One-Shot Object Parsing in Newborn Chicks

Samantha M. W. Wood[1] and Justin N. Wood[1, 2, 3]

[1] Department of Informatics, Indiana University
[2] Center for the Integrated Study of Animal Behavior, Indiana University
[3] Cognitive Science Program, Indiana University

Controlled-rearing studies provide the unique opportunity to examine which psychological mechanisms are present at birth and which mechanisms emerge from experience. Here we show that one core component of visual perception—the ability to parse objects from backgrounds—is present when newborn animals see their first object. We reared newborn chicks in strictly controlled environments containing a single object on a single background, then tested the chicks' object parsing and recognition abilities. We found that chicks can parse objects from natural backgrounds at the onset of vision, allowing chicks to recognize objects equally well across familiar and novel backgrounds. We also found that the development of object parsing requires motion cues, akin to the development of object parsing in human infants and newly sighted blind patients. These results demonstrate that newborn brains are capable of "one-shot object parsing" and show that motion cues scaffold object perception from the earliest stages of learning. We conclude that prenatal developmental programs build brain architectures with an object-based inductive bias, allowing animals to solve object perception tasks immediately without extensive experience with objects. We discuss the implications of this finding for developmental psychology, computational neuroscience, and artificial intelligence.

Keywords: visual parsing, object perception, newborn, controlled rearing, chicks

Supplemental materials: https://doi.org/10.1037/xge0001043.supp

To perceive objects successfully, newborn brains must perform a difficult task: They must parse objects from natural backgrounds and build invariant object representations that generalize across new viewing situations (e.g., changes in viewpoint, background, and illumination conditions). This is a difficult computational task because real-world images typically contain many regions with different colors and luminances, so the visual system must decide how to carve and combine these regions into meaningful object percepts. While many studies have examined how mature subjects parse objects, methodological barriers have prevented detailed examination of object parsing in newborn organisms. Thus, the origins of this ability remain unknown. How do newborn brains learn to parse objects from backgrounds? What mechanisms underlie this core visual ability?[1]

Statistical Learning

One candidate learning mechanism for visual parsing is statistical learning. The term "statistical learning" refers to processes that enable detection of regularity in sensory inputs (e.g., co-occurrence frequencies, transitional probabilities, nonadjacent dependencies). Studies from developmental psychology (Krogh et al., 2012; Saffran et al., 1996; Smith et al., 2014) and computational neuroscience (DiCarlo et al., 2012; Wiskott & Sejnowski, 2002) suggest that visual systems learn to interpret sensory input through statistical learning, by associating features that co-occur in the input stream. Infants and adults, for example, extract statistical regularities from sensory input to construct higher-order object concepts (Fiser et al., 2007; Kirkham et al., 2002; Saffran & Kirkham, 2018; Turk-Browne et al., 2005).

Statistical learning also plays a central role in the development of artificial visual systems. For example, deep artificial neural net-

Samantha M. W. Wood ⓘ https://orcid.org/0000-0002-2219-0285
Justin N. Wood ⓘ https://orcid.org/0000-0002-2219-0285

[1] "Object parsing" refers to the ability to segment objects from backgrounds and bind object features into representations for recognition. Object parsing is one of many abilities involved in object perception. Object perception also involves object individuation (e.g., Xu & Carey, 1996), object permanence (e.g., Bremner et al., 2015), amodal completion (e.g., Kellman & Spelke, 1983), the discovery of object affordances (e.g., Bushnell & Boudreau, 1993), and so forth.

works (ANNs) trained to extract statistical regularities from millions of diverse training images can learn to recognize objects across novel viewing situations—achieving human-level performance on challenging object recognition tasks (Krizhevsky et al., 2012; Yamins & DiCarlo, 2016). Moreover, ANN models can accurately predict neural responses to novel images and be used to control the activity state of populations of neurons (e.g., Bashivan et al., 2019). Thus, ANN models provide a mechanistic understanding of object recognition in mature visual systems.

While statistical learning of spatial features is a powerful framework for understanding object recognition, this type of learning is insufficient to support object recognition when a newborn has limited experience with the world. For instance, if a newborn's visual experience consists of a single object seen on a single background, then the object and background will have a 100% concurrence rate in the newborn's visual input. Statistical learning mechanisms bind together features with high concurrence rates, so a newborn that relies solely on statistical learning of spatial features would build an 'incorrect' object representation that includes features from the background.

Indeed, ANN models typically require thousands to millions of diverse training images to successfully dissociate object and background features, and many still suffer from this "statistical concurrence problem" (Mordvintsev et al., 2015). For example, when ANN models were trained to recognize objects (e.g., dumbbells), but the objects were presented in similar scenes across the training images (e.g., the dumbbells were always held by weightlifters), then the resulting visual representations incorrectly linked object and scene features (e.g., the dumbbell representation incorrectly included hand and arm features; for details see Mordvintsev et al., 2015). Thus, to overcome this statistical concurrence problem, biological visual systems might use additional mechanisms to parse objects from backgrounds.

Motion-Based Visual Parsing

A second candidate learning mechanism for visual parsing is motion-based parsing. Studies of infants suggest that motion-based parsing plays an important role in the development of object perception. Young infants use motion cues to determine the three-dimensional shape of objects (Arterberry & Yonas, 2000; Owsley, 1983) and to integrate spatially separated parts into unified object concepts (Johnson et al., 2002; Kellman & Spelke, 1983; Kellman et al., 1986). Furthermore, the ability to parse objects using motion cues tends to emerge *before* the ability to parse objects using static cues (e.g., color, shape; Spelke, 1990). During early stages of visual learning, motion appears to be critical for both parsing objects and binding their constituent features into representations for recognition.

Complementary evidence for motion-based parsing comes from studies exploring visual learning after recovery from blindness. Like young infants, newly sighted patients do not rely on prominent figural cues of grouping (e.g., good continuation, junction structure) to parse objects (Ostrovsky et al., 2009). Rather, they use motion cues to parse objects and develop representations that permit recognition in static images. For both infants and patients recovering from blindness, dynamic information provides a key organizing influence for early visual processing. Motion-based parsing might therefore be a "primitive" of object perception—a program of visual learning that enables the brain to assemble fragmented features into unified object representations.

One attractive feature of motion-based parsing is that it can overcome the statistical concurrence problem described above. Motion can act as a selective-gating mechanism, constraining learning to moving features, rather than the whole visual field. Since objects typically move separately from backgrounds, motion-based parsing can segment objects from backgrounds without the need for diverse visual experiences with objects.

While the studies cited above indicate that motion-based parsing facilitates early visual learning, it is unclear what role experience plays in the development of this ability. Both the infants and the patients recovering from blindness had already acquired weeks to months of uncontrolled visual experience before they were tested, so in principle, motion-based parsing might be learned from early visual experience. While a few studies have explored whether newborn humans use motion cues to integrate spatially separated parts into unified object representations, the results from these studies have been mixed, with the majority of studies reporting no evidence for this ability (Slater et al., 1990; Slater et al., 1996; Valenza & Bulf, 2011; Valenza et al., 2005). Furthermore, to our knowledge, no studies have tested whether newborn organisms can parse objects from natural backgrounds—a critical skill for object perception in the real world. Consequently, it is not clear (a) whether motion-based parsing is a primitive of object perception (i.e., present and functional at the onset of vision); or (b) what role experience plays in the development of object parsing. To address these questions, we used a controlled-rearing approach.

Using Automated Controlled Rearing to Study the Origins of Object Parsing

When animals are raised in uncontrolled environments, it is difficult (if not impossible) to characterize the precise role that experience plays in visual development. Conversely, with controlled-rearing experiments, researchers can systematically manipulate the visual experiences provided to newborn subjects and measure the effects of those manipulations on perceptual and cognitive development. This is important because visual experience can rapidly shape biological neural networks (Arcaro et al., 2017; Li & DiCarlo, 2008). Controlled-rearing experiments therefore provide an experimental avenue for probing the core learning mechanisms in newborn brains and for characterizing the role of experience in visual development.

