Attention, Perception, & Psychophysics
2009, 71 (4), 712-723
doi:10.3758/APP.71.4.712

The role of sequence order in determining view canonicality for novel wire-frame objects

BENJAMIN J. BALAS AND PAWAN SINHA
Massachusetts Institute of Technology, Cambridge, Massachusetts

Objects are best recognized from so-called "canonical" views. The characteristics of canonical views of arbitrary objects have been qualitatively described using a variety of different criteria, but little is known regarding how these views might be acquired during object learning. We address this issue, in part, by examining the role of object motion in the selection of preferred views of novel objects. Specifically, we adopt a modeling approach to investigate whether or not the sequence of views seen during initial exposure to an object contributes to observers' preferences for particular images in the sequence. In two experiments, we exposed observers to short sequences depicting rigidly rotating novel objects and subsequently collected subjective ratings of view canonicality (Experiment 1) and recall rates for individual views (Experiment 2). Given these two operational definitions of view canonicality, we attempted to fit both sets of behavioral data with a computational model incorporating 3-D shape information (object foreshortening), as well as information relevant to the temporal order of views presented during training (the rate of change for object foreshortening). Both sets of ratings were reasonably well predicted using only 3-D shape; the inclusion of terms that capture sequence order improved model performance significantly.

For almost any object, there exists a set of views from which object identity can be recognized faster and more accurately than from others (Palmer, Rosch, & Chase, 1981). These views are called *canonical views* to reflect that they are the best views of an object for recognition. Broadly speaking, the existence of privileged views is consistent with view-dependent models of object recognition, in which complex objects are represented by a limited set of 2-D images, with some generalization function permitting interpolation between stored views (Bülthoff & Edelman, 1992; Tarr, 1995; Tarr & Bülthoff, 1995). In this article, we investigate how such privileged views of an object are determined during initial exposure to a moving object. Specifically, we examine the extent to which the sequence of views presented to observers contributes to the observer's selection of canonical views for recognition.

What makes a particular view of an object canonical? Most descriptions of the properties of canonical object views focus on properties of the orientation of the 3-D object, or of the 2-D image, ignoring the possibility that temporal factors (such as the order of views seen by the observer) might make an independent contribution. For example, it has been reported that canonical views tend to limit foreshortening of the object along major axes (Humphrey & Jolicœur, 1993; Lawson & Humphreys, 1996). Exceptions to this rule of thumb have been reported (Perret, Harries, & Looker, 1992), suggesting that, in some cases, *plan views* of objects (views obtained by looking directly down a major orientation axis) may be favored, though perhaps only for particular objects and subject to particular task requirements. A three-quarters view, often from a slightly elevated position, tends to be selected very frequently by observers who are asked to produce the "best" image of an object (Blanz, Tarr, & Bülthoff, 1999). It has been suggested that this vantage point provides the most information about 3-D form for a wide range of natural objects viewed in typical settings. Another explanation is that such views tend to minimize 2-D image redundancies (such as symmetry), making them information rich, as defined within an information theory framework. Computationally, pose robustness has been suggested as a principle governing the selection of canonical views for arbitrary objects (Peters, Zitova, & von der Malsburg, 2002). Under this proposal, the attributes of any individual image are irrelevant. Instead, it is the extent to which small perturbations of object position lead to substantial changes in 2-D appearance that determines canonicality. A canonical view, as defined by pose robustness, is a view of the object that will change little when the object moves slightly. Such stability over small pose changes might make it likely that this view would be selected via unsupervised clustering algorithms, for example. A similar notion is embodied in the decomposition of an object's viewing sphere into an *aspect graph* (Koenderink & van Doorn, 1979), which is a formal geometrical structure describing the object's appearance in terms of a limited set of views.

In the present study, we ask a question that has heretofore not been addressed by any of the aforementioned

B. J. Balas, bjbalas@gmail.com

descriptions of canonicality or their associated theoretical proposals: Does object motion contribute to the selection of canonical views? A growing literature suggests that object motion is a critical component in object recognition and learning. Object motion can be an independent feature for recognition (Newell, Wallraven, & Huber, 2004), and, in many cases, there is an advantage for matching and recognizing moving objects relative to static ones. Moreover, disruptions of a sequence of object views (such as reversal) incur a recognition cost even when the set of static images is fully preserved (Stone, 1998, 1999; Vuong & Tarr, 2004), indicating that the spatiotemporal signature of an object is a key part of its representation. Finally, temporal association between views of an object contributes both to the achievement of robust invariance over the observed views (Cox, Meier, Oertelt, & DiCarlo, 2005; Wallis, 2002; Wallis & Bülthoff, 2001) and to the achievement of fine-grained sensitivity to subtle differences in appearance (Balas & Sinha, 2008). The sequence of object views seen by an observer clearly has wide-ranging consequences for their ultimate representation of that object's appearance.

Our goal was to specify the extent to which object motion contributes to determining what object views are better (more canonical) than others. To do this, we chose to adopt a modeling approach so that we could define distinct features that reflected form-based and sequence-based aspects of the spatiotemporal inputs that we presented to observers. Our assessment of the relative contributions of form and sequence information thus hinged on our ability to specify a useful model of view canonicality for the set of stimuli that we would use to obtain behavioral data from observers. This is, in general, very challenging, and previous attempts to accurately predict canonicality for arbitrary objects have met with only limited success (Tarr & Kriegman, 2001). For our purposes, we chose to use a set of simplified artificial objects to make it more likely that we could develop a usable model.

Specifically, we used *paper clip* objects in both of our behavioral tasks (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992). These objects are made up of several line segments oriented in 3-D space and joined together to form a wiry shape. Although these are highly unnatural stimuli, they offer several important advantages. For example, there has been a great deal of previous work with these stimuli suggesting that observers recognize them in a view-dependent way (Bricolo, Poggio, & Logothetis, 1997; Edelman, 1999; Logothetis, Pauls, Bülthoff, & Poggio, 1994; Logothetis, Pauls, & Poggio, 1995), which is a critical requirement for our experiments to be informative. Paper clip objects are also highly unfamiliar, providing us with the opportunity to study canonical view selection independent from cognitive factors that we cannot hope to model solely with object appearance. For example, objects with writing or faces on their surfaces are likely to be oriented so that they are readable or so that the face is visible, possibly regardless of the geometric qualities of that view. Additionally, the 3-D form of a paper clip is sufficiently simple that we were able to formalize a straightforward model of view canonicality that is both plausible and easily testable. In particular, paper clips do not undergo self-occlusion as they rotate, obviating the need to consider the influence of features that disappear from view during particular viewing sequences. Obviously, the choice to use such unnatural and highly constrained objects potentially limits our ability to generalize our findings to natural stimuli. We take up this issue in the General Discussion section, where we outline how our model can be usefully extended to more complex, natural objects. For now, we point out that employing artificial objects in our experiments has the benefit of allowing us to build a detailed quantitative model that we can meaningfully apply to behavioral data.

In two experiments, we measured view canonicality following exposure to rigidly rotating paper clips (Figure 1). In Experiment 1, we obtained observers' subjective judgments of view goodness on a 1–7 scale following exposure to a rotating object. In Experiment 2, we used recall rates for individual frames in an *old–new* memory task as a more objective measure of canonicality. In both cases, we compared the behavioral data with the predictions made by a form-based model of canonical view se-

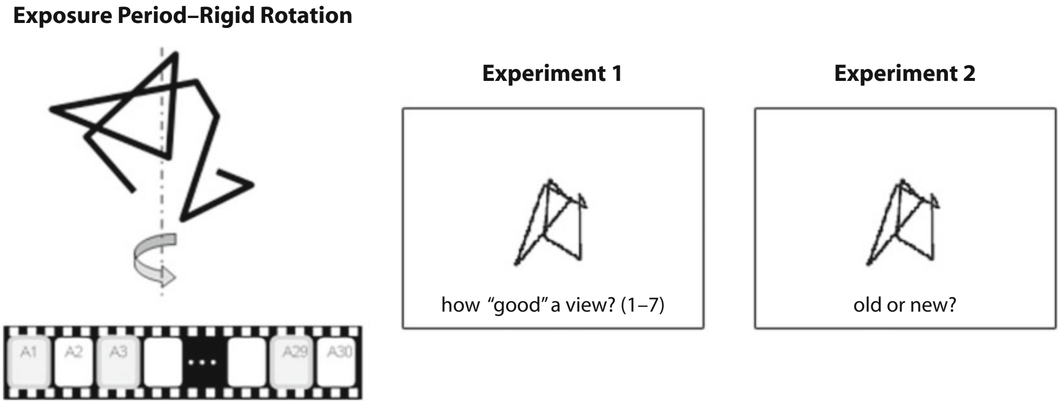

Figure 1. A schematic view of the two experiments that we conducted. Observers watched novel paper clip objects rotate about their vertical axis during an exposure period, after which they were asked either to provide subjective ratings of the goodness to individual views (Experiment 1) or to perform an *old–new* memory judgment for appearances observed during the exposure period (Experiment 2).

lection, defined in terms of projection ratios that could, in principle, be applied to arbitrary objects. Finally, we asked whether or not the fit of the model could be meaningfully improved by including sequence-based information, which would suggest that object motion (or at least sequence order) played a role in determining which object views were preferred.

In examining the data from both experiments, we asked three key questions:

1. Are consistent canonical views evident across subjects in the behavioral data?
2. If so, can we approximate the behavioral data with a purely form-based model?
3. Does the inclusion of information regarding sequence order improve our model's goodness of fit?

We began by examining observers' subjective ratings of object views following brief exposure to novel paper clip objects.

EXPERIMENT 1

In our first experiment, observers were asked to make subjective judgments about image canonicality for novel paper clip objects after viewing short sequences depicting rigid rotation of the objects. These judgments were then compared with a computational model of view canonicality, which we describe in more detail below.

Method

Subjects. Twenty-four members of the Massachusetts Institute of Technology (MIT) community volunteered to participate in this experiment. All observers reported normal or corrected-to-normal vision and were naive to the purposes of the experiment. None of the observers had previous experience with paper