



PAPER

Characterizing the information content of a newly hatched chick's first visual object representation

Justin N. Wood

Department of Psychology, University of Southern California, USA

Abstract

*How does object recognition emerge in the newborn brain? To address this question, I examined the information content of the first visual object representation built by newly hatched chicks (*Gallus gallus*). In their first week of life, chicks were raised in controlled-rearing chambers that contained a single virtual object rotating around a single axis. In their second week of life, I tested whether subjects had encoded information about the identity and viewpoint of the virtual object. The results showed that chicks built object representations that contained both object identity information and view-specific information. However, there was a trade-off between these two types of information: subjects who were more sensitive to identity information were less sensitive to view-specific information, and vice versa. This pattern of results is predicted by iterative, hierarchically organized visual processing machinery, the machinery that supports object recognition in adult primates. More generally, this study shows that invariant object recognition is a core cognitive ability that can be operational at the onset of visual object experience.*

Research highlights

- This study measured the information content of the first visual object representation built by newly hatched chicks.
- The results showed that (1) the first visual object representation built by chicks contains both object identity information and view-specific information, and (2) there is a trade-off between these two types of information.
- These results indicate that newly hatched chicks build similar object representations as adult primates.
- More generally, these results show that invariant object recognition is a core cognitive ability that is operational at the onset of visual object experience.

Introduction

Humans are remarkably good at recognizing objects. Although any given object can cast an infinite number of different images on the retina (due to variation in viewpoint, scale, lighting conditions, etc.), we recognize objects quickly and accurately in a manner that is largely

invariant to these transformations (DiCarlo, Zoccolan & Rust, 2012). To date, however, the origins of invariant object recognition are poorly understood. Due to the challenges associated with testing newborns experimentally, it has generally not been possible to study the initial state of object recognition (i.e. the state of object recognition machinery at the onset of visual object experience).

Two major limitations have hindered the study of newborns' object recognition abilities. First, it is typically possible to collect just a few test trials from each newborn subject. This limitation has prevented researchers from obtaining precise measurements of newborns' first visual object representations. Second, most newborn animals cannot be raised in controlled environments from birth. This limitation has prevented researchers from characterizing how specific visual experiences shape the initial state of object recognition.

To overcome these limitations, I used a *complete data* controlled-rearing technique (Wood, 2013) with a newborn¹ animal model – the domestic chicken (*Gallus gallus*).

¹ The term 'newborn' is used to refer to an animal at the beginning of the post-embryonic phase of their life cycle.

Chickens were used as an animal model for four reasons: (1) newly hatched chicks can learn to recognize objects (e.g. Vallortigara, 2012; Wood, 2013); (2) newly hatched chicks can be raised in highly controlled environments from the onset of visual object experience (e.g. Wood, 2013), (3) newly hatched chicks imprint to objects (e.g. Bateson, 2000; Horn, 2004; Vallortigara, 2012), providing a naturally occurring behavior that can be used to test object recognition abilities without training, and (4) chickens and humans use nearly identical neural circuits to process sensory information (reviewed by Karten, 2013). Thus, controlled-rearing experiments with newly hatched chicks can be used to study the initial state of object recognition. The *complete data* controlled-rearing technique involves raising newly hatched chicks for several weeks within controlled-rearing chambers that provide complete control over all visual object experiences. We use the term *complete data* because the chambers record *all* of the subjects' behavior (9 samples/second, 24 hours/day, 7 days/week). This approach makes it possible to measure a newborn's first visual object representation with high precision, because each subject can be presented with hundreds of test trials during the experiment (Wood, 2013).

In a previous study, Wood (2013) used this controlled-rearing method to investigate the origins of invariant object recognition. In their first week of life, chicks were raised in environments that contained a single object that could only be seen from a limited 60° viewpoint range. In their second week of life, subjects were tested on their ability to recognize that object from novel viewpoints. The majority of the subjects successfully recognized the object across the viewpoint changes, which shows that newborn animals can build a viewpoint-invariant representation of the first object they see in their life. Thus, at the onset of visual object experience, object recognition generalizes far beyond the available visual input.

To examine the mechanisms underlying this ability, in the present study I measured the informational content of the first visual object representation built by newly hatched chicks. Specifically, I tested whether chicks build object representations that contain two types of information: (1) object identity information; and (2) view-specific information (i.e. information about the specific visual features used to build the representation). Building an invariant object representation requires transforming patterns of retinal activity (view-specific information) into an abstract representation that is tolerant to retinal image changes and selective for a particular object (identity information). Thus, measuring sensitivity to identity and view-specific information reveals how successful the subject has been in building an invariant representation.

In the first week of life (the input phase), subjects were raised in environments that contained a single virtual object rotating around a single axis. Chicks imprint to moving objects experienced in the first few days of life, so subjects were expected to imprint to this object. In the second week of life (the test phase), I measured the informational content of the object representation built by each subject, by using an automated two-alternative forced-choice testing procedure.

Methods

Subjects

Sixteen domestic chickens (*Gallus gallus*) of unknown sex were tested. No subjects were excluded from the analyses. The eggs were incubated in darkness to ensure that no visual input would reach the subject through the shell. After hatching, the chicks were moved from the incubation room to the controlled-rearing chambers in darkness. Each chick was raised singly within its own chamber. All care of the subjects (e.g. replenishment of food and water) was performed in darkness with the aid of night vision goggles.

Controlled-rearing chambers

Subjects were raised within controlled-rearing chambers that measured 66 cm (length) × 42 cm (width) × 69 cm (height). For a picture of the chambers, see Figure 1 in Wood (2013). The chambers contained extended surfaces only and were constructed from white, high-density plastic. The virtual objects were projected onto two display walls (19" liquid crystal display monitors, 1440 × 900 pixel resolution) situated on opposite sides of the chamber. Food and water were available *ad libitum* within transparent holes in the ground that measured 66 cm (length) × 2.5 cm (width) × 2.7 cm (height). Grain was used as food because it does not behave like an object (i.e. grain does not maintain a rigid, bounded shape). The floors of the chambers consisted of black wire mesh suspended over a black surface.

Subjects' movements were tracked continuously by micro-cameras embedded in the ceilings of the chambers and analyzed with software (Ethovision XT) that calculated the amount of time chicks spent within zones (22 cm × 42 cm) next to the left and right display walls. The entire data collection process was automated, which allowed a large number of test trials to be collected from each subject.

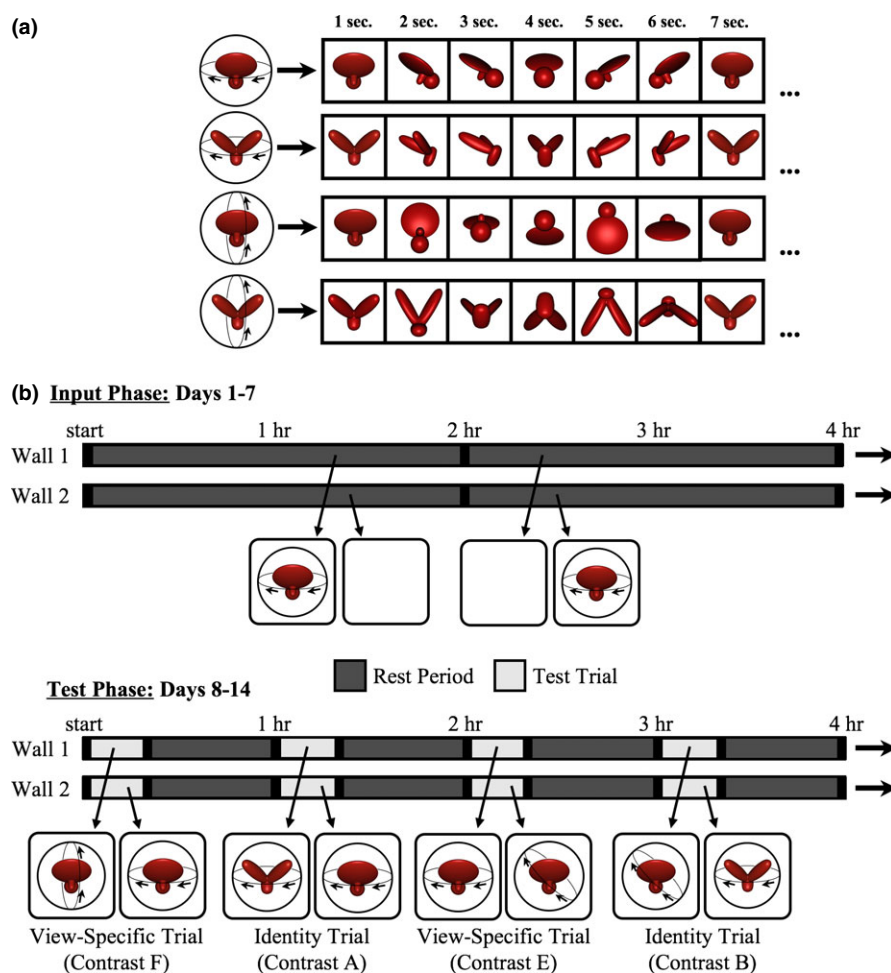


Figure 1 (a) The four object animations presented during the input phase. Each subject was shown only one of the four animations. The objects moved smoothly, completing a full rotation every 6 seconds. (b) (Top) A schematic showing a 4-hour presentation schedule of the virtual object during the input phase. (Bottom) A schematic showing a 4-hour presentation schedule of the virtual objects during the test phase.

Procedure

The input phase

In their first week of life (the input phase), subjects were raised in controlled-rearing chambers that contained a single virtual object rotating around a single axis. There were two possible imprinting objects and two possible axes of rotation (Figure 1a). The object moved continuously, completing a full rotation every 6 seconds. The object rotated 2.5° between each visual frame at 24 frames/second. On average, the objects measured 8 cm (length) \times 7 cm (height) and were displayed on a uniform white background. The object appeared for an equal amount of time on the left and right display wall, switching walls every 2 hours (following a 1-minute period of darkness, Figure 1b). The objects were modeled after those used in

previous studies that tested for invariant object recognition in adult rats (Alemi-Neissi, Rosselli & Zoccolan, 2013; Zoccolan, Oertelt, DiCarlo & Cox, 2009).

The test phase

In their second week of life (the test phase), subjects received 168 test trials (24 test trials per day). During the test trials, I projected an object animation onto each display wall and measured the amount of time subjects spent in proximity to each object (Figure 1b). Subjects were expected to spend a greater proportion of time in proximity to the object that they perceived to be their imprinted object. Each test trial lasted 20 minutes. The test trials were separated from one another by 40-minute rest periods. During the rest periods, the animation from

the input phase was projected onto one display wall and a white screen was projected onto the other display wall. Rest periods were included in the design for two reasons. First, they provided periods of rest because the subjects

were able 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 each subject generally wanted to spend

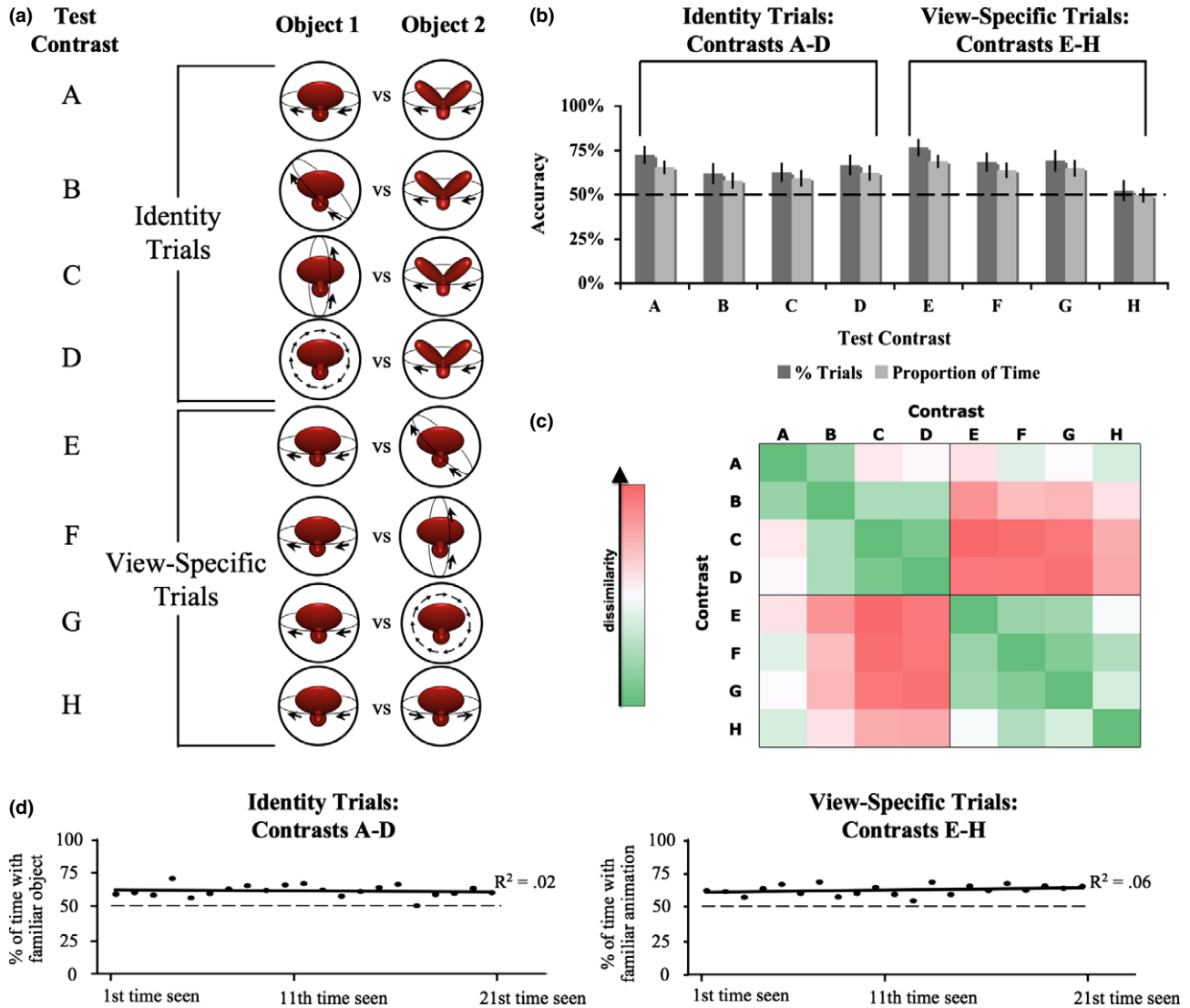


Figure 2 (a) The eight test contrasts presented during the test phase. The specific object animations depicted here were presented to the subjects who were imprinted to the first animation shown in Figure 1a. (b) Subjects' mean performance on the eight test contrasts. The dark gray bars denote the percentage of correct trials, and the light gray bars denote the proportion of time subjects spent with the familiar object on the Identity Trials and the familiar animation on the View-Specific Trials. Chance performance was 50%. Error bars denote standard error. (c) The representational dissimilarity matrix. Each cell of the matrix compares the behavioral response elicited by two test contrasts. Low dissimilarities (green colors) indicate that subjects exhibited similar behavior when presented with those test contrasts. High dissimilarities (red colors) indicate that subjects exhibited the opposite behavior when presented with those test contrasts. The dissimilarity between two test contrasts was measured as 1 minus the Pearson correlation between the set of performance values obtained across all subjects for the two different contrasts. The color scale reflects the full range of correlation values. (d) Change over time results. The graphs illustrate group mean performance for the Identity Trials and View-Specific Trials, computed for the first, second, third, etc., presentation of the test contrasts. The y-axis indicates the proportion of time subjects spent with the familiar object on the Identity Trials and the familiar animation on the View-Specific Trials. Chance performance was 50%.

in proximity to their imprinted object (see Results section for more details).

Subjects were presented with two types of test trials, each containing four different test contrasts (Figure 2a).

Identity Trials (Contrasts A–D). The imprinted object was paired with an unfamiliar object that had a similar size, color, and motion speed as the imprinted object. In Contrast A, the imprinted object and the unfamiliar object rotated around the same axis. In Contrasts B–D, the imprinted object rotated around a novel axis. When an object rotates around a novel axis, it projects novel retinal images to the subject (Figure 1a). Thus, to recognize their imprinted object, subjects needed to generalize across large, novel, and complex changes in the retinal images produced by the object.

View-Specific Trials (Contrasts E–H). One display wall showed the imprinted object rotating around the familiar axis, while the other display wall showed the imprinted object rotating around a novel axis. The imprinted object rotated around an axis tilted 45° (Contrast E), a perpendicular axis (Contrast F), the picture plane (Contrast G), or in the opposite direction (Contrast H). Contrasts E–G presented novel images of the imprinted object to the subject, whereas Contrast H presented familiar images of the imprinted object but in the opposite order from the input animation. Thus, Contrasts E–G tested whether subjects could distinguish between familiar versus novel viewpoints, whereas Contrast H tested whether subjects recognize objects by using information about the temporal order of object views. If chicks build object representations that contain information about the specific retinal input used to build the representation, then they should prefer the familiar animation over the unfamiliar animation.

Analysis of object stimuli

To investigate whether subjects needed invariant representations to succeed on the Identity Trials, I quantified the similarity between the input animations and the test animations in two ways. First, I measured the overall brightness of each animation by computing the sum of the pixel intensities for each frame in each animation. As shown in Figure 3a, the unfamiliar object was brighter than the imprinted object for two of the groups (*top* panels) and darker than the imprinted object for the other two groups (*bottom* panels). Thus, subjects could not succeed in this task simply by having a preference for brighter objects or darker objects. Another possibility is that subjects might encode the overall brightness of the imprinted animation and

perceive objects as similar to the imprinted object when those objects have a similar overall brightness. This possibility is difficult to assess in the current study because the design only included four animations of the imprinted object and those animations had similar brightness values (Figure 3a). Accordingly, I performed new analyses on the stimuli and data from Wood (2013). Wood (2013) used the same two virtual objects as the present study, but presented the objects from 12 different 60° viewpoint ranges in the test phase. When objects are presented from different viewpoint ranges they produce different brightness values. The brightness values and results for each viewpoint range are shown in Figure 3b. Importantly, subjects' object recognition performance did not vary as a function of the brightness similarity between the test animation of the imprinted object and the input animation of the imprinted object (Pearson correlation: $r = -.06$, $p = .67$). Thus, chicks do not simply use overall brightness as a low-level cue to distinguish between objects.

Second, I computed the amount of image variation between the input animations and the test animations (from a pixel-level perspective). This was achieved by (1) converting each 6-second animation into a sequence of images (at 24 frames/second), (2) measuring the brightness level of each pixel in each image, and (3) comparing each image in the test animation to each image in the input animation (i.e. by comparing the brightness level of each corresponding pixel across the images). These comparisons were then plotted as a time series. Each point along the time series reflects the mean pixel-level difference between one image in the test animation and all of the images in the input animation. The time series for each test animation is shown in Figure 4. This analysis confirmed that changing the axis of rotation of the imprinted object produced large and complex changes in the pixel-level similarity between the test animation and the input animation. At some points in time the within-object image difference (i.e. the pixel-level difference between the test animation of the imprinted object and the input animation of the imprinted object) was greater than the between-object image difference (i.e. the pixel-level difference between the test animation of the unfamiliar object and the input animation of the imprinted object), whereas at other points in time the between-object image difference was greater than the within-object image difference. Thus, a representation that only contained information about the object's pixel-level appearance (i.e. a pictorial-like representation) would be insufficient for reliably distinguishing the imprinted object from the unfamiliar object. Rather, to succeed in this task, subjects needed to build

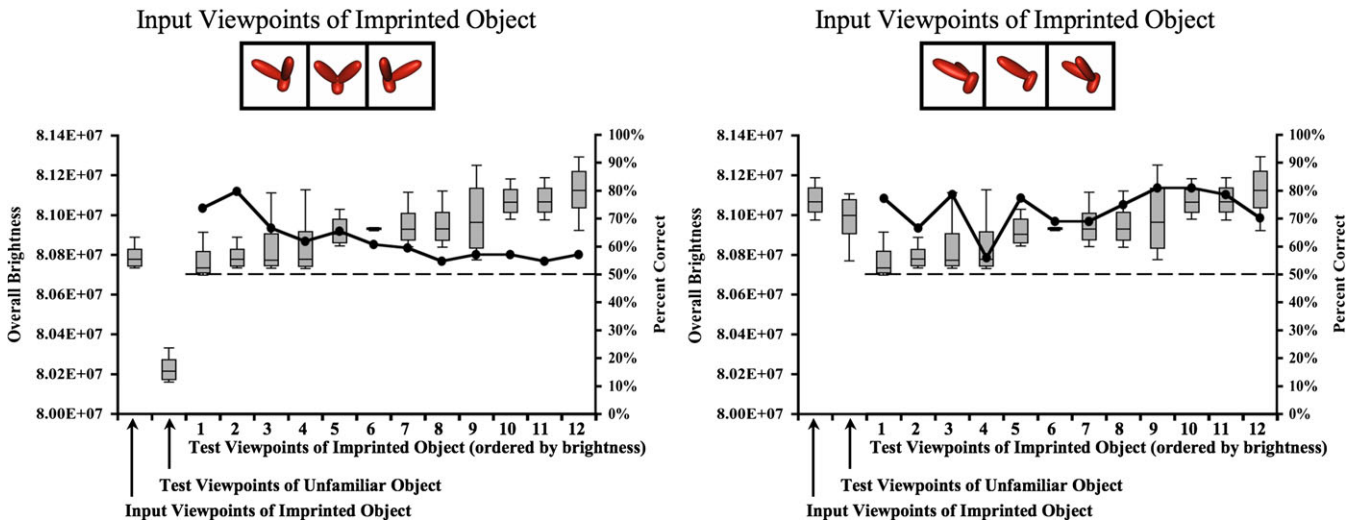
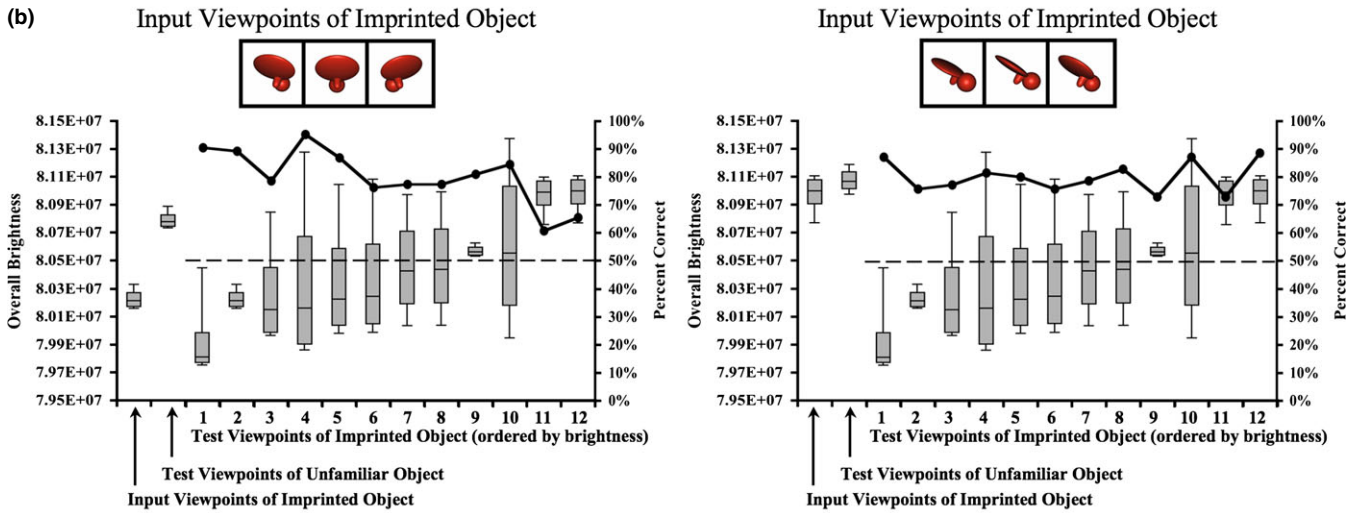
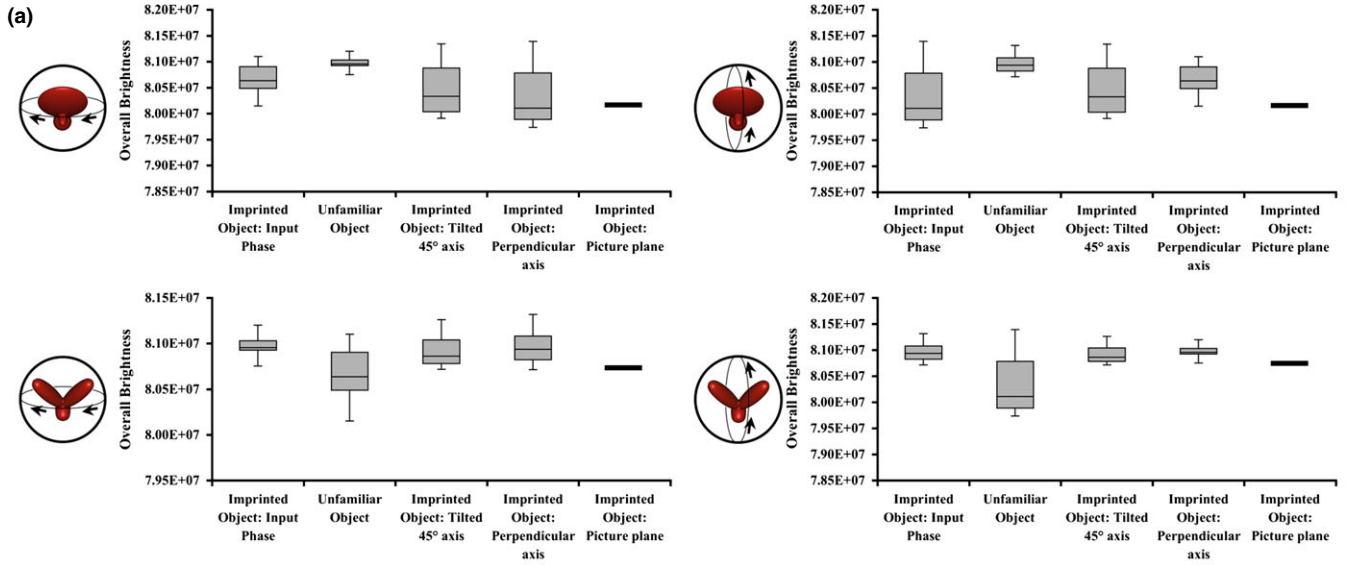


Figure 3 (a) The brightness levels of the object animations. The bottom and top of each box plot depict the first and third quartile brightness levels, the band inside the box depicts the median brightness level, and the whiskers depict the minimum and maximum brightness levels. (b) New analyses of the stimuli and results from Experiments 1 and 2 in Wood (2013), which used the same two virtual objects as the present study. Chicks were imprinted to a single virtual object rotating through a 60° viewpoint range, and then tested on their ability to recognize that object across 12 different 60° viewpoint ranges. The box plots show the brightness levels of each viewpoint range, and the line graphs show subjects' object recognition performance across the viewpoint ranges. Subjects' object recognition performance did not vary as a function of the brightness similarity between the test animation of the imprinted object and the input animation of the imprinted object (Pearson correlation: $r = -.06$, $p = .67$). Thus, chicks do not simply use overall brightness as a low-level cue to recognize objects.

invariant object representations, such as those found in high-level cortical visual areas.

Results

For each test trial, I computed the amount of time subjects spent within zones located next to the left and right display walls. The Identity Trials were scored as 'correct' when subjects spent a greater proportion of time with the imprinted object compared to the unfamiliar object, and 'incorrect' when they spent a greater proportion of time with the unfamiliar object compared to the imprinted object. The View-Specific Trials were scored as 'correct' when subjects spent a greater proportion of time with the familiar animation compared to the unfamiliar animation, and 'incorrect' when they spent a greater proportion of time with the unfamiliar animation compared to the familiar animation. These responses were then analyzed with hierarchical Bayesian methods that provided probabilistic estimates of recognition performance for each subject and the overall group (Kruschke, 2011).

The Bayesian analysis first required specifying a prior distribution. I used a conservative prior consisting of one correct trial and one incorrect trial. This is a conservative prior because it consists of just two trials at chance level (50%). The prior distribution also includes a parameter – kappa – that represents the consistency across subjects. I used a uniform density kappa (Gelman, 2006) that ranged from 0.000001 (i.e. very little consistency across subjects) to the maximum reasonable kappa. The maximum reasonable kappa was estimated from subjects' performance during the rest periods in the test phase. The rest periods were expected to produce the greatest consistency across subjects because they presented the easiest choice: subjects chose whether to spend time with the input animation versus a white screen. Critically, all of these prior distribution parameters were broad and vague, thereby expressing great prior uncertainty in the values of the parameters. This type of prior has minimal influence on the estimates of success and is quickly overwhelmed by even a modest amount of data with Bayesian parameter estimation (Kruschke, 2011). The model used Markov

Chain Monte Carlo sampling to estimate the probability that performance was above chance level.

Overall performance

Figure 2b shows subjects' average performance on each test contrast. On the Identity Trials (Contrasts A–D), subjects successfully distinguished their imprinted object from the unfamiliar object on 66% (SEM = 4%) of the test trials. The probability that performance was above chance was greater than 99.9%. Across the 16 subjects, the probability that performance was above chance was 99% or greater for 10 of the subjects, between 97 and 99% for two of the subjects, and 72% for another subject. The probability that performance was above chance was below 50% for the remaining three subjects.

On the View-Specific Trials (Contrasts E–H), subjects successfully distinguished the familiar animation from the unfamiliar animation on 67% (SEM = 4%) of the test trials. The probability that performance was above chance was greater than 99.9%. Across the 16 subjects, the probability that performance was above chance was 99% or greater for nine of the subjects, and 97%, 86%, and 80% for three other subjects. The probability that performance was above chance was below 70% for the remaining four subjects.

Overall, these newborn subjects were sensitive to both identity information and view-specific information. However, there was considerable variation in subjects' object recognition performance. To examine this variation directly, I constructed a representational dissimilarity matrix (RDM) that revealed the relationship between the different types of information contained within the subjects' representations.

Representational dissimilarity matrix

An RDM characterizes the information contained within a representation, revealing which distinctions matter to the representation (Kriegeskorte, Mur & Bandettini, 2008). To construct the RDM, I computed the dissimilarity in performance for each pair of test contrasts. The dissimilarity between two test contrasts was measured as 1 minus the

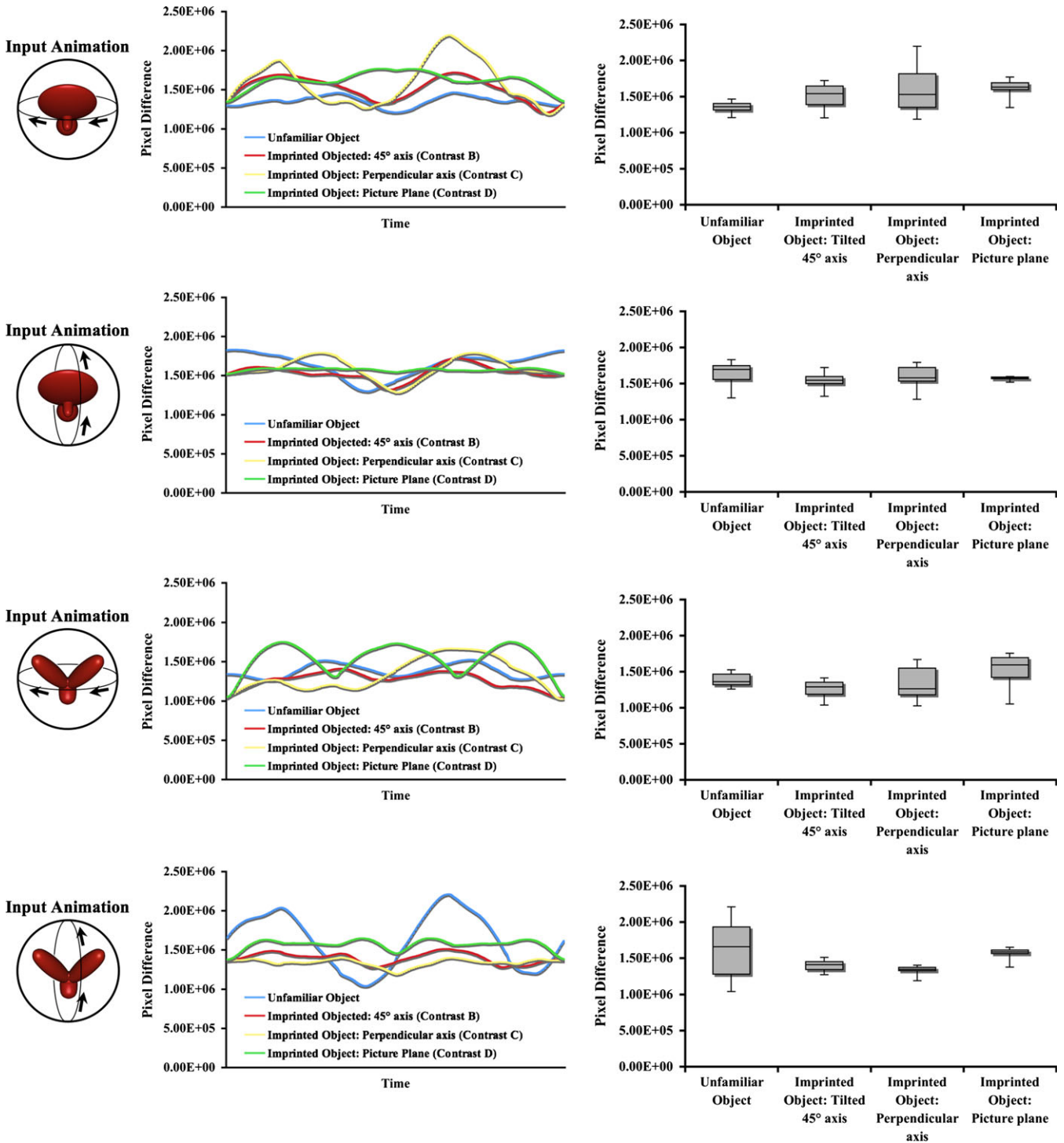


Figure 4 Results of the pixel-by-pixel comparisons of the virtual objects presented in the input phase and the virtual objects presented in the test phase. Each point along a time series (left panels) reflects the mean pixel-level difference between one image in the test object animation and all of the images in the input object animation. The box plots (right panels) show the overall pixel-level differences between the test animations and the input animations.

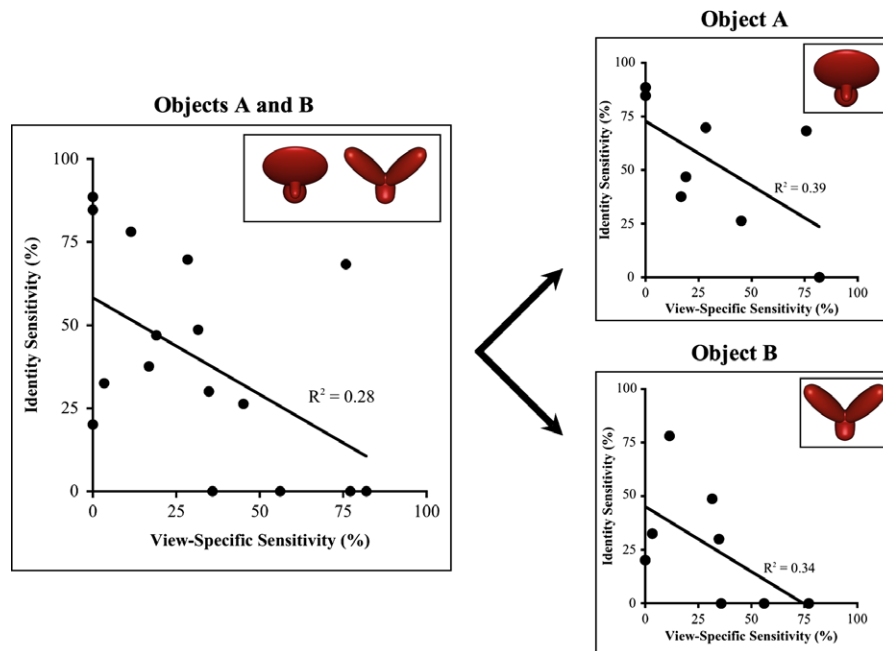


Figure 5 The abstraction space. Each dot depicts the object representation built by one subject. The y-axis reflects the representation's sensitivity to identity information (measured from Contrasts A–D) and the x-axis reflects the representation's sensitivity to view-specific information (measured from Contrasts E–H). The abstraction space shown on the left includes the representations built by all of the subjects, whereas the smaller abstraction spaces shown on the right include the representations built by the subjects who were imprinted to each of the two objects.

Pearson correlation between the set of performance values obtained across all subjects for the two different contrasts (with 0 indicating a perfect positive correlation, 1 indicating no correlation, and 2 indicating a perfect negative correlation). The dissimilarities for all pairs of test contrasts were then assembled in the RDM, with each cell in the matrix comparing the behavioral response elicited by the two test contrasts. The RDM is shown in Figure 2c. Low dissimilarities (green colors) indicate that subjects exhibited similar behavior when presented with those test contrasts (e.g. subjects performed well on both contrasts or poorly on both contrasts). High dissimilarities (red colors) indicate that subjects exhibited the opposite behavior when presented with those test contrasts (e.g. subjects performed well on one contrast and poorly on the other contrast). There was high dissimilarity between the Identity Trials (Contrasts A–D) and the View-Specific Trials (Contrasts E–H). Thus, in a newborn's first visual object representation, identity information and view-specific information are negatively correlated with one another.

Abstraction space

To visualize the pattern of variation across subjects, I charted the position of each subject's object representation within a two-dimensional abstraction space (Figure 5).

This space contained sensitivity to identity information and sensitivity to view-specific information as axes. The position of the representation within the space reflects its degree of abstraction (i.e. the representation's sensitivity to identity information and tolerance to retinal image variation). Representations occupying the top left area of the space would be highly sensitive to identity information and highly insensitive to view-specific information. Representations occupying the top right area of the space would be highly sensitive to both identity and view-specific information. Representations occupying the bottom right area of the space would be highly sensitive to view-specific information and highly insensitive to identity information. And representations occupying the bottom left area of the space would not be sensitive to identity or view-specific information.

To calculate the position of each subject's representation in the space, I divided the time spent with the correct animation by the time spent with both animations, collapsing across Contrasts A–D (Identity Trials) and across Contrasts E–H (View-Specific Trials). This yielded separate performance scores for the Identity Trials and the View-Specific Trials. These performance scores were then scaled to fit on the axes, such that '0% sensitivity' marked chance performance and '100% sensitivity' marked perfect performance. Perfect performance was equal to

each subject's performance on the rest periods. The rest periods measured the amount of time subjects generally wanted to spend in proximity to their imprinted object. During the rest periods, subjects spent 85% (SEM = 1%) of their time with the imprinted object, showing that chicks spend most (but not all) of their time with their imprinted object.

The abstraction space is shown in Figure 5. Each dot depicts the object representation built by one subject. Visual inspection shows that the object representations lie within a constrained space within the theoretically possible space. This variation produced a significant negative correlation between sensitivity to identity information and sensitivity to view-specific information (Pearson correlation: $r = -.52$, $p = .04$).² Subjects who were more sensitive to identity information were less sensitive to view-specific information, and vice versa.

The abstraction space also shows that subjects built many different kinds of representations. Some subjects built representations that were highly sensitive to identity information but not at all sensitive to view-specific information (the representations occupying the top left area of the space). Conversely, other subjects built representations that were highly sensitive to view-specific information but not at all sensitive to identity information (the representations occupying the bottom right area of the space). The majority of subjects generated representations that were somewhere in-between these extremes, building representations that were sensitive to both identity and view-specific information. This pattern of variation occurred for both imprinted objects (Figure 5).

Interestingly, only one of the subjects built an object representation that was highly sensitive to both identity information and view-specific information (i.e. only one representation occupied the top right area of the space). Similarly, none of the subjects built an object representation that was highly insensitive to both identity information and view-specific information (i.e. no representations occupied the extreme bottom left area of the space). Thus, it was not the case that subjects were simply good or bad at object recognition tasks in general.

² To test whether this correlation was driven by one or a few outliers, I created 16 datasets, each containing the data from 15 of the 16 subjects (i.e. the subjects were iteratively removed from the datasets). The correlations ranged from $r = -.71$ to $r = -.44$ (mean correlation: $r = -.52$). Further, when only one-half of the subjects were included in the analysis (i.e. the dataset only included the subjects imprinted to one of the two objects), the correlation was similar to the correlation obtained when all of the subjects were included in the analysis (see Figure 5). Thus, the strong correlation obtained from the full dataset was not driven by one or a few outliers.

Rather, there was a trade-off between the ability to recognize familiar object identities and the ability to recognize familiar view-specific features.

Change over time analysis

To test whether subjects' performance changed over the course of the test phase, I analyzed the proportion of time subjects spent in proximity to their imprinted object as a function of trial number (e.g. first presentation of test stimuli, second presentation, etc.). Performance on both the Identity Trials and the View-Specific Trials remained stable over the course of the test phase (see Figure 2d), with little variation as a function of trial number: Identity Trials, $F(20, 335) = 0.86$, $p = .63$; View-specific Trials, $F(20, 335) = 0.82$, $p = .69$. The test-retest reliability was also high: Cronbach's $\alpha = .93$ for the Identity Trials and Cronbach's $\alpha = .95$ for the View-Specific Trials. Thus, subjects' performance was stable over time and cannot be explained by learning taking place across the test phase.

Discussion

This study examined the nature of a newborn animal's first visual object representation. Specifically, newly hatched chicks were raised for 2 weeks within controlled-rearing chambers. In their first week of life, subjects were shown a single virtual object rotating around a single axis. In their second week of life, I measured the information content of the object representation built by each subject. Three main findings emerged.

First, newly hatched chicks build object representations that contain both object identity information and view-specific information. On the Identity Trials, subjects were able to recognize their imprinted object across large and novel changes in the object's pixel-level appearance. Thus, newborns can build object representations that are selective to an object's identity and tolerant to variation in the retinal images produced by the object (see also Wood, 2013). On the View-Specific Trials, subjects preferred the imprinted object that had the same visual features as the object animation from the input phase. Thus, newborns can also maintain some sensitivity to the specific visual features that were used to build an object representation.³ This pattern shows that

³ Interestingly, subjects failed to distinguish between the animations showing the imprinted object rotating in the familiar direction versus unfamiliar direction (Contrast H). This indicates that newly hatched chicks do not use information about the temporal order of object views in order to recognize objects.

newborns do not build purely 'viewpoint-invariant' or 'viewpoint-dependent' object representations. Rather, they build object representations that contain a mix of both identity and view-specific information.

Importantly, subjects who were more sensitive to identity information were less sensitive to view-specific information, and vice versa. This trade-off is consistent with iterative, hierarchically organized visual processing machinery, the neural machinery underlying object recognition in adult primates (e.g. Sharpee, Kouh & Reynolds, 2013; Zoccolan, Kouh, Poggio & DiCarlo, 2007). Specifically, a large body of physiological evidence has shown that primates build invariant object representations through a cascade of iterative, largely feed-forward computations along the ventral visual stream (reviewed by DiCarlo *et al.*, 2012). These computations build increasingly selective and tolerant object representations by progressively integrating convergent input from lower layers. This feed-forward architecture produces a gradual increase in selectivity to an object's shape and tolerance to image variation along the hierarchy (e.g. representations in V2 are more selective and tolerant than representations in V1, while representations in V4 are more selective and tolerant than representations in V2). Thus, this architecture progressively transforms a retinal representation (view-specific information) into a new representation that is selective to an object's identity and tolerant to image changes (identity information). However, since organisms have a finite amount of neural tissue, they will only be able to perform a finite number of iterative transformations on the retinal input. Consequently, some retinal representations will not be fully transformed into a selective and tolerant object representation. Such representations might be partially invariant to image changes while still being sensitive to some of the specific visual features used to generate the representation. The present study shows that newborns build such 'partially invariant' representations.

Second, these results provide additional evidence that newborn animals can build invariant object representations from extremely sparse data (see also Wood, 2013). These subjects could not touch or interact with the virtual object, and their visual experience of the object was limited to movement around a single axis; nevertheless, some subjects were able to build highly invariant representations that contained no apparent memory trace of the specific visual features used to build the representation (i.e. the subjects who built representations that occupied the top left area of the abstraction space in Figure 5). This is a striking example of the power of newborns' object recognition abilities. Newborns do not need experience with a wide variety of objects in order to build invariant representations. The first object

representation built by newborn animals can be selective to an object's identity and tolerant to variation in the retinal images produced by the object.

These results also show that object identity representations can develop in the absence of sensory-motor experience with objects, which has long been argued to play a central role in the development of object concepts (e.g. Piaget, 1954). Since the objects in this study were virtual objects, they did not provide any tactile experiences to the subjects; nevertheless, subjects were able to build identity representations that generalized across large changes in the pixel-level appearances of the objects.

Third, these results show that a stable representation of an object can be built from a continuous flux of input. Since the imprinted object rotated continuously and was never stationary, the retinal images produced by the object were constantly changing. Nevertheless, subjects perceived that these changing retinal signals were produced by the same object and could generalize to novel appearances of the object. Thus, newborns can perceive constancy in object identity even when the sensory input from that object (and from all objects ever experienced) is always in flux.

In conclusion, this study shows that (1) the first visual object representation built by newly hatched chicks contains both object identity information and view-specific information, and (2) there is a trade-off between these two types of information. These findings are consistent with studies of visual object recognition in adult primates (e.g. Afraz & Cavanagh, 2008; Kravitz, Kriegeskorte & Baker, 2010; Kravitz, Vinson & Baker, 2008; Logothetis, Pauls, Bülhoff & Poggio, 1994; Zoccolan *et al.*, 2007). Thus, the initial state of object recognition in the avian brain produces similar representations as the mature state of object recognition in the primate brain. More generally, these results indicate that invariant object recognition is a core cognitive ability that can be present and functional at the onset of visual object experience.

Acknowledgements

I thank Samantha M.W. Wood for assistance on this manuscript. This work was supported by the University of Southern California.

References

- Afraz, S.R., & Cavanagh, P. (2008). Retinotopy of the face aftereffect. *Vision Research*, **48**, 42–54.

- Alemi-Neissi, A., Rosselli, B.F., & Zoccolan, D. (2013). Multifetural shape processing in rats engaged in invariant visual object recognition. *Journal of Neuroscience*, **33**, 5939–5956.
- Bateson, P. (Ed.) (2000). *What must be known in order to understand imprinting?* Cambridge, MA: MIT Press.
- DiCarlo, J.J., Zoccolan, D., & Rust, N.C. (2012). How does the brain solve visual object recognition? *Neuron*, **73** (3), 415–434.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1** (3), 515–533.
- Horn, G. (2004). Pathways of the past: the imprint of memory. *Nature Reviews Neuroscience*, **5** (2), 108–120.
- Karten, H.J. (2013). Neocortical evolution: neuronal circuits arise independently of lamination. *Current Biology*, **23** (1), R12–R15.
- Kravitz, D.J., Kriegeskorte, N., & Baker, C.I. (2010). High-level visual object representations are constrained by position. *Cerebral Cortex*, **20** (12), 2916–2925.
- Kravitz, D.J., Vinson, L.D., & Baker, C.I. (2008). How position dependent is visual object recognition? *Trends in Cognitive Sciences*, **12** (3), 114–122.
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, **2** (4), 1–28.
- Kruschke, J.K. (2011). *Doing bayesian data analysis: A tutorial with R and BUGS*. Burlington, MA: Academic Press.
- Logothetis, N.K., Pauls, J., Bülthoff, H.H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology*, **4** (5), 401–414.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- Sharpee, T.O., Kouh, M., & Reynolds, J.H. (2013). Trade-off between curvature tuning and position invariance in visual area V4. *Proceedings of the National Academy of Sciences, USA*, **110** (28), 11618–11623.
- Vallortigara, G. (2012). Core knowledge of object, number, and geometry: a comparative and neural approach. *Cognitive Neuropsychology*, **29** (1–2), 213–236.
- Wood, J.N. (2013). Newborn chickens generate invariant object representations at the onset of visual object experience. *Proceedings of the National Academy of Sciences, USA*, **110** (34), 14000–14005.
- Zoccolan, D., Kouh, M., Poggio, T., & DiCarlo, J.J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *Journal of Neuroscience*, **27**, 12292–12307.
- 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, USA*, **106** (21), 8748–8753.

Received: 30 August 2013

Accepted: 31 March 2014