We used newborn chicks (*Gallus gallus*) as a model system because they are uniquely suited for studying the earliest stages of visual learning. Unlike commonly used animal models in psychology (e.g., rats, pigeons, monkeys), chickens are a precocial species (mobile in the first day of life) and can be raised in strictly controlled environments immediately after hatching (Wood, 2013; Wood & Wood, 2015). While previous studies have shown that newborn chicks can perceive partly occluded objects (Lea et al., 1996) and build invariant representations of objects seen on homogeneous white backgrounds (Wood, 2013, 2016; Wood & Wood, 2015, 2016, 2018, 2020), no studies have examined whether (and how) newborn chicks learn to parse objects from natural backgrounds. Controlled-rearing studies of chicks provide a valuable opportunity to explore how newborn brains solve one of the central challenges of object perception.

Recently, it has also become possible to fully automate controlled-rearing studies of newborn chicks. Automation allows

newborn chicks to be observed continuously (24/7), producing large amounts of data per subject. For example, whereas nonautomated methods typically produce a few minutes of test data per chick (e.g., 6 minutes), we collected 10,080 minutes of test data per chick, a 1,680-fold increase in data quantity. As a result, we can perform statistical analyses of each chick's performance to assess whether each chick succeeded on the task. This is valuable from a theoretical perspective. If motion-based object parsing is a core learning mechanism in newborn brains, then it should be present in most (if not all) newborn animals. By collecting enough data to analyze each individual's performance, our automated approach provides a strong test of the hypothesis that motion-based object parsing is a core learning mechanism in newborn brains.

Automation also offers many additional benefits over nonautomated methods (Wood & Wood, 2019). First, because computers present the stimuli and code the behavior, automated methods are free of experimenter bias. Second, because automation produces large amounts of data per subject, we can substantially *decrease* measurement error and substantially *increase* effect sizes, resulting in high-powered experiments. Third, by monitoring behavior 24/7, automation permits longitudinal analyses of how visual learning changes over time. Taken together, automation can be used to illuminate the core learning mechanisms in newborn brains with an unprecedented level of precision.

The Present Study

To investigate the origins of visual parsing, we tested whether newborn chicks are capable of "one-shot object parsing." Specifically, we presented chicks with a challenging task, in which their visual world contained a single object on a single background (input phase, Figure 1A). As a result, the object and background had a 100% concurrence rate. During the test phase (Figure 1B), we then tested whether the chicks could recognize their imprinted object when the object was presented on familiar and novel backgrounds.

If newborn chicks depend solely on statistical learning to parse and recognize objects, then they should bind the object and background features together during the input phase. Accordingly, the chicks' object recognition performance should be high when the object is presented on the familiar background and low when the object is presented on novel backgrounds. Conversely, if newborn chicks are capable of one-shot object parsing (i.e., they have mechanisms for segmenting objects from backgrounds at the onset of vision), then their recognition performance should be high, regardless of whether the object is presented on familiar or novel backgrounds.

Experiment 1

Our controlled-rearing method was designed to mimic the experimental design from machine learning studies of ANN models. During the training phase, ANN models are provided with a set of training data for learning. During the test phase, learning is turned off, and the ANN model is tested on a variety of familiar and novel stimuli. With this approach, researchers can measure what an ANN model has learned from a specific set of training data. Likewise, in our study, we presented chicks with a set of training data

during the input phase (a single object moving on a single background). Then, during the test phase, we measured what the chicks had learned from the training data provided in the input phase. By using the same experimental design as machine learning studies, this approach allows researchers to make direct comparisons between the learning abilities of newborn animals and ANN models: an essential step for building neural network models of visual development.

Unlike traditional controlled-rearing studies of newborn chicks, in which the chick typically receives a short input phase (e.g., 2 hr) and test phase (e.g., 6 min), our design included a long input phase (5 days) and test phase (7 days). We used a long input phase because imprinting in chicks typically ends ~3 days after hatching. By exposing the chicks to the imprinting stimulus for 5 days, we could ensure that the imprinting period had ended and that the chicks would not imprint to the test stimuli (Wood & Wood, 2015). We used a long test phase because we wanted to collect precise measurements of performance from each subject. In previous work, we found that increasing the amount of data collected from each chick can increase the effect size by a factor of four and reduce measurement error by a factor of three (Wood & Wood, 2019). Thus, collecting more data per chick significantly improves the precision of the data, while also producing high-powered experiments.

Using long input and test phases 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 two-hour imprinting window). Moreover, 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.

Method

Subjects

Thirty-one Rhode Island Red chicks of unknown sex were tested. The sample size was determined before the experiment was conducted. The sample size was based on previous automated controlled-rearing experiments with newborn chicks (Wood, 2013, 2015). Specifically, a minimum sample size for achieving 80% power was determined based on prior studies (Wood, 2013); and the final sample size was about three times larger than the minimum sample size to accommodate counterbalancing of the stimuli. Ultimately, this experiment was powered to over 99.9% for detecting effect sizes (Cohen's d) of $d = 1.0$ or above.[2]

No subjects were excluded from the analyses. The eggs were obtained from a local distributer and incubated in darkness in an OVA-Easy incubator (Brinsea Products Inc., Titusville, FL). To avoid

---

[2] While an effect size of $d = 1.0$ is considered a very large effect in the psychology literature, this is a typical effect size estimate in automated controlled-rearing studies with newborn chicks. For example, Wood (2013) and Wood (2015) found effect sizes with an average of $d = 1.4$ for view-invariant object recognition tasks. Automated methods allow large amounts of data to be collected from each newborn chick, which reduces measurement error and increases the effect size of experiments (Wood & Wood, 2019).

Figure 1
The Experimental Design
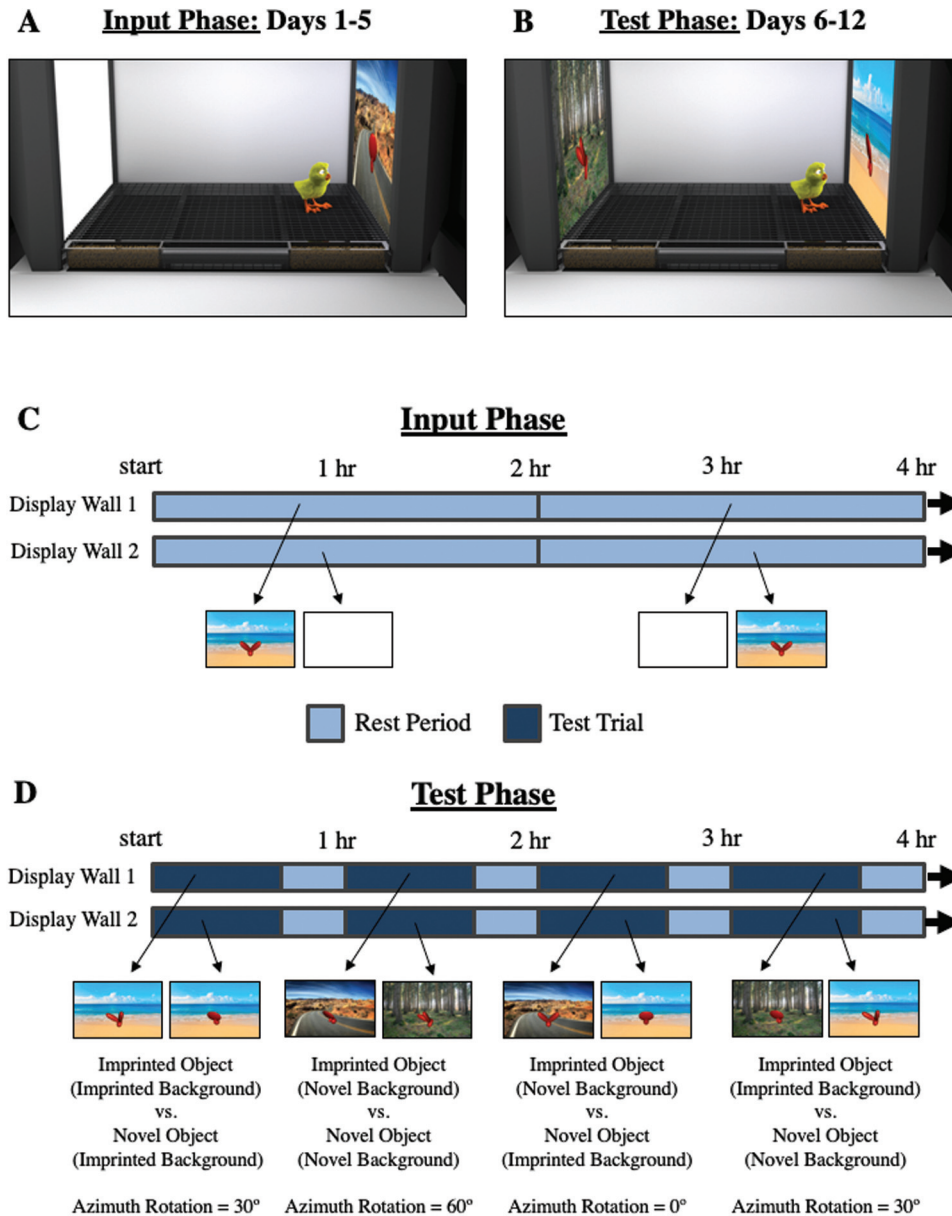
Note. (A) During the input phase, the chicks were raised with a single virtual object moving on a single background. (B) During the test phase, the chicks were presented with test trials in which the imprinted object appeared on one display wall and a novel object appeared on the opposite display wall. The timelines show how the virtual objects were presented on the two display walls during sample 4-hr periods during the (C) input phase and (D) test phase. Across the test trials, the imprinted object and novel object were presented on the familiar background and on novel backgrounds, in a fully crossed design. The object was also presented from familiar viewpoints (0° azimuth rotation) and novel viewpoints (30° and 60° azimuth rotations) on different test trials. This schematic illustrates examples of each background test condition; the actual order of the test trials was randomized across the experiments. See the online article for the color version of this figure.

exposing the chicks to any extraneous visual input, we used night vision goggles to move the chicks in darkness from the incubation room to the controlled-rearing chambers. Each chick was raised within its own chamber. This research was approved by The University of Southern California Institutional Animal Care and Use Committee.

Procedure

The chicks were raised in automated controlled-rearing chambers (66 cm length × 42 cm width × 69 cm height) that were constructed from white, high-density polyethylene. The chambers

contained no real-world (solid, bounded) objects. To present object stimuli to the chicks, virtual objects were projected on two display walls situated on opposite sides of the chamber. The display walls were 19" liquid crystal display monitors (1,440 × 900 pixel resolution). Food and water were provided *ad libidum* in transparent troughs in the ground. We used grain 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 were constructed from black wire mesh supported by transparent beams. Microcameras in the ceilings of the chambers recorded all of the chicks' behavior, and the video feed was analyzed with automated image-based tracking software (EthoVision XT, Noldus Information Technology, Leesburg, VA). This software calculated the amount of time each chick spent within zones (22 cm × 42 cm) next to the left and right display walls. All of the chicks' behavior (9 samples/s, 24 hr/day, 7 days/week) was tracked and recorded across the 12 days of the experiment. In total, we collected 8,928 hr of video footage for this experiment (24 hr per day × 12 days × 31 subjects).

During the input phase (Days 1–5), the chicks were raised in controlled-rearing chambers that contained a single virtual object rotating around a fronto-parallel horizontal axis (see Figure 1). The object rotated continuously, completing a full rotation every 15 s. The object was presented on one of three background scenes. Each chick only saw the imprinted object on a single background, and the background scene was counterbalanced across chicks. The object appeared on one display wall at a time and switched to the opposite display wall every 2 hr (following a 1-min period of darkness). The display wall that was not showing the imprinted object was white. Figures 1A and 1C illustrate how the imprinting stimuli were presented on the display walls during the input phase. Supplementary Movie 1 shows a sample animation presented during the input phase.

During the test phase (Days 6–12), we tested whether the chicks could recognize their imprinted object when the object was presented on familiar and novel backgrounds. On each test trial, the imprinted object appeared on one display wall, and a novel object appeared on the opposite display wall (Figures 1B and 1D). The novel object was the same color as the imprinted object and moved at the same speed along the same motion trajectory. The two objects were modeled after those used in previous studies that tested for invariant object recognition in adult rats (Zoccolan et al., 2009) and newborn chicks (Wood, 2013, 2015). Fifteen of the chicks were imprinted to Object 1, with Object 2 serving as the novel object, and 16 of the chicks were imprinted to Object 2, with Object 1 serving as the novel object.

Across the test trials, the objects were presented on all possible combinations of the three background scenes (Background 1 vs. Background 1, Background 1 vs. Background 2, Background 1 vs. Background 3, etc.). Accordingly, we grouped the test trials into four background conditions. In the "imprinted object (imprinted background) versus novel object (imprinted background)" condition, both the imprinted object and the novel object were shown on the imprinted background from the input phase. In the "imprinted object (novel background) versus novel object (novel background)" condition, both the imprinted object and the novel object were shown on a novel background. In the "imprinted object (imprinted background) versus novel object (novel background)" condition, the imprinted object was shown on the

imprinted background, and the novel object was shown on a novel background. Finally, in the "imprinted object (novel background) versus novel object (imprinted background)" condition, the imprinted object was shown on a novel background, and the novel object was shown on the imprinted background.

In each of the background conditions, the objects were also shown from three possible viewing angles: 0° change in azimuth rotation, 30° change in azimuth rotation, and 60° change in azimuth rotation. We included these viewpoint changes to test whether the chicks were simply encoding and recognizing familiar features versus building high-level object representations that generalize across novel viewpoints. If chicks simply encode familiar features, then their recognition performance should be higher when the object is shown from the familiar viewpoint (when familiar features are present) and lower when the object is shown from novel viewpoints (when novel features are present). Conversely, if chicks build high-level (view-invariant) object representations, then their recognition performance should be high regardless of whether the object is shown from familiar or novel viewpoints.

The order of the test trials was randomized across the test phase. The imprinted object and novel object were shown from the same viewing angle within a test trial. During the trials, the objects rotated 360° around a fronto-parallel horizontal axis (as in the input phase). Supplementary Movie 2 shows sample animations presented during the test phase.

The chicks received 24 test trials per day at the rate of one trial per hour. Each test trial lasted 40 min and was followed by a 20-min rest period. During the rest periods, the animation from the input phase appeared on one display wall, and a white screen appeared on the other display wall. We included rest periods in the design for two reasons. First, the rest periods allowed the chicks to reunite with their imprinted object without needing to make a choice between two objects. Second, the rest periods provided a baseline measure of the amount of time the chicks generally preferred to spend in proximity to their imprinted stimuli.

Results

Overall Recognition Performance

The results are shown in Figure 2. Our dependent variable was the proportion of time the chick spent by their imprinted object versus the novel object. To compute performance, we used the following formula: time by imprinted object/[time by imprinted object + time by novel object]. Performance was well above chance level (50%) in all four background conditions (one sample t-tests, all $ps < 10^{-7}$, all Cohen's $ds > 1.2$). Performance was also well above chance level in all three viewpoint conditions (one sample t-tests, all $ps < 10^{-10}$, all Cohen's $ds > 1.7$). The chicks successfully recognized their imprinted object, regardless of whether the object was presented on familiar or novel backgrounds or whether the object was presented from familiar or novel viewpoints. In all background and viewpoint conditions, the analyses exceeded five sigma levels of statistical significance (the threshold for new discoveries in theoretical physics).

