recognition from face detection. Face recognition involves identifying whether a specific face is familiar versus novel, whereas face detection involves identifying whether a stimulus is a face or not. The majority of studies on newborn face perception have focused on face detection rather than face recognition.
2. We computed the average of these two values because performance was higher on the test trials where the imprinted face stayed on the same display wall after the rest period compared to the test trials where the imprinted face switched display walls after the rest period. This is a natural consequence of tracking all of the chicks' behavior. For example, if the chick was sleeping when the experiment switched from a rest period to a test trial, then they would remain on the "correct" side of the chamber on some trials and the "incorrect" side of the chamber on other trials, without having made a choice between the faces. Computing the average of these two values corrected for this issue.

Acknowledgments

Funded by NSF CAREER Grant BCS-1351892 and a James S. McDonnell Foundation Understanding Human Cognition Scholar Award. This research was conducted at the University of Southern California.

References

- Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., & Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature Neuroscience*, *20*(10), 1404.
- Bogale, B. A., Aoyama, M., & Sugita, S. (2011). Categorical learning between 'male' and 'female' photographic human faces in jungle crows (*Corvus macrorhynchos*). *Behavioural Processes*, *86*(1), 109–118.
- Bulf, H., & Turati, C. (2010). The role of rigid motion in newborns' face recognition. *Visual Cognition*, *18*(4), 504–512.
- Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development*, *10*(1–2), 67–74.
- Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, *195*(4275), 312–314.
- de Haan, M., Johnson, M. H., Maurer, D., & Perrett, D. I. (2001). Recognition of individual faces and average face prototypes by 1- and 3-month-old infants. *Cognitive Development*, *16*(2), 659–678.
- De Heering, A., Turati, C., Rossion, B., Bulf, H., Goffaux, V., & Simion, F. (2008). Newborns' face recognition is based on spatial frequencies below 0.5 cycles per degree. *Cognition*, *106*(1), 444–454.
- DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341.
- Duchaine, B., Yovel, G., Butterworth, E., & Nakayama, K. (2004). Elimination of all domain-general hypotheses of prosopagnosia in a single individual: Evidence for an isolated deficit in 2nd order configural face processing. *Journal of Vision*, *4*(8), 214–214.
- Farah, M. J. (2004). *Visual agnosia*. MIT Press.
- Farah, M. J., Rabinowitz, C., Quinn, G. E., & Liu, G. T. (2000). Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology*, *17*(1–3), 117–123.
- Foldiak, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, *3*, 194–200.
- Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*(2), 160–170.
- Gauthier, I., Tarr, M. J., Moylan, J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2000). Does visual subordinate-level categorisation engage the functionally defined fusiform face area? *Cognitive Neuropsychology*, *17*(1–3), 143–164.
- Gibson, B. M., Wasserman, E. A., Gosselin, F., & Schyns, P. G. (2005). Applying bubbles to localize features that control pigeons' visual discrimination behavior. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(3), 376–382.
- Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, *56*(5), 872.
- Hestrin, S., Sah, P., & Nicoll, R. A. (1990). Mechanisms generating the time course of dual component excitatory synaptic currents recorded in hippocampal slices. *Neuron*, *5*(3), 247–253.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*(1), 106–154.
- Isik, L., Leibo, J. Z., & Poggio, T. (2012). Learning and disrupting invariance in visual recognition with a temporal association rule. *Frontiers in Computational Neuroscience*, *6*, 37.
- Jayaraman, S., & Smith, L. B. (2019). Faces in early visual environments are persistent not just frequent. *Vision Research*, *157*, 213–221.

- Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, *107*(25), 11163–11170.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L. Z., & Pascalis, O. (2007). The other-race effect develops during infancy — Evidence of perceptual narrowing. *Psychological Science*, *18*(12), 1084–1089.
- Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, *321*(5895), 1502–1507.
- Masquelier, T., & Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology*, *3*(2), e31.
- Matteucci, G., & Zoccolan, D. (2020). Unsupervised experience with temporal continuity of the visual environment is causally involved in the development of V1 complex cells. *Science Advances*, *6*(22), eaba3742.
- McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*(5), 605–610.
- 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.
- Mobahi, H., Collobert, R., & Weston, J. (2009). Deep learning from temporal coherence in video. Paper presented at the Proceedings of the 26th Annual International Conference on Machine Learning.
- Pascalis, O., Deschonen, S., Morton, J., Deruelle, C., & Fabregrenet, M. (1995). Mothers face recognition by neonates — A replication and an extension. *Infant Behavior & Development*, *18*(1), 79–85.
- Roder, B., Ley, P., Shenoy, B. H., Kekunnaya, R., & Bottari, D. (2013). Sensitive periods for the functional specialization of the neural system for human face processing. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(42), 16760–16765.
- Rolls, E. T. (2012). Invariant visual object and face recognition: Neural and computational bases, and a model, VisNet. *Frontiers in Computational Neuroscience*, *6*, 35.
- Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral-cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *257*(1348), 9–15.
- Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *Journal of Cognitive Neuroscience*, *11*(3), 300–311.
- Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., & Azzopardi, P. (1994). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Experimental Brain Research*, *101*(3), 473–484.
- Srihasam, K., Mandeville, J. B., Morocz, I. A., Sullivan, K. J., & Livingstone, M. S. (2012). Behavioral and anatomical consequences of early versus late symbol training in macaques. *Neuron*, *73*(3), 608–619.
- Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(1), 394–398.
- Troje, N. F., Huber, L., Loidolt, M., Aust, U., & Fieder, M. (1999). Categorical learning in pigeons: The role of texture and shape in complex static stimuli. *Vision Research*, *39*(2), 353–366.
- Turati, C., Macchi Cassia, V., Simion, F., & Leo, I. (2006). Newborns' face recognition: Role of inner and outer facial features. *Child Development*, *77*(2), 297–311.
- Walk, R., Gibson, E. J., & Tighe, T. (1957). Behavior of light-and dark-reared rats on a visual cliff. *Science*, *126*, 80–81.
- Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(8), 4800–4804.
- Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology*, *51*(2), 167–194.
- 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.
- Wilkinson, N., Paikan, A., Gredeback, G., Rea, F., & Metta, G. (2014). Staring us in the face? An embodied theory of innate face preference. *Developmental Science*, *17*(6), 809–825.

- Wiskott, L., & Sejnowski, T. J. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation*, 14(4), 715–770.
- Wood, J. N. (2013). Newborn chickens generate invariant object representations at the onset of visual object experience. *Proceedings of the National Academy of Sciences of the United States of America*, 110(34), 14000–14005.
- Wood, J. N. (2014). Newly hatched chicks solve the visual binding problem. *Psychological Science*, 25(7), 1475–1481.
- Wood, J. N. (2015). Characterizing the information content of a newly hatched chick's first visual object representation. *Developmental Science*, 18(2), 194–205.
- Wood, J. N. (2016). A smoothness constraint on the development of object recognition. *Cognition*, 153, 140–145.
- Wood, J. N., Prasad, A., Goldman, J. G., & Wood, S. M. W. (2016). Enhanced learning of natural visual sequences in newborn chicks. *Animal Cognition*, 19(4), 835–845.
- Wood, J. N., & Wood, S. M. W. (2016). The development of newborn object recognition in fast and slow visual worlds. *Proceedings of the Royal Society B-Biological Sciences*, 283(1829), 20160166.
- Wood, J. N., & Wood, S. M. W. (2017). Measuring the speed of newborn object recognition in controlled visual worlds. *Developmental Science*, 20(4), e12470.
- Wood, J. N., & Wood, S. M. W. (2018). The development of invariant object recognition requires visual experience with temporally smooth objects. *Cognitive Science*, 42(4), 1391–1406.
- Wood, S. M. W., & Wood, J. N. (2015a). A chicken model for studying the emergence of invariant object recognition. *Frontiers in Neural Circuits*, 9, 7.
- Wood, S. M. W., & Wood, J. N. (2015b). Face recognition in newly hatched chicks at the onset of vision. *Journal of Experimental Psychology: Animal Learning and Cognition*, 41(2), 206.
- Wood, S. M. W., & Wood, J. N. (2019). Using automation to combat the replication crisis: A case study from controlled-rearing studies of newborn chicks. *Infant Behavior and Development*, 57, 101329.
- Wood, S. M., & Wood, J. N. (2021). One-shot object parsing in newborn chicks. *Journal of Experimental Psychology: General*, <https://psycnet.apa.org/doi/10.1037/xge0001043>
- Wyss, R., Konig, P., & Verschure, P. F. M. J. (2006). A model of the ventral visual system based on temporal stability and local memory. *PLoS Biology*, 4(5), 836–843.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting information