

Cognitive Science 42 (2018) 1391–1406
Copyright © 2018 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12595

The Development of Invariant Object Recognition Requires Visual Experience With Temporally Smooth Objects

Justin N. Wood, Samantha M. W. Wood

Department of Psychology, University of Southern California

Received 29 January 2017; received in revised form 10 January 2018; accepted 15 January 2018

Abstract

How do newborns learn to recognize objects? According to temporal learning models in computational neuroscience, the brain constructs object representations by extracting smoothly changing features from the environment. To date, however, it is unknown whether newborns depend on smoothly changing features to build invariant object representations. Here, we used an automated controlled-rearing method to examine whether visual experience with smoothly changing features facilitates the development of view-invariant object recognition in a newborn animal—the domestic chick (*Gallus gallus*). When newborn chicks were reared with a virtual object that moved smoothly over time, the chicks created view-invariant representations that were selective for object identity and tolerant to viewpoint changes. Conversely, when newborn chicks were reared with a temporally non-smooth object, the chicks developed less selectivity for identity features and less tolerance to viewpoint changes. These results provide evidence for a “smoothness constraint” on the development of invariant object recognition and indicate that newborns leverage the temporal smoothness of natural visual environments to build abstract mental models of objects.

Keywords: Controlled rearing; Newborn; Object recognition; Smoothness; *Gallus gallus*

1. Introduction

To perceive the world successfully, newborns must build abstract object representations that generalize far beyond the input coming in through the retina. This ability is known as “invariant object recognition.” Invariant object representations are selective for object identity and tolerant to identity-preserving image changes (e.g., changes in viewpoint, size, position, and illumination). Mature visual systems can build invariant object

Correspondence should be sent to Justin N. Wood, Department of Psychology, University of Southern California, 3620 South McClintock Ave., Los Angeles, CA 90089. E-mail: justin.wood@usc.edu

representations effortlessly. For example, after just a brief glimpse of an object, human adults build invariant representations that generalize across tremendous variation in the retinal images produced by the object (Biederman & Bar, 1999; DiCarlo, Zoccolan, & Rust, 2012; Peissig & Tarr, 2007). To date, however, the development of this ability is poorly understood. How do newborns learn to transform raw retinal inputs into abstract (invariant) object representations?

According to unsupervised temporal learning models in computational neuroscience, the brain builds invariant object representations by extracting smoothly changing features from the visual environment (DiCarlo et al., 2012; Feldman & Tremoulet, 2006; Földiák, 1991; Rolls, 2012; Stone, 1996; Wiskott & Sejnowski, 2002). The key assumption underlying these models is that perceptually salient visual parameters (e.g., curvature, depth, surface orientation, and texture) vary smoothly over time in natural visual environments. Thus, newborn brains could build up invariance to these visual parameters by encoding statistical regularities across temporally smooth changes in the environment. This process is thought to be accomplished by unsupervised temporal learning mechanisms that associate patterns of neuronal activity produced by successive retinal images of an object.

A number of behavioral studies provide evidence that human adults use temporal association mechanisms to build object representations (Cox, Meier, Oertelt, & DiCarlo, 2005; Liu, 2007; Wallis, Backus, Langer, Huebner, & Bulthoff, 2009; Wallis & Bulthoff, 2001). Temporal association effects have also been found on the neurophysiological level in adult monkeys (Li & DiCarlo, 2008; Meyer & Olson, 2011; Miyashita, 1988). In this study, we tested whether newborn animals leverage the temporal smoothness of natural visual objects to build invariant object representations. Importantly, we examined the *first* visual object representation built by newborn animals, before their visual systems had been shaped by any prior visual experience with objects. Thus, our experiment provided a direct test of the initial state of object recognition machinery (i.e., machinery that has not been shaped by postnatal visual experience).

1.1. Using automated controlled rearing to explore the origins of object recognition

To assess the role of temporal smoothness in the development of invariant object recognition, we used an automated controlled-rearing method with a newborn animal model—the domestic chick. Unlike newborn humans, newborn chicks can be raised in strictly controlled environments from the onset of vision. Consequently, it is possible to systematically manipulate the visual experiences provided to newborn chicks and measure the effects of those manipulations on the development of object recognition (Wood & Wood, 2015). Moreover, fueled by innovation in image-based tracking technology, controlled-rearing experiments can now be fully automated (Wood, 2013). Automation makes it possible to collect large amounts of precise behavioral data from each subject, while eliminating the risk of experimenter error and bias during stimuli presentation and data collection.

Previous automated controlled-rearing studies have shown that newborn chicks develop object recognition rapidly. For example, newborn chicks can begin binding color and shape features into integrated object representations at the onset of vision (Wood, 2014).

Thus, newborn chicks can solve the “visual binding problem.” Newborn chicks can also solve the “invariance problem,” building object representations that generalize across novel viewing situations. For example, in Wood (2013), newborn chicks were reared with a single virtual object rotating through a 60° viewpoint range. Despite seeing the object from a limited range of views, the chicks were able to recognize the object across a wide range of novel viewpoints. Moreover, after learning about an object, chicks can recognize that object rapidly, within a fraction of a second (125 ms; Wood & Wood, 2016a). Thus, newborn chicks show many of the signatures of adult object recognition, including color-shape binding, view-invariant recognition, and rapid recognition. These studies provide an existence proof that newborn brains can develop advanced object recognition abilities within the first few days of life.

Automated controlled-rearing studies have also begun to reveal the role of visual experience in the development of object recognition. By systematically manipulating the visual experiences provided to newborn chicks, it has been possible to characterize the experiential ingredients necessary for the development of object recognition. First, there is evidence for a “slowness constraint” on the development of invariant object recognition (Wood & Wood, 2016b). Newborn chicks need experience with slowly moving objects in order to build view-invariant and speed-invariant object representations (Wood & Wood, 2016b). When chicks are reared with objects that move too quickly, they develop inaccurate object concepts that are distorted in the direction of object motion. Second, there is evidence for a “smoothness constraint” on the development of color recognition, shape recognition, and color-shape binding (Wood, 2016; Wood, Prasad, Goldman, & Wood, 2016). Newborn chicks need experience with smoothly moving objects in order to build accurate color and shape representations. Critically, however, it is unknown whether the development of invariant object recognition is also subject to a smoothness constraint. This study addresses this issue by testing whether newborn chicks need experience with smoothly moving objects to develop view-invariant object recognition.

In the first week of life (input phase), newborn chicks were reared in controlled-rearing chambers (Fig. 1) that contained no objects other than a single virtual object. For one group of chicks, the virtual object moved smoothly over time (Temporally Smooth Condition), whereas for another group of chicks, the virtual object moved non-smoothly over time (Temporally Non-Smooth Condition). In the second week of life (test phase), we used an automated two-alternative forced-choice procedure to measure each chick’s sensitivity to viewpoint features and identity features (Wood, 2015).

We measured sensitivity to these two feature types because building an invariant object representation requires transforming patterns of retinal activity (viewpoint features) into an abstract representation that is tolerant to retinal image variation and selective for a particular object (identity features). Measuring sensitivity to viewpoint features and identity features therefore reveals how successful the chick has been in building an invariant representation. If the chick has high sensitivity to identity features and low sensitivity to viewpoint features, then the chick successfully built an invariant object representation (i.e., a representation that is selective for the object’s identity and tolerant to viewpoint changes). Conversely, if the chick has high sensitivity to viewpoint features and low

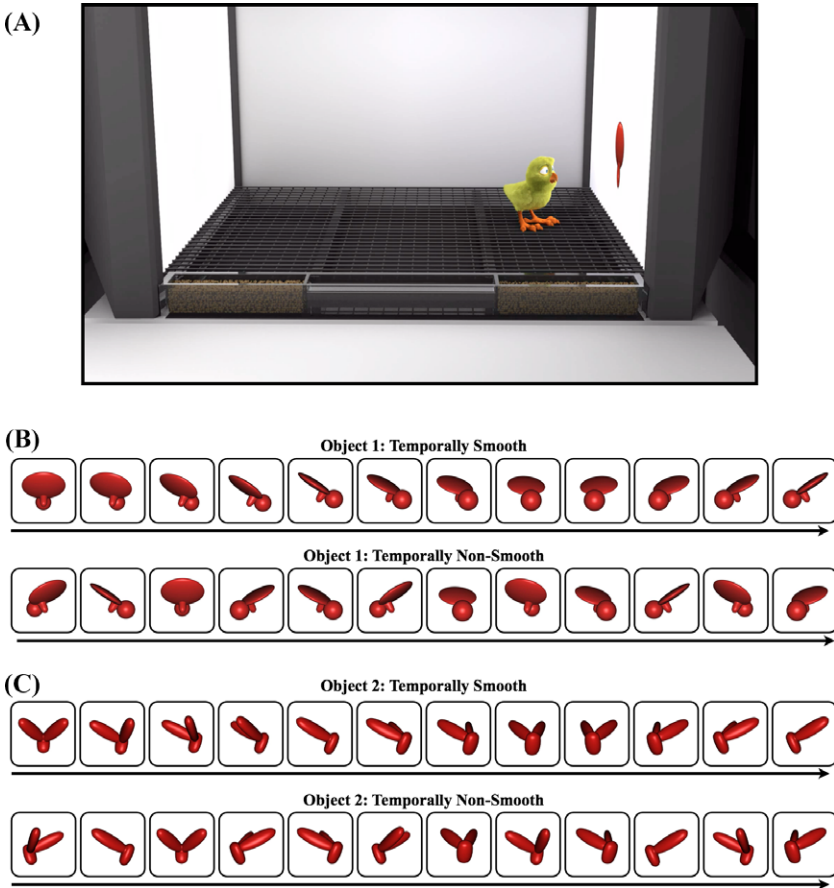


Fig. 1. (A) Illustration of a controlled-rearing chamber. The chambers contained no real-world objects. To present object stimuli to the chicks, virtual objects were projected on two display walls situated on opposite sides of the chamber. During the input phase (first week of life), newborn chicks were raised with a single virtual object moving smoothly or non-smoothly over time. Half of the chicks were raised with the object shown in panel (B) and half of the chicks were raised with the object shown in panel (C). The figures show successive images of the object presented to the chick.

sensitivity to identity features, then the chick built a view-based representation (i.e., a representation that is selective for an object's viewpoint).

2. Method

2.1. Subjects

Forty-one Rhode Island Red domestic chicks (*Gallus gallus*) of unknown sex were tested. No subjects were excluded from the analyses. The sample size was determined

before the experiments were conducted based on previous automated controlled-rearing experiments with newborn chicks (Wood, 2013, 2014). The chicks were randomly assigned to the conditions. The eggs were obtained from a local distributor and incubated in darkness in an OVA-Easy incubator (Brinsea Products Inc., Titusville, FL). The incubation room was kept in complete darkness to avoid exposing the chicks to light through their shells. After hatching, the chicks were moved from the incubation room to the controlled-rearing chambers in darkness with the aid of night vision goggles. Each chick was reared singly within its own chamber. This research was approved by The University of Southern California Institutional Animal Care and Use Committee.

2.2. Procedure

Newborn chicks were reared for 2 weeks within specially designed controlled-rearing chambers. The chambers measured 66 cm (length) \times 42 cm (width) \times 69 cm (height) and contained no real-world (solid, movable) objects. To present object stimuli to the chicks, we projected virtual 3D objects on two display walls (19" LCD monitors with 1,440 \times 900 pixel resolution) situated on opposite sides of the chamber. Food and water were available within transparent troughs in the ground that measured 66 cm (length) \times 2.5 cm (width) \times 2.7 cm (height). Grain was used as food because a heap of grain does not behave like an object (i.e., a heap of grain does not maintain a rigid, bounded shape). The floors of the chambers consisted of black wire mesh supported over a black matte surface.

The chambers recorded all of the chicks' behavior (9 samples/second, 24 hours/day, 7 days/week) via micro-cameras in the ceilings and automated image-based tracking software (EthoVision XT, Noldus Information Technology, Leesburg, VA). This automated data collection approach allowed large numbers of test trials (up to 168 trials) to be collected from each chick. As a result, it was possible to measure each chick's first visual object representation with high precision. In total, \sim 13,776 h of video footage (14 days \times 24 hours/day \times 41 subjects) were collected for this experiment.

In the first week of life (input phase), the chicks were reared in controlled-rearing chambers that contained no objects other than a single virtual object (Fig. 1). The object appeared on one display wall at a time, switching walls every 2 hours (Fig. 2A, Movies S1–S4). On average, the object measured 9 cm (length) \times 7 cm (height) and was displayed on a uniform white background. Half of the chicks were imprinted to the object shown in Fig. 1B, and half of the chicks were imprinted to the object shown in Fig. 1C. The virtual objects were modeled after those used in previous studies that tested for invariant object recognition in adult rats (Zoccolan, Oertelt, DiCarlo, & Cox, 2009) and newborn chicks (Wood, 2013).

We experimentally manipulated both the smoothness of the object's motion and the presentation rate of the successive images. In Wood (2016), there was a significant impairment in color and shape recognition when chicks were reared with objects that moved non-smoothly over time. However, since the temporally smooth objects were presented at a rate of 30 frame/s and the temporally non-smooth objects were presented at a

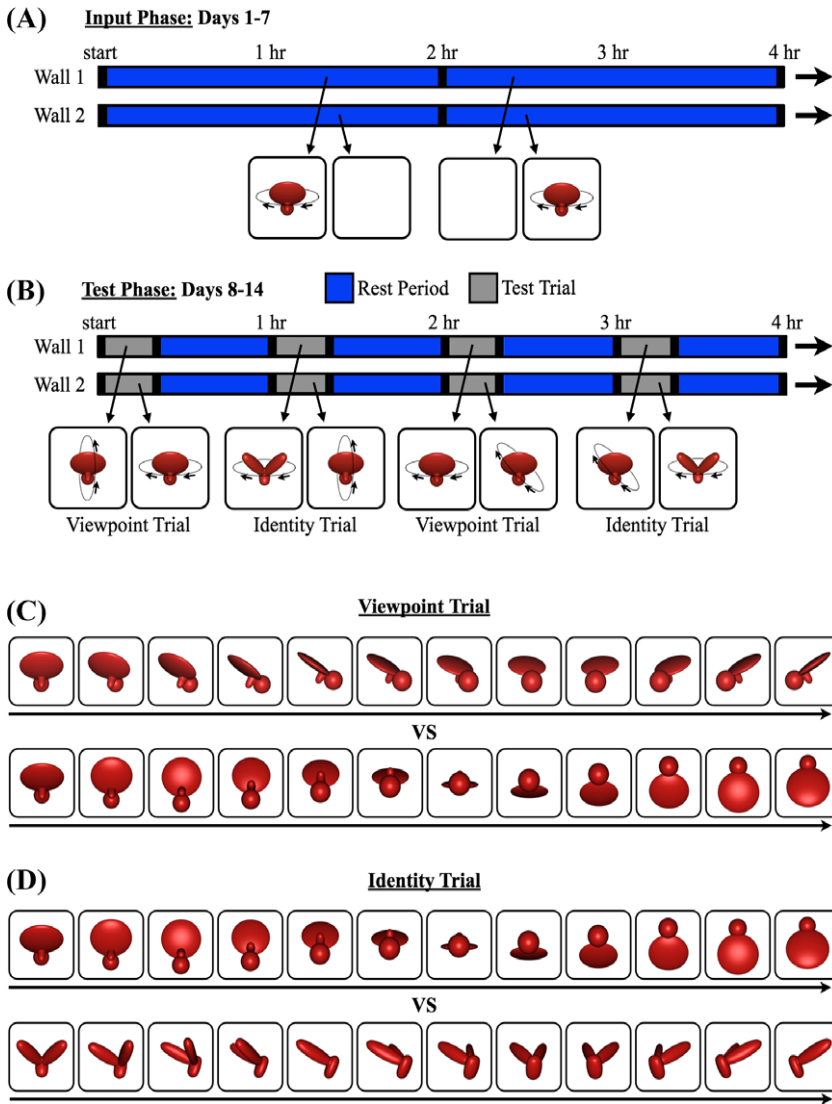


Fig. 2. The experimental procedure. These schematics illustrate how the virtual stimuli were presented for sample 4-h periods during the (A) input phase and (B) test phase. During the input phase, newborn chicks were reared with a single virtual object. The object appeared on one wall at a time (indicated by blue segments on the timeline), switching walls every 2 h, after a 1-min period of darkness (black segments). During the test phase, two object animations were shown simultaneously, one on each display wall, for 20 min per hour (gray segments). The illustrations below the timeline are examples of paired test objects displayed in four of the test trials. Each test trial was followed by a 40-min rest period (blue segments). During the rest periods, the animation from the input phase was shown on one display wall, and the other display wall was blank. These illustrations show the displays seen by the subjects raised with Imprinted Object 1 (Fig. 1). (C) Illustrations of a Viewpoint Trial and (D) an Identity Trial in the Temporally Smooth Condition.

rate of 1 frame/s, the factors of smoothness and rate of image change were confounded with one another. Thus, it is not clear whether the decrease in recognition performance was due to the lack of smoothness of the successive images or the slower presentation rate of the images (or a combination of these factors). To disentangle these factors, this study used a 2×2 design to manipulate the smoothness of the object motion (Temporally Smooth vs. Temporally Non-Smooth) and the presentation rate of the images (1 frame/s vs. 10 frame/s). See SI Movies S1–S4 for sample animations (all SI movies online: http://buildingamind.com/temporal_smoothness.html).

In the Temporally Smooth Condition, the virtual object rotated smoothly around a frontoparallel vertical axis, completing a 360° rotation in the depth plane every 15 s. The object rotated either 2.4° with each successive frame (10 frame/s presentation rate, Movie S1) or 24° with each successive frame (1 frame/s presentation rate, Movie S2).

In the Temporally Non-Smooth Condition, the chicks were shown the same virtual object, but the images of the object were presented in a scrambled order (Fig. 1B and C). Specifically, we took the 150 unique images (10 frame/s \times 15 s, from the 10 frame/s Temporally Smooth Condition) and randomized their order. On average, the object rotated 131.2° across successive images and the minimum difference between two successive images was 96° . As in the Temporally Smooth Condition, the successive images were presented at a rate of 10 frame/s (Movie S3) or 1 frame/s (Movie S4).

In the second week of life (test phase), we measured each chick's sensitivity to viewpoint features and identity features. On the Viewpoint Trials (Movies S5–S6), one display wall showed the imprinted object rotating around the familiar axis, whereas the other display wall showed the imprinted object rotating around a novel axis (i.e., the axis was tilted 45° or 90° ; Fig. 2C). If the chicks built view-based object representations, then they should have preferred the object rotating around the familiar axis (which presented familiar views of the object) compared to the novel axis (which presented novel views of the object).

On the Identity Trials (Movies S7–S8), one display wall showed the imprinted object rotating around a novel axis (i.e., the axis was tilted 45° or 90°), whereas the other display wall showed a novel object rotating around the familiar axis (Fig. 2D). Importantly, for this pair of test animations, the imprinted object and the unfamiliar object were equally different from the imprinting stimulus from both low-level retina-like (pixel-level) and V1-like perspectives (for details see Wood & Wood, 2016b). Thus, to recognize their imprinted object, the chicks needed to build high-level invariant representations that were selective for object identity and tolerant to identity-preserving image transformations (i.e., changes in viewpoint).

During the test phase, the objects moved smoothly in the Temporally Smooth Condition and non-smoothly in the Temporally Non-Smooth Condition. Additionally, the presentation rate of the objects in the test phase (10 frame/s vs. 1 frame/s) was the same as the presentation rate of the object in the input phase. The chicks received 168 test trials (1 trial per hour). Each test trial lasted 20 min and was followed by a 40-min rest period. During the rest periods, the animation from the input phase was shown on one display

wall and the other display wall showed a white screen. Fig. 2 illustrates the testing schedule and the trial types.

3. Results

The results are shown in Fig. 3. For the Identity Trials, the correct animation showed the imprinted object, while the incorrect animation showed the unfamiliar object. For the Viewpoint Trials, the correct animation showed the familiar viewpoint range, while the incorrect animation showed the unfamiliar viewpoint range. To analyze the data, we first computed the proportion of time each chick spent with the correct animation compared to

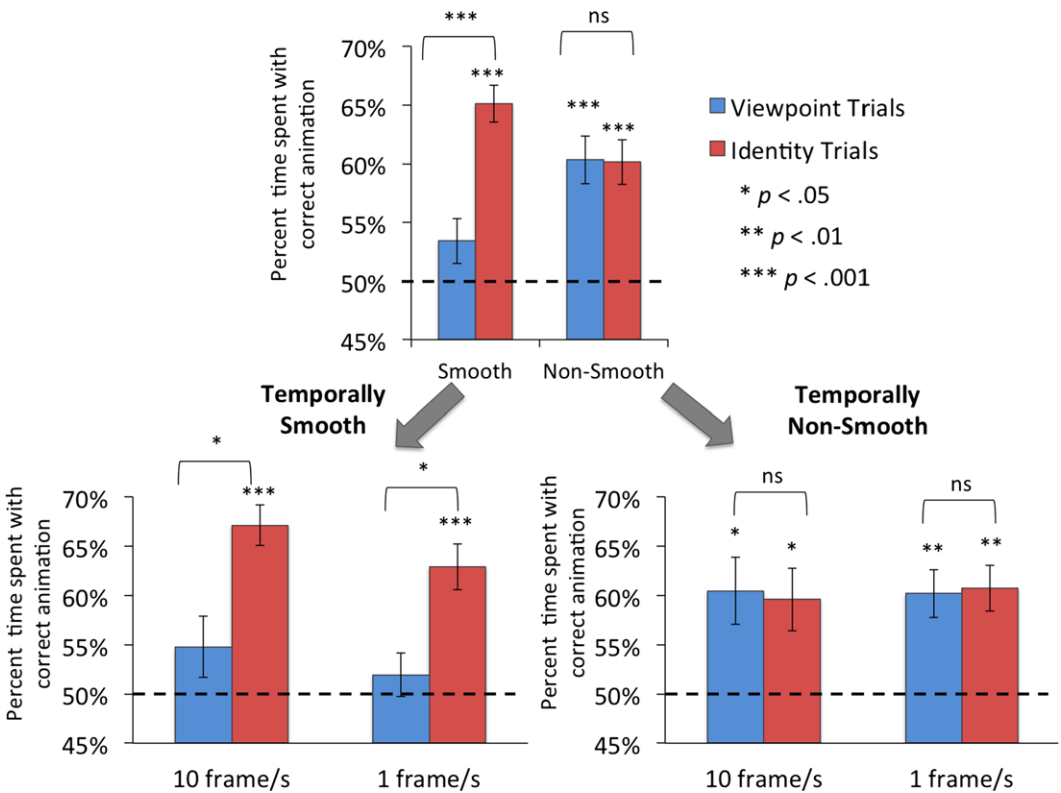


Fig. 3. Results from the test phase. The top graph shows the mean proportion of time the chicks spent by the familiar object on the Identity Trials and the familiar viewpoint range on the Viewpoint Trials. The bottom graphs show performance by Presentation Rate in the Temporally Smooth Condition (left) and Temporally Non-Smooth Condition (right). Chance performance was 50%. Error bars denote $\pm 1 SE$. When newborn chicks were reared with a temporally smooth object, the chicks built invariant representations that were selective for object identity (high performance on Identity Trials) and tolerant to viewpoint changes (low performance on Viewpoint Trials). Conversely, when newborn chicks were reared with a temporally non-smooth object, the chicks' object representations were both less selective for object identity and less tolerant to viewpoint changes.

the incorrect animation for the trials in which the imprinted object switched display walls after the rest period (switch trials) and for the trials in which the imprinted object stayed on the same display wall after the rest period (stay trials).¹ We then computed the average of these two values for the Identity Trials and Viewpoint Trials to obtain a performance score for each chick in each trial type.

A repeated measures ANOVA with the within-subjects factor of Trial Type (Identity vs. Viewpoint Trials) and between-subjects factor of Condition (Temporally Smooth vs. Temporally Non-Smooth) and Presentation Rate (10 frame/s vs. 1 frame/s) revealed a significant main effect of Trial Type ($F(1,37) = 9.560, p = .004, \eta_p^2 = .205$) and a significant interaction between Condition and Trial Type ($F(1,37) = 10.138, p = .003, \eta_p^2 = .215$). No other main effects or interactions were significant (all $ps > .30$).

Post hoc paired-samples t tests showed that performance on the Identity Trials was significantly higher than performance on the Viewpoint Trials for the chicks in the Temporally Smooth Condition ($t(20) = 4.289, p = .0004, d = 0.936$), but not for the chicks in the Temporally Non-Smooth Condition ($t(19) = 0.072, p = .943, d = 0.016$). Thus, when chicks were reared with a temporally smooth object, they built object representations showing the signatures of view-invariance (greater sensitivity to identity features and less sensitivity to familiar viewpoint features), while chicks reared with a temporally non-smooth object did not show these signatures of view-invariant recognition.

On the Viewpoint Trials, performance was significantly higher in the Temporally Non-Smooth Condition than the Temporally Smooth Condition ($t(39) = 2.47, p = .018, d = 0.771$). Thus, the chicks reared with the temporally non-smooth object were more likely to build representations that contained view-based features. Conversely, on the Identity Trials, performance was marginally higher in the Temporally Smooth Condition than the Temporally Non-Smooth Condition ($t(39) = 2.011, p = .051, d = 0.627$). Together, these results indicate that newborn chicks achieve greater levels of view-invariance when reared with temporally smooth objects (i.e., the representations are both more selective for object identity and more tolerant to viewpoint changes).

3.1. Change in performance over time

The chicks' recognition performance across the test phase is shown in Fig. 4. To test whether performance changed across the test phase, we performed a repeated measures ANOVA with the within-subjects factors of Trial Type (Identity vs. Viewpoint Trials) and Test Day (Days 1–7) and the between-subjects factors of Condition (Temporally Smooth vs. Temporally Non-Smooth) and Presentation Rate (10 frame/s vs. 1 frame/s). As in the analysis above, the main effect of Trial Type was significant ($F(1,37) = 7.365, p = .010, \eta_p^2 = .166$) as was the interaction between Trial Type and Condition ($F(1,37) = 10.134, p = .003, \eta_p^2 = .215$). In addition, the ANOVA revealed a significant main effect of Test Day ($F(6,222) = 2.251, p = .040, \eta_p^2 = .057$). Notably, recognition performance was significantly above chance levels across all days of the test phase (one-sample t tests collapsing across all conditions and trial types, all $ps < 10^{-7}$ with Holm-Bonferroni correction for multiple comparisons).

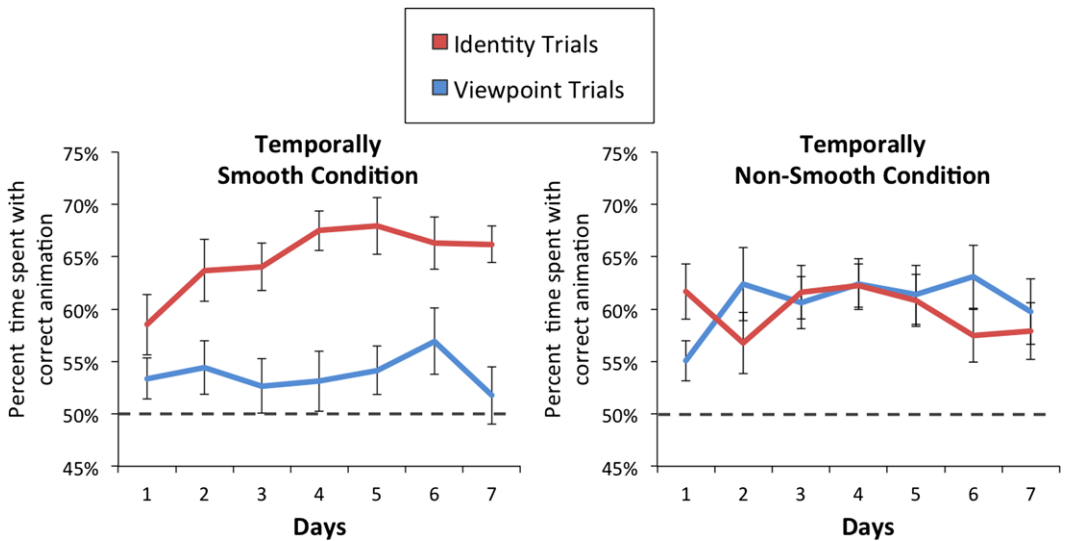


Fig. 4. Change in recognition performance across the test phase. The graphs show the mean performance on the Identity and Viewpoint Trials for each day of the test phase. Performance was higher on the Identity Trials than the Viewpoint Trials for each day of the test phase in the Temporally Smooth Condition (*left*), but not in the Temporally Non-Smooth Condition (*right*). Chance performance was 50%. Error bars denote ± 1 SE.

3.2. Individual subject performance

To test for the presence of individual differences across subjects, we tested whether subject identity was a significant predictor of performance on both the Viewpoint Trials and the Identity Trials. One-way ANOVAS showed that the identity of the subject was a significant predictor of performance on the Viewpoint Trials (Temporally Smooth Condition (10 frame/s): $F(10,895) = 13.686$, $p < 10^{-21}$, $\eta_p^2 = .133$; Temporally Smooth Condition (1 frame/s): $F(9,830) = 3.418$, $p < 10^{-3}$, $\eta_p^2 = .036$; Temporally Non-Smooth Condition (10 frame/s): $F(9,830) = 13.159$, $p < 10^{-19}$, $\eta_p^2 = .125$; Temporally Non-Smooth Condition (1 frame/s): $F(9,830) = 7.337$, $p < 10^{-9}$, $\eta_p^2 = .074$), and the Identity Trials (Temporally Smooth Condition (10 frame/s): $F(10,889) = 4.844$, $p < 10^{-6}$, $\eta_p^2 = .052$; Temporally Smooth Condition (1 frame/s): $F(9,815) = 4.181$, $p < 10^{-4}$, $\eta_p^2 = .044$; Temporally Non-Smooth Condition (10 frame/s): $F(9,820) = 8.465$, $p < 10^{-11}$, $\eta_p^2 = .085$; Temporally Non-Smooth Condition (1 frame/s): $F(9,829) = 4.153$, $p < 10^{-4}$, $\eta_p^2 = .043$). Thus, there were significant individual differences across the subjects.

To visualize this pattern of variation, we charted the position of each chick's object representation within a two-dimensional abstraction space (Fig. 5), using the method reported in Wood and Wood (2016b). Each point in the space depicts the object representation built by one chick (with sensitivity to identity features and viewpoint features charted on the x - and y -axis, respectively). Thus, the position of the representation within the abstraction space reflects its degree of abstraction (i.e., the representation's sensitivity to identity features and tolerance to viewpoint changes). A representation that is fully

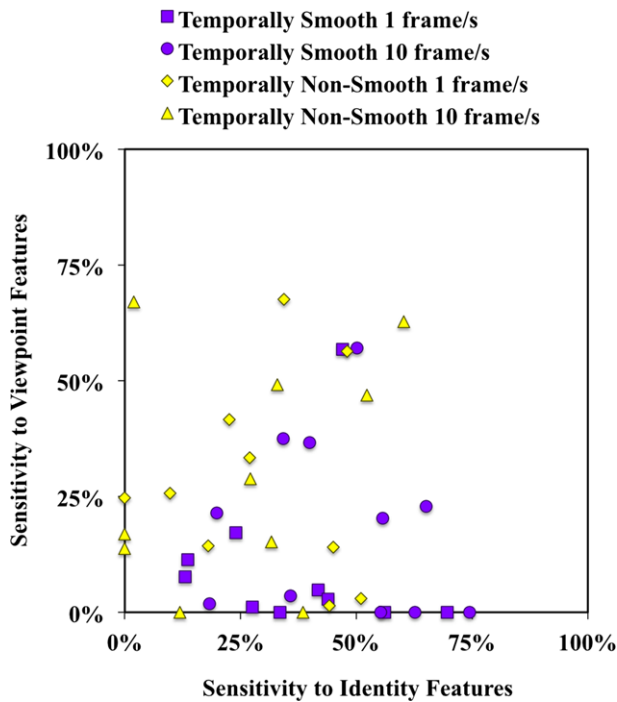


Fig. 5. The abstraction space for visualizing the patterns of variation across chicks. Each dot depicts the object representation built by one chick. The x - and y -axes show the representation's sensitivity to identity features and viewpoint features, respectively. The axes range from 0% (chance performance) to 100% (perfect performance). Perfect performance was equal to each chick's performance on the rest periods, which measured the amount of time the chicks generally preferred to spend in proximity to their imprinted object. The position of the representations in the abstraction space varied as a function of the temporal smoothness of the objects, with chicks building more abstract (invariant) representations when raised with temporally smooth objects.

view-dependent would fall in the upper left corner (high sensitivity to viewpoint features and low sensitivity to identity features), while a representation that is fully view-invariant would fall in the bottom right corner (high sensitivity to identity features and low sensitivity to viewpoint features). Other types of representations could also be positioned in this space. For example, a chick could build a representation that is highly sensitive to both identity features and viewpoint features (upper right corner).

The chicks reared with temporally smooth objects tended to build abstract object representations that were sensitive to identity features but not viewpoint features (representations closer to the bottom right area of the abstraction space). Conversely, the chicks reared with temporally non-smooth objects tended to build object representations that were more sensitive to viewpoint features and less sensitive to identity features (representations closer to the top left area of the abstraction space). These results indicate that newborn chicks can build many different types of object representations, with the degree of abstraction varying as a function of the object's temporal smoothness (and speed of movement; see Wood & Wood, 2016b).

4. Discussion

Our goal in this study was to explore the experiential factors that enable newborn brains to build invariant object representations. Based on previous theoretical and behavioral work, we hypothesized that visual experience with temporally smooth objects would facilitate the development of invariant object recognition. To test this hypothesis, we reared newborn chicks in strictly controlled environments that contained a single virtual object moving smoothly or non-smoothly over time. When newborn chicks were reared with a temporally smooth object, the chicks successfully built view-invariant object representations that were selective for object identity and tolerant to viewpoint changes. In contrast, the chicks reared with a temporally non-smooth object tended to build representations that were more view-based (i.e., there was lower sensitivity to object identity and higher sensitivity to familiar views). Consequently, these results provide evidence for a “smoothness constraint” on the development of invariant object recognition. Newborn chicks develop more abstract object representations when exposed to temporally smooth objects. These results provide three main contributions to the literature.

First, this study builds on previous studies showing that temporal smoothness facilitates the development of object recognition (Wood, 2016; Wood et al., 2016) and extends this phenomenon to the domain of invariant object recognition. Invariant recognition is an important ability to understand both because it is required for real-world object recognition and because it is a complex computational task (DiCarlo & Cox, 2007). The computational challenge of object recognition is that retinal representations are not well suited for reading out behaviorally relevant dimensions (e.g., object identity) since changing the viewpoint, size, or retinal position of an object causes complex nonlinear changes in the object’s appearance on the retina. Thus, the brain must “reformat” retinal representations into higher level representations that allow behaviorally relevant properties to be read out by downstream neurons. The present results suggest that newborn vertebrate brains solve this problem, at least in part, by leveraging the temporal smoothness of natural visual objects.

It is worth emphasizing that the chicks reared with the temporally non-smooth objects built object representations that contained more view-based features despite being reared in environments that contained some temporally smooth features. For instance, the chicks acquired visual experience with the temporally smooth extended surfaces of the chamber, such as the walls, floor, and ceiling. The chicks also acquired visual experience with temporally smooth grain and water during feeding. Nevertheless, when the *objects* in the chicks’ visual environment failed to move smoothly over time, the chicks developed less selectivity for identity features and more selectivity for viewpoint features (compared to the chicks raised with temporally smooth objects).

Second, these results support unsupervised temporal learning models from computational neuroscience and computer vision (DiCarlo et al., 2012; Földiák, 1991; Rolls, 2012; Stone, 1996; Wiskott & Sejnowski, 2002). These models propose that biological visual systems learn perceptually relevant features by extracting smoothly and slowly changing features from the environment. Consistent with these models, this study provides evidence that the development of invariant object recognition is facilitated by visual

experience with smoothly changing features. Notably, there is also evidence that the development of invariant object recognition requires visual experience with slowly changing features (Wood & Wood, 2016b). Together, these findings confirm a central prediction of unsupervised temporal learning theories, by showing that newborn brains build view-invariant object representations by extracting smoothly and slowly changing features from the environment.

In the developmental psychology literature, there is also growing evidence for temporal learning abilities in human infants. For example, infants build visual and auditory representations by encoding spatiotemporal regularities in sensory input (Bulf, Johnson, & Valenza, 2011; Johnson et al., 2009; Kellman & Short, 1987; Kirkham, Slemmer, & Johnson, 2002; Saffran, 2003; Saffran, Aslin, & Newport, 1996; Wu, Gopnik, Richardson, & Kirkham, 2011). While it may seem surprising that humans and chicks would use common temporal learning strategies to learn about the world (given that their last common ancestor lived over 200 million years ago), humans and chicks do rely on homologous neural circuits to process sensory input. This “canonical” cortical circuit is thought to have evolved in stem amniotes over 300 million years ago and to underlie the computations of high-level perception and cognition (Güntürkün & Bugnyar, 2016; Jarvis et al., 2005; Karten, 2013). Since humans and chicks use homologous neural circuits to process sensory input, controlled-rearing experiments with newborn chicks provide a powerful and unique experimental avenue for probing how vision emerges in the brain.

Third, these results demonstrate that automated controlled-rearing experiments can be used to measure—with high precision—how specific visual inputs relate to specific behavioral outputs in a newborn animal. These input-output patterns are valuable because they provide specific targets for computational models. For instance, in order to successfully account for the development of object recognition in newborn chicks, a computational model would need to produce two patterns. First, the model should be capable of building invariant object representations from sparse visual input (e.g., from input of a single object seen from a limited viewpoint range) in an unsupervised manner. Second, the model should develop high selectivity for identity features, and low selectivity for viewpoint features, when provided with smooth and slow visual object input. As such, these input-output patterns place significant constraints on computational models of newborn object recognition.

It is important to note that temporal smoothness is a broad term that can refer to many different types of change across images (e.g., brightness-level change, pixel-level change, feature-level change, and so forth). As a result, additional research is needed to characterize how particular types of temporal smoothness influence the development of object recognition. Furthermore, temporal smoothness is a continuous variable rather than a binary variable. Thus, additional research is needed to characterize the precise amount of temporal smoothness that is needed for the development of invariant object recognition. In this study, for instance, the chicks reared with the temporally non-smooth objects still performed above chance levels on the Identity Trials, indicating that their representations were partially view-invariant. It would be interesting to examine whether chicks fail to develop view-invariance completely when raised with objects that are even less

temporally smooth than the objects used here (e.g., objects with zero overlap in pixels across successive frames).

In conclusion, this study provides evidence for a smoothness constraint on the development of invariant object recognition in a newborn animal. Newborn chicks can begin building view-invariant object representations within the first few days of life, but this ability does not emerge automatically. Rather, visual experience with temporally smooth objects facilitates the development of invariant object recognition.

Acknowledgments

This research was funded by National Science Foundation CAREER Grant BCS-1351892. We thank Brian W. Wood for helpful comments on the manuscript.

Note

1. A natural consequence of tracking all of the chicks' behavior is that the chicks could be sleeping, eating, or resting when the experiment switched from a rest period to a test trial. If this occurred, then the chicks 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 two objects. Computing the average of the "switch trials" and the "stay trials" corrected for this issue.

References

- Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research*, 39(17), 2885–2899.
- Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, 121(1), 127–132. <https://doi.org/10.1016/j.cognition.2011.06.010>.
- Cox, D. D., Meier, P., Oertelt, N., & DiCarlo, J. J. (2005). "Breaking" position-invariant object recognition. *Nature Neuroscience*, 8(9), 1145–1147. <https://doi.org/10.1038/nm1519>.
- DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11(8), 333–341. <https://doi.org/10.1016/j.tics.2007.06.010>.
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3), 415–434. <https://doi.org/10.1016/j.neuron.2012.01.010>.
- Feldman, J., & Tremoulet, P. D. (2006). Individuation of visual objects over time. *Cognition*, 99(2), 131–165. <https://doi.org/10.1016/j.cognition.2004.12.008>.
- Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, 3(2), 194–200. <https://doi.org/10.1162/neco.1991.3.2.194>.
- Güntürkün, O., & Bugnyar, T. (2016). Cognition without cortex. *Trends in Cognitive Sciences*, 20(4), 291–303. <https://doi.org/10.1016/j.tics.2016.02.001>.
- Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., Medina, L., Paxinos, G., Perkel, D. J., Shimizu, T., Striedter, G., Wild, J. M., Ball, G. F., Dugas-Ford, J., Durand, S. E., Hough, G.

- E., Husband, S., Kubikova, L., Lee, D. W., Mello, C. V., Powers, A., Siang, C., Smulders, T. V., Wada, K., White, S. A., Yamamoto, K., Yu, J., Reiner, A., Butler, A. B.; Avian Brain Nomenclature Consortium. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, 6(2), 151–159. <https://doi.org/10.1038/nrn1606>.
- Johnson, S., Fernandes, K., Frank, M., Kirkham, N., Marcus, G., Rabagliati, H., & Slemmer, J. (2009). Abstract rule learning for visual sequences in 8- and 11-Month-Olds. *Infancy*, 14(1), 2–18. <https://doi.org/10.1080/15250000802569611>.
- Karten, H. J. (2013). Neocortical evolution: Neuronal circuits arise independently of lamination. *Current Biology*, 23(1), R12–R15. <https://doi.org/10.1016/j.cub.2012.11.013>.
- Kellman, P. J., & Short, K. R. (1987). Development of three-dimensional form perception. *Journal of Experimental Psychology: Human Perception and Performance*, 13(4), 545–557.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83(2), B35–B42.
- Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, 321(5895), 1502–1507. <https://doi.org/10.1126/science.1160028>.
- Liu, T. (2007). Learning sequence of views of three-dimensional objects: The effect of temporal coherence on object memory. *Perception*, 36(9), 1320–1333. <https://doi.org/10.1068/p5778>.
- Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences*, 108(48), 19401–19406. <https://doi.org/10.1073/pnas.1112895108>.
- Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, 335(6193), 817–820. <https://doi.org/10.1038/335817a0>.
- Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology*, 58(1), 75–96. <https://doi.org/10.1146/annurev.psych.58.102904.190114>.
- Rolls, E. T. (2012). Invariant visual object and face recognition: Neural and computational bases, and a model, VisNet. *Frontiers in Computational Neuroscience*, 6, 35. <https://doi.org/10.3389/fncom.2012.00035>.
- Saffran, J. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12, 110–114.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- Stone, J. V. (1996). Learning perceptually salient visual parameters using spatiotemporal smoothness constraints. *Neural Computation*, 8(7), 1463–1492.
- Wallis, G., Backus, B. T., Langer, M., Huebner, G., & Bulthoff, H. (2009). Learning illumination- and orientation-invariant representations of objects through temporal association. *Journal of Vision*, 9(7), 6. <https://doi.org/10.1167/9.7.6>.
- Wallis, G., & Bulthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences of the USA*, 98(8), 4800–4804. <https://doi.org/10.1073/pnas.071028598>.
- Wiskott, L., & Sejnowski, T. J. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation*, 14(4), 715–770. <https://doi.org/10.1162/089976602317318938>.
- Wood, J. N. (2013). Newborn chickens generate invariant object representations at the onset of visual object experience. *Proceedings of the National Academy of Sciences*, 110(34), 14000–14005. <https://doi.org/10.1073/pnas.1308246110>.
- Wood, J. N. (2014). Newly hatched chicks solve the visual binding problem. *Psychological Science: A Journal of the American Psychological Society/APS*, 25(7), 1475–1481. <https://doi.org/10.1177/0956797614528955>.
- Wood, J. N. (2015). Characterizing the information content of a newly hatched chick's first visual object representation. *Developmental Science*, 18(2), 194–205. <https://doi.org/10.1111/desc.12198>.
- Wood, J. N. (2016). A smoothness constraint on the development of object recognition. *Cognition*, 153, 140–145. <https://doi.org/10.1016/j.cognition.2016.04.013>.

- 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. <https://doi.org/10.1007/s10071-016-0982-5>.
- Wood, S. M. W., & Wood, J. N. (2015). A chicken model for studying the emergence of invariant object recognition. *Frontiers in Neural Circuits*, 9(89), 7. <https://doi.org/10.3389/fncir.2015.00007>.
- Wood, J. N., & Wood, S. M. W. (2016a). Measuring the speed of newborn object recognition in controlled visual worlds. *Developmental Science*, 20, e12470. <https://doi.org/10.1111/desc.12470>.
- Wood, J. N., & Wood, S. M. W. (2016b). The development of newborn object recognition in fast and slow visual worlds. *Proceedings of the Royal Society B: Biological Sciences*, 283 (1829), pii:20160166. <https://doi.org/10.1098/rspb.2016.0166>
- Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, 47, 1220. <https://doi.org/10.1037/a0024023>.
- Zoccolan, D., Oertelt, N., DiCarlo, J. J., & Cox, D. D. (2009). A rodent model for the study of invariant visual object recognition. *Proceedings of the National Academy of Sciences*, 106(21), 8748–8753. <https://doi.org/10.1073/pnas.0811583106>.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Movie S1. The imprinting animation from the Input Phase showing Object 2 moving smoothly at a presentation rate of 10 frame/s.

Movie S2. The imprinting animation from the Input Phase showing Object 2 moving smoothly at a presentation rate of 1 frame/s.

Movie S3. The imprinting animation from the Input Phase showing Object 2 moving non-smoothly at a presentation rate of 10 frame/s.

Movie S4. The imprinting animation from the Input Phase showing Object 2 moving non-smoothly at a presentation rate of 1 frame/s.

Movie S5. A Viewpoint Trial from the Test Phase showing Object 2 moving smoothly at a presentation rate of 10 frame/s.

Movie S6. A Viewpoint Trial from the Test Phase showing Object 2 moving non-smoothly at a presentation rate of 1 frame/s.

Movie S7. An Identity Trial from the Test Phase showing the objects moving smoothly at a presentation rate of 10 frame/s.

Movie S8. An Identity Trial from the Test Phase showing the objects moving non-smoothly at a presentation rate of 1 frame/s.