To examine whether performance differed across the background and viewpoint conditions, we performed a repeated-measures ANOVA with the within-subjects factors of background

Figure 2
Experiment 1 Results
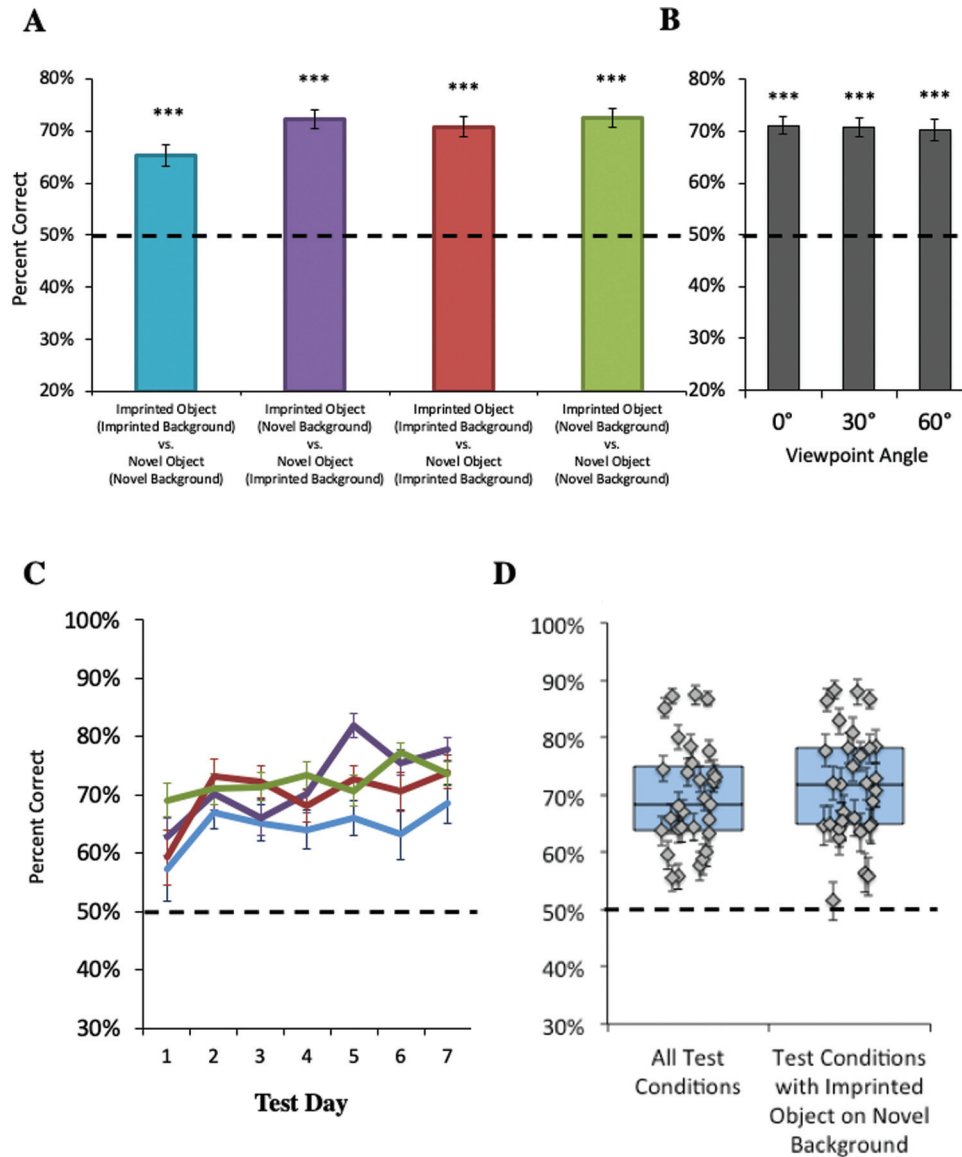
Note. (A) Recognition performance in the four background conditions. (B) Recognition performance in the three viewpoint conditions. The chicks' performance (percent time spent with the imprinted object versus novel object) was well above chance level in all test conditions. *** denotes $p < .001$ (C) Recognition performance in each background condition on each test day. While performance improved across the test phase, the chicks successfully recognized their imprinted object on all test days, including Day 1 of testing. (D) Individual subject performance. The graphs show the percent of time each chick spent with the imprinted object versus the novel object. Each chick is represented by a gray marker. The blue boxes indicate the 25th to 50th percentile and the 50th to 75th percentile of performance. Dashed lines indicate chance performance. Error bars show ± 1 standard error (*SEM*). See the online article for the color version of this figure.

condition and viewpoint angle. The ANOVA revealed a significant main effect of background condition, $F(3, 90) = 9.455$, $p = .00002$, $\eta^2 = .240$, and a significant interaction between background condition and viewpoint angle, $F(6, 180) = 2.509$, $p = .023$, $\eta^2 = .077$. The main effect of viewpoint angle was not significant, $F(2, 60) = .032$, $p = .969$, $\eta^2 = .001$. The significant interaction between background condition and viewpoint angle was driven by slightly

lower performance when the imprinted object was presented on the familiar background from 0° (with the novel object presented on the novel background). Importantly, this should have been the condition with the *highest* performance if the chicks suffered from the statistical concurrence problem and failed to segment object and background features. We suspect that the slightly lower performance in this condition may reflect newborn chicks' competing

motivations to spend time with the imprinted object and to explore novel information (i.e., the novel background), although additional research is needed to confirm this hypothesis.

Analysis of Change in Performance Over Time

Overall recognition performance by test day is shown in Figure 2C. To determine whether performance changed across the test phase, we performed a repeated-measures ANOVA with the within-subjects factor of test day. The ANOVA revealed a significant main effect of test day, $F(6, 180) = 8.585$, $p < 10^{-7}$, $\eta^2 = .223$. A posthoc correlation between test day and the average performance for that day revealed a significant positive relationship between test day and performance ($r = .888$, $p = .008$). Thus, the chicks' performance improved across the test phase. Critically, however, performance was significantly above chance level even on Day 1 (one-sample t-test, $t(30) = 6.987$, $p < 10^{-7}$, Cohen's $d = 1.255$). In fact, when the analysis only included the test trials in which the imprinted object was shown on novel backgrounds, performance was still significantly above chance level on Day 1 (one-sample t-test, $t(30) = 6.736$, $p < 10^{-6}$, Cohen's $d = 1.210$). Therefore, the chicks' ability to segment object features from background features cannot be explained solely in terms of learning across the test phase.

There are three potential reasons why performance increased across the test phase. First, in addition to an imprinting response, the chicks might have been motivated by a novelty response. A novelty response would have reduced performance early in the test phase (when the novel object was less familiar) and improved performance as the novel object became more familiar. If this were the case, then we would expect to find an increase in performance across the test phase in our prior studies of object recognition in newborn chicks (e.g., Wood, 2013, 2015; Wood & Wood, 2016, 2018). However, in our prior studies, performance generally did not increase across the test phase, which provides evidence against this account.

Second, the chicks might have continued to develop a stronger imprinting response across the test phase. If the imprinting response did become stronger across the test phase, then performance during the rest periods of the test phase should have increased as well. To test this hypothesis, we examined the proportion of time the chicks spent by the imprinted object during the rest periods. We found that rest period performance significantly *decreased* across the test phase (one-way ANOVA: $F(6, 180) = 6.070$, $p = .0002$; $\eta^2 = .168$; correlation: $r = -.794$, $p = .033$). Thus, the chicks' improvement in performance does not appear to be caused by an increase in imprinting strength across the test phase.

Third, the chicks might have developed stronger object parsing abilities across the test phase. When objects are seen on natural backgrounds, object parsing might continue to develop over time, even after the imprinting period has ended. While prenatal developmental programs might lay the foundations for object parsing, experience seeing objects on different backgrounds might provide additional training data for optimizing the visual system. This is our working hypothesis, but future research is needed to test this possibility directly.

Analysis of Individual Subject Performance

Because we collected a large number of test trials from each subject, we were able to analyze each chick's object recognition performance with high precision (Figure 2D). Collapsing across the test trials, all of the chicks spent more time with the imprinted object than the novel object (one-sample t-tests, 27 subjects, ps < .001; two subjects, ps < .01; two subjects, ps < .05; all p values survived Holm-Bonferroni correction for multiple comparisons). After limiting the analysis to the test trials in which the imprinted object was shown on a novel background, 28 of the 31 chicks spent more time with the imprinted object than the novel object (one-sample t-tests, 28 subjects, ps < .001; all 28 significant p values survived Holm-Bonferroni correction for multiple comparisons). Thus, the majority of the chicks successfully segmented the object and background features.

Discussion

Experiment 1 demonstrates that newborn chicks can successfully parse objects from natural backgrounds, even after acquiring visual experience with just a single object on a single background. The chicks also successfully recognized their imprinted object across novel viewpoints. Thus, newborn chicks are capable of one-shot learning of both background-invariant and view-invariant object representations.

These results rule out the alternative account that the chicks simply encoded and recognized clusters of features, without actually having built *object* representations per se. If the chicks had simply encoded familiar features, then their recognition performance should have been higher when the object was shown from familiar viewpoints (when familiar features were present) and lower when the object was shown from novel viewpoints (when novel features were present). In contrast, recognition performance was high across both the familiar and novel viewpoints, indicating that the chicks successfully built high-level (view-invariant) object representations. A growing body of work in the neurosciences has demonstrated that complex computational systems are needed to solve view-invariant and background-invariant recognition tasks (e.g., DiCarlo & Cox, 2007; Yamins & DiCarlo, 2016). Because our chicks succeeded on both view-invariant and background-invariant recognition tasks, these results indicate that newborn brains can be equipped with powerful visual processing machinery at the onset of vision.

These results raise the question of *how* newborn chicks parse objects from backgrounds. One possibility is that chicks selectively encode moving objects. As discussed above, prior studies have shown that motion cues are critical for object parsing and recognition in human infants (Spelke, 1990) and patients recovering from blindness (Ostrovsky et al., 2009). Similarly, in Experiment 1, the object moved while the background was stationary. Thus, the chicks could have selectively encoded moving objects to parse the object from the background.

A second possibility is that the 360° rotation of the object provided sufficient image variation for a statistical learning mechanism to segment objects from backgrounds using spatial cues. As the object rotated, parts of the object moved across different features of the background, creating many object-background feature changes at the boundaries of the object. In principle, this image variation might be sufficient for object parsing through statistical learning.

A third possibility is that the object features provided sufficient cues to parse the object from the background. The objects were

salient due to their homogeneous color, relatively large size, and placement (superimposed on the middle of the background). Thus, these object features may have allowed the chicks to select the object through feature-based attention and parse the object from the background.

To distinguish between these possibilities, we repeated Experiment 1 with one crucial change: Rather than presenting a moving object, we reduced the motion speed of the object so that, from an observer's perspective, the object appeared to be stationary (i.e., the object rotated extremely slowly, at the rate of 1° per min). Over time, however, the object rotated 360° (as in Experiment 1), so the chicks were presented with all of the individual images shown in Experiment 1. Thus, the chicks received the same statistical and object feature information in Experiments 1 and 2, but the object moved in Experiment 1 and appeared stationary in Experiment 2.

If newborn chicks parse objects by selectively encoding moving objects, then the chicks should fail to recognize their imprinted object in Experiment 2 (because the motion cues were removed). Conversely, if the displays used in Experiment 1 provided sufficient image variation for object parsing through statistical learning and/or the object features were sufficiently salient for object parsing, then the chicks should successfully recognize their imprinted object in Experiment 2 (because the chicks in Experiments 1 and 2 were presented with the same image variation and object features).

Experiment 2

Method

The methods in Experiment 2 were identical to those used in Experiment 1, with two key changes. First, we tested a new group of 36 chicks. Second, we used the same individual images as in Experiment 1, but the images changed at a rate of 1 frame/min (as opposed to 24 frames/s). Thus, while the object completed a full rotation every 15 s in Experiment 1, the object completed a full rotation every 360 min in Experiment 2 (Figure 3A). The objects rotated at the same rate during the input phase and test phase.

Results

Overall Recognition Performance

The results are shown in Figures 3B and 3C. Overall recognition performance barely exceeded chance level (50%) in Experiment 2 (one-sample t-test, $M = 53\%$, $t(35) = 2.152$, $p = .038$, Cohen's $d = .359$). Comparing across the experiments, overall recognition performance was significantly lower in Experiment 2 compared with Experiment 1, $t(56.7) = 8.662$, $p < 10^{-11}$. When motion cues were removed from the displays, newborn chicks were severely impaired at recognizing their imprinted object.

In Experiment 2, performance exceeded chance level in the imprinted object (novel background) versus novel object (novel background) condition ($M = 53\%$, $t(35) = 2.365$, $p = .024$, Cohen's $d = .394$) and the imprinted object (imprinted background) versus novel object (imprinted background) condition ($M = 54\%$, $t(35) = 2.247$, $p = .031$, Cohen's $d = .374$). However, neither survived Holm-Bonferroni correction for multiple comparisons. For each

background condition, recognition performance was significantly lower in Experiment 2 compared with Experiment 1 (all ps $< .02$; all ps survive Holm-Bonferroni correction for multiple comparisons). In sum, in each background condition, the chicks were significantly impaired at recognizing objects when motion cues were removed.

Similarly, the chicks' performance in Experiment 2 marginally exceeded chance level in two of the viewpoint conditions (one-sample t-tests, 0°: $M = 53\%$, $t(35) = 2.163$, $p = .037$, Cohen's $d = .360$; 30°: $M = 53\%$, $t(35) = 2.152$, $p = .038$, Cohen's $d = .359$). However, neither survived Holm-Bonferroni correction for multiple comparisons. For each viewpoint condition, recognition performance was significantly lower in Experiment 2 compared with Experiment 1 (all ps $< 10^{-8}$; all ps survive Holm-Bonferroni correction for multiple comparisons). Together, these results support the hypothesis that newborn chicks selectively encode moving features. When an object lacks motion cues, newborn chicks develop impaired object recognition, regardless of whether the object is presented in a familiar context (imprinted background and viewpoint) or a novel context (new background and viewpoint).

#### Analysis of Rest Period Performance

The results from Experiments 1 and 2 indicate that motion cues drive the development of object parsing and recognition in newborn chicks. Accordingly, motion cues might drive the imprinting response in chicks, such that chicks imprint more strongly to moving objects than to stationary objects. While some prior studies indicate that newborn chicks will imprint to stationary stimuli (e.g., Regolin & Vallortigara, 1995; Rosa-Salva et al., 2010), other studies have emphasized the importance of motion cues for imprinting (Horn, 2004; Wood, 2017). To examine whether motion cues influenced the strength of the imprinting response in our study, we measured the proportion of time the chicks spent by the imprinted object during the rest periods.

Overall, the chicks in Experiment 2 spent the majority of their time with the imprinted object during the rest periods ($M = 75\%$, $t(35) = 13.182$, $p < 10^{-14}$). However, when we compared rest period performance across Experiments 1 and 2, we found that performance differed significantly across experiments ($t(50.02) = 3.914$, $p = .0003$; Cohen's $d = .915$), with the chicks spending more time with their imprinted object in Experiment 1. This finding raises the possibility that the chicks' impaired recognition performance in Experiment 2 occurred simply because they were less interested in the imprinting stimulus.

To test this hypothesis, we reanalyzed the results from Experiment 2, controlling for differences in rest period performance across experiments. Specifically, we removed the chicks in Experiment 2 with the lowest rest period performance until the average rest period performance matched across experiments (84% rest period performance; 19 chicks removed). Then, we repeated our Experiment 2 analyses with this subset of high-rest-period-performance chicks.

The overall pattern of results was the same after controlling for rest period performance. Performance exceeded chance level on one background condition, imprinted object (novel background) versus novel object (novel background) condition ($M = 55\%$, $t(16) = 2.154$, $p = .047$, Cohen's $d = .522$), but this did not survive Holm-Bonferroni correction for multiple comparisons. For all but

Figure 3
Experiment 2 Results
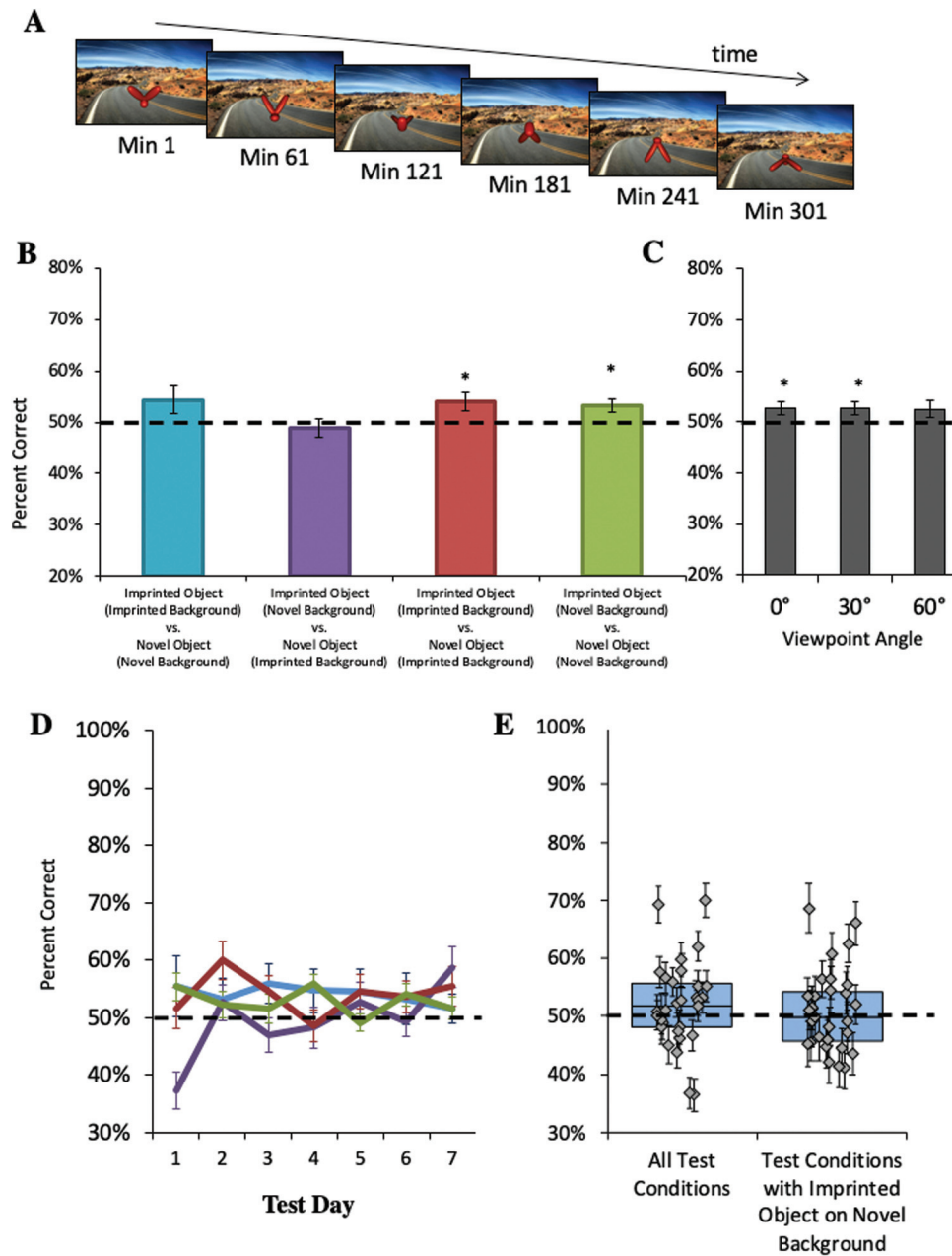
Note. (A) Sample frames from the object displays in Experiment 2. The object rotated 360° at a rate of 1°/min. The chicks in Experiment 2 saw all of the same images as the chicks in Experiment 1, but without clear motion cues. (B) Recognition performance in the four background conditions. (C) Recognition performance in the three viewpoint conditions. * denotes $p < .05$ (D) Recognition performance in each background condition on each test day. Performance remained low and stable across the test days. (E) Individual subject performance. The graphs show the percent of time each chick spent with the imprinted object versus the novel object. Each chick is represented by a gray marker. The blue boxes indicate the 25th to 50th percentile and the 50th to 75th percentile of performance. Dashed lines indicate chance performance. Error bars show ±1 standard error (*SEM*). See the online article for the color version of this figure.

one background condition, recognition performance was significantly lower in Experiment 2 compared with Experiment 1 ($ps <$ .001; all significant ps survive Holm-Bonferroni correction for multiple comparisons).

Similarly, the chicks' performance exceeded chance level in two viewpoint conditions (one-sample t-tests, 0°: $M = 55\%$, $t(16) =$ 2.265, $p = .038$, Cohen's $d = .549$; 30°: $M = 55\%$, $t(16) = 2.301$, $p = .035$, Cohen's $d = .558$), but neither survived Holm-Bonferroni correction for multiple comparisons. For each viewpoint condition, recognition performance was significantly lower in Experiment 2 compared with Experiment 1 (all $ps < 10^{-4}$; all ps survive Holm-Bonferroni correction for multiple comparisons). Thus, chicks' recognition impairment in Experiment 2 cannot be explained solely in terms of a reduced imprinting response or a lack of interest in the imprinting displays.

Discussion

Experiment 2 indicates that motion cues facilitate the development of object parsing and recognition in newborn chicks. When motion cues were removed from the displays, the chicks were highly impaired at recognizing their imprinted object across familiar and novel backgrounds and across familiar and novel viewpoints. These results rule out the second and third possibilities mentioned above. If the object displays used in Experiment 1 provided sufficient image variation for object parsing through statistical learning, then the chicks should have succeeded in Experiment 2 (because the chicks in Experiments 1 and 2 saw the same set of images). Similarly, if the object features themselves were sufficiently salient to drive object parsing and recognition, then the chicks should have successfully recognized their imprinted object in Experiment 2 (because the chicks in Experiments 1 and 2 saw the same object features). We conclude that motion cues play a critical role in the development of object parsing and recognition in complex real-world scenes.

General Discussion

Controlled-rearing studies provide a unique opportunity to examine which psychological mechanisms are present at birth and which mechanisms emerge from experience. Here we show that one core mechanism of visual perception—the ability to parse objects from backgrounds—is present when newborn animals see their first object. Despite acquiring experience with just one object on one background (100% object-background concurrence rate), newborn chicks built accurate object representations that generalized across novel backgrounds and novel viewpoints. Thus, newborn brains can perform "one-shot object parsing." These results provide two important contributions to the literature.

First, these results demonstrate that motion-based object parsing can be present and functional at the onset of vision. For decades, researchers have speculated that motion-based parsing is foundational to early visual learning, based on studies of human infants (Kellman & Spelke, 1983) and newly sighted blind patients (Ostrovsky et al., 2009). However, because the infants and patients had already acquired weeks to months of natural visual experience before they were tested, it was not possible to determine whether motion-based parsing is a primitive of object perception or whether this capacity is learned from early visual experience. Our

results provide clear evidence—both on the group level and on the individual level—that motion-based parsing scaffolds object perception from the earliest stages of visual learning. From a computational perspective, this is a valuable strategy. Restricting learning to moving objects (rather than the whole visual field) reduces the computational complexity of object perception, while also allowing objects to be segmented from backgrounds.

Second, these results provide a unique comparison with ANN models in artificial intelligence. With controlled rearing (as with ANN models) we have a record of the full set of training data available for learning, so we can directly compare the learning abilities of ANN models and biological systems. ANN models are "data hungry" (Marcus, 2018), typically requiring thousands to millions of labeled training images to build invariant object representations. Conversely, newborn chicks can build invariant object representations from visual input of a single object on a single background. Thus, newborn brains appear to be far more efficient at building object representations than ANN models. This efficiency may be due, in part, to a motion-based parsing mechanism that restricts learning to moving objects, rather than the whole visual field. Consistent with this hypothesis, researchers have shown that modifying ANN models to operate over motion cues substantially improves the networks' performance, especially when training data are limited (Goel et al., 2018; Pathak et al., 2017). For ANN models to be biologically plausible (from a developmental perspective), they will need to learn from motion cues during the earliest stages of object perception.

More generally, we theorize that newborn visual systems have an object-based inductive bias at the onset of vision. Inductive biases permit rapid learning by constraining the space of input-output functions that can be learned. We speculate that an object-based inductive bias predisposes newborn visual systems to process visual input in terms of bounded objects that persist over space and time. Accordingly, an object-based inductive bias would allow animals to solve challenging object perception tasks without extensive object experience.

If newborn animals have an object-based inductive bias, then we should observe two patterns. First, newborn animals should succeed in object perception tasks early in development, without extensive experience with objects. Second, an object-based inductive bias should constrain the input-output functions that newborn animals can learn. Specifically, newborn animals should learn best from objects that have the same spatiotemporal properties as objects in the real world.

Evidence for these two patterns comes from automated controlled-rearing experiments with newborn chicks. First, these studies show that newborn chicks are rapid learners: chicks are capable of one-shot object parsing (as shown here), one-shot visual binding (Wood, 2014), one-shot view-invariant object recognition (Wood, 2013, 2015; Wood & Wood, 2020), and one-shot object permanence (Prasad et al., 2019). Second, these studies show that the development of object perception requires experience with natural object motion. As we show here, the development of object parsing requires experience with moving objects, and in prior work, we discovered that the development of visual binding (Wood, 2016), view-invariant object recognition (Wood & Wood, 2016, 2018), and object permanence (Prasad et al., 2019) require experience with objects moving slowly and smoothly over time, akin to natural objects. Thus, automated

controlled-rearing studies provide strong evidence for an object-based inductive bias in newborn brains. Prenatal developmental programs build an object-based inductive bias that allows animals to learn rapidly about objects and generalize far beyond the visual experiences that they acquire during postnatal development.

Limitations of Our Findings and Directions for Future Research

Our results raise additional questions for future research. First, how do newborn brains build an object-based inductive bias during prenatal development? One possibility is that developing brains generate their own "visual" experiences for learning. For instance, waves of spontaneous activity in the retina are present long before eye opening (Katz & Shatz, 1996) and are necessary for the normal development of the visual cortex (Chou et al., 2013). These waves allow the visual cortex of newborn animals to develop properties similar to those found in the adult cortex (Cang et al., 2005), potentially because retinal waves share many characteristics with real-world objects. Both are bounded, cohesive units that move on continuous spatiotemporal paths (Albert et al., 2008). These prenatal "object-like" experiences provide early training data for building an object-based inductive bias in developing brains. Since these retinal waves have been discovered in both birds and mammals (Wong, 1999), we suspect that this prenatal developmental program endows most newborn animals with an object-based inductive bias.

Second, what role does attention play in motion-based object parsing? The core problem of object parsing is determining which features are part of objects versus backgrounds. We suspect that attention will be heavily involved in this process, with motion cues driving attention toward object features over background features. We also suspect that other temporal cues that drive attention toward object features (such as blinking or color changes) could be used to parse objects from backgrounds. Of course, in the real world, motion is typically the most reliable cue for determining object boundaries (Spelke, 1990), so we focused on motion as a starting point in our investigation of object parsing.

It would be interesting for future studies to explore whether motion holds a privileged status in visual development compared with other attention-grabbing cues. From an optics perspective, motion provides valuable information. For instance, reverse projections from static visual inputs typically produce inaccurate representations of three-dimensional (3D) space, whereas optic flow resulting from object (or observer) motion provides a successive combination of inputs for building accurate 3D models of the world (Gibson, 1966; Johansson, 1970). Accordingly, newborn brains might be predisposed to orient toward motion cues to drive rapid visual development.

Third, what specific mechanisms underlie motion-based object parsing? Our results are consistent with two (nonmutually exclusive) possibilities. First, newborn brains might selectively encode moving objects by filtering information through a motion-detection bottleneck during early visual processing. Evidence for this possibility comes from studies showing that motion detection circuits can be hardwired into the retina (Kim et al., 2015; Zhang et al., 2012), suggesting that newborn brains could selectively encode moving objects during the earliest stages of visual learning. Second, newborn brains might use unsupervised temporal learning (UTL) mechanisms to parse and recognize objects. When an object moves, object features transition rapidly across background features, providing a useful set of training data for UTL mechanisms (e.g., predictive coding). By encoding which features move together versus separately, UTL mechanisms could (in principle) learn to detect object boundaries, even after experience with just a single object. Notably, both of these accounts depend heavily on motion cues to drive the rapid emergence of object parsing and recognition.

Ultimately, a deep understanding of the origins and computational foundations of object perception will require building task-performing models of visual development. Task-performing models can simulate the complex interaction between newborn brains and visual experience. Researchers across computational neuroscience (Kriegeskorte & Douglas, 2018), cognitive science (Newell, 1973), and developmental psychology (Dupoux, 2018) have long argued that task-performing models are essential for gaining a mechanistic understanding of how brains process information. Unlike verbal models, task-performing models can serve as formal hypotheses that operate over raw sensory inputs and make behavioral decisions; as such, these models make unequivocal predictions that can be falsified by empirical data (Hasson et al., 2020; Richards et al., 2019; Schrimpf et al., 2020).

Automated controlled-rearing studies will be essential for building accurate task-performing models of visual development because these studies produce precise measurements of behavior while simultaneously providing strict control over all of the animal's experiences. Both biological and artificial neural networks are heavily shaped by their experiences (training data). Accordingly, by providing the same training data to newborn animals and task-performing models, we can directly compare their learning abilities. For instance, to account for the development of object parsing, a task-performing model must learn to parse and recognize objects from sparse training data (e.g., input of a single object on a single background). By collecting precise benchmarks of visual learning in controlled environments, our study takes an important step toward building task-performing models of the origins and development of object perception.

Conclusion

Our study demonstrates that newborn animals are capable of one-shot object parsing. By leveraging motion cues, newborn chicks can parse objects from natural backgrounds and build invariant representations that generalize across novel backgrounds and novel viewpoints. As such, we argue that prenatal developmental programs equip newborn brains with an object-based inductive bias that allows animals to generalize far beyond the visual experiences acquired during postnatal development. These findings illuminate the origins of a foundational visual ability and provide benchmarks for building task-performing models of visual development.

These results also compliment a growing body of work showing that humans and other animals are capable of one-shot learning (e.g., Lake et al., 2015; Landau et al., 1988; Xu & Tenenbaum, 2007). Our study indicates that some forms of one-shot learning do not require extensive postnatal experience to develop. One-shot object parsing scaffolds object perception during the earliest stages of visual learning.

References

Albert, M. V., Schnabel, A., & Field, D. J. (2008). Innate visual learning through spontaneous activity patterns. *PLoS Computational Biology*, *4*(8), e1000137. https://doi.org/10.1371/journal.pcbi.1000137

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), 1404–1412. https://doi.org/10.1038/nn.4635

Arterberry, M. E., & Yonas, A. (2000). Perception of three-dimensional shape specified by optic flow by 8-week-old infants. *Perception & Psychophysics*, *62*(3), 550–556. https://doi.org/10.3758/BF03212106

Bashivan, P., Kar, K., & DiCarlo, J. J. (2019). Neural population control via deep image synthesis. *Science*, *364*(6439), eaav9436. https://doi.org/10.1126/science.aav9436

Bremner, J. G., Slater, A. M., & Johnson, S. P. (2015). Perception of object persistence: The origins of object permanence in infancy. *Child Development Perspectives*, *9*(1), 7–13. https://doi.org/10.1111/cdep.12098

Bushnell, E. W., & Boudreau, J. P. (1993). Motor development and the mind: The potential role of motor abilities as a determinant of aspects of perceptual development. *Child Development*, *64*(4), 1005–1021. https://doi.org/10.2307/1131323

Cang, J., Rentería, R. C., Kaneko, M., Liu, X., Copenhagen, D. R., & Stryker, M. P. (2005). Development of precise maps in visual cortex requires patterned spontaneous activity in the retina. *Neuron*, *48*(5), 797–809. https://doi.org/10.1016/j.neuron.2005.09.015

Chou, T.-H., Park, K. K., Luo, X., & Porciatti, V. (2013). Retrograde signaling in the optic nerve is necessary for electrical responsiveness of retinal ganglion cells. *Investigative Ophthalmology & Visual Science*, *54*(2), 1236–1243. https://doi.org/10.1167/iovs.12-11188

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341.

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

Dupoux, E. (2018). Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition*, *173*, 43–59. https://doi.org/10.1016/j.cognition.2017.11.008

Fiser, J., Scholl, B. J., & Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychonomic Bulletin & Review*, *14*(1), 173–178. https://doi.org/10.3758/BF03194046

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin.

Goel, V., Weng, J., & Poupart, P. (2018). Unsupervised video object segmentation for deep reinforcement learning. *Advances in Neural Information Processing Systems*, *31*, 5683–5694.

Hasson, U., Nastase, S. A., & Goldstein, A. (2020). Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron*, *105*(3), 416–434. https://doi.org/10.1016/j.neuron.2019.12.002

Horn, G. (2004). Pathways of the past: The imprint of memory. *Nature Reviews Neuroscience*, *5*(2), 108–120. https://doi.org/10.1038/nrn1324

Johansson, G. (1970). "On theories for visual space perception: A letter to Gibson." *Scandinavian Journal of Psychology*, *11*(2), 67–74.

Johnson, S. P., Bremner, J. G., Slater, A. M., Mason, U. C., & Foster, K. (2002). Young infants' perception of unity and form in occlusion displays. *Journal of Experimental Child Psychology*, *81*(3), 358–374. https://doi.org/10.1006/jecp.2002.2657

Katz, L. C., & Shatz, C. J. (1996). Synaptic activity and the construction of cortical circuits. *Science*, *274*(5290), 1133–1138. https://doi.org/10.1126/science.274.5290.1133

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*(4), 483–524. https://doi.org/10.1016/0010-0285(83)90017-8

Kellman, P. J., Spelke, E. S., & Short, K. R. (1986). Infant perception of object unity from translatory motion in depth and vertical translation. *Child Development*, *57*(1), 72–86. https://doi.org/10.2307/1130639

Kim, T., Soto, F., & Kerschensteiner, D. (2015). An excitatory amacrine cell detects object motion and provides feature-selective input to ganglion cells in the mouse retina. *Elife*, *4*, e08025.

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. https://doi.org/10.1016/S0010-0277(02)00004-5

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148–1160. https://doi.org/10.1038/s41593-018-0210-5

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012, December). *ImageNet classification with deep convolutional neural networks* [Paper presentation]. 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV.

Krogh, L., Vlach, H. A., & Johnson, S. P. (2012). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology*, *3*, 598. https://doi.org/10.3389/fpsyg.2012.00598

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338. https://doi.org/10.1126/science.aab3050

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299–321. https://doi.org/10.1016/0885-2014(88)90014-7

Lea, S. E. G., Slater, A. M., & Ryan, C. M. E. (1996). Perception of object unity in chicks: A comparison with the human infant. *Infant Behavior and Development*, *19*(4), 501–504. https://doi.org/10.1016/S0163-6383(96)90010-7

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

Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv*. https://arxiv.org/abs/1801.00631

Mordvintsev, A., Olah, C., & Tyka, M. (2015). Inceptionism: Going deeper into neural networks. https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). Academic Press. https://doi.org/10.1016/B978-0-12-170150-5.50012-3

Ostrovsky, Y., Meyers, E., Ganesh, S., Mathur, U., & Sinha, P. (2009). Visual parsing after recovery from blindness. *Psychological Science*, *20*(12), 1484–1491. https://doi.org/10.1111/j.1467-9280.2009.02471.x

Owsley, C. (1983). The role of motion in infants' perception of solid shape. *Perception*, *12*(6), 707–717. https://doi.org/10.1068/p120707

Pathak, D., Girshick, R., Dollár, P., Darrell, T., & Hariharan, B. (2017). Learning features by watching objects move. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2701–2710). https://openaccess.thecvf.com/content_cvpr_2017/html/Pathak_Learning_Features_by_CVPR_2017_paper.html

Prasad, A., Wood, S. M. W., & Wood, J. N. (2019). Using automated controlled rearing to explore the origins of object permanence. *Developmental Science*. Advance online publication. https://doi.org/10.1111/desc.12796

Regolin, L., & Vallortigara, G. (1995). Perception of partly occluded objects by young chicks. *Perception & Psychophysics*, *57*(7), 971–976. https://doi.org/10.3758/BF03205456

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Ganguli, S., Gillon, C. J., Hafner, D., Kepecs, A., Kriegeskorte, N., Latham, P., Lindsay, G. W., Miller, K. D., Naud, R., Pack, C. C., . . . Kording, K. P. (2019). A deep learning framework for neuroscience. *Nature Neuroscience*, *22*(11), 1761–1770. https://doi.org/10.1038/s41593-019-0520-2

Rosa-Salva, O., Regolin, L., & Vallortigara, G. (2010). Faces are special for newly hatched chicks: Evidence for inborn domain-specific

mechanisms underlying spontaneous preferences for face-like stimuli. *Developmental Science*, *13*(4), 565–577. https://doi.org/10.1111/j.1467-7687.2009.00914.x

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*(1), 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Schrimpf, M., Kubilius, J., Lee, M. J., Murty, N. A. R., Ajemian, R., & DiCarlo, J. J. (2020). Integrative benchmarking to advance neurally mechanistic models of human intelligence. *Neuron*, *108*(3), 413–423. https://doi.org/10.1016/j.neuron.2020.07.040

Slater, A., Johnson, S. P., Brown, E., & Badenoch, M. (1996). Newborn infant's perception of partly occluded objects. *Infant Behavior and Development*, *19*(1), 145–148. https://doi.org/10.1016/S0163-6383(96)90052-1

Slater, A., Morison, V., Somers, M., Mattock, A., Brown, E., & Taylor, D. (1990). Newborn and older infants' perception of partly occluded objects. *Infant Behavior and Development*, *13*(1), 33–49. https://doi.org/10.1016/0163-6383(90)90004-R

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word-referent learning. *Trends in Cognitive Sciences*, *18*(5), 251–258. https://doi.org/10.1016/j.tics.2014.02.007

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*(1), 29–56. https://doi.org/10.1207/s15516709cog1401_3

Turk-Browne, N. B., Junge, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Valenza, E., & Bulf, H. (2011). Early development of object unity: Evidence for perceptual completion in newborns. *Developmental Science*, *14*(4), 799–808. https://doi.org/10.1111/j.1467-7687.2010.01026.x

Valenza, E., Zulian, L., & Leo, I. (2005). The role of perceptual skills in newborns' perception of partly occluded objects. *Infancy*, *8*(1), 1–20. https://doi.org/10.1207/s15327078in0801_1

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

Wong, R. O. (1999). Retinal waves and visual system development. *Annual Review of Neuroscience*, *22*(1), 29–47. https://doi.org/10.1146/annurev.neuro.22.1.29

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. https://doi.org/10.1073/pnas.1308246110

Wood, J. N. (2014). Newly hatched chicks solve the visual binding problem. *Psychological Science*, *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. (2017). Spontaneous preference for slowly moving objects in visually naïve animals. *Open Mind*, *1*(2), 111–122. https://doi.org/10.1162/OPMI_a_00012

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. https://doi.org/10.1098/rspb.2016.0166

Wood, J. N., & Wood, S. M. W. (2018). A smoothness constraint on the development of abstract object representations. *Cognitive Science*. Advance online publication. https://doi.org/10.1111/cogs.12595

Wood, J. N., & Wood, S. M. W. (2020). One-shot learning of view-invariant object representations in newborn chicks. *Cognition*. Advance online publication. https://doi.org/10.1016/j.cognition.2020.104192

Wood, S. M. W., & Wood, J. N. (2015). A chicken model for studying the emergence of invariant object recognition. *Frontiers in Neural Circuits*, *9*, 7. https://doi.org/10.3389/fncir.2015.00007

Wood, S. M., & 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. https://doi.org/10.1016/j.infbeh.2019.101329

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, *30*(2), 111–153. https://doi.org/10.1006/cogp.1996.0005

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245

Yamins, D. K. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365. https://doi.org/10.1038/nn.4244

Zhang, Y., Kim, I. J., Sanes, J. R., & Meister, M. (2012). The most numerous ganglion cell type of the mouse retina is a selective feature detector. *Proceedings of the National Academy of Sciences*, *109*(36), E2391–E2398.

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 of the United States of America*, *106*(21), 8748–8753. https://doi.org/10.1073/pnas.0811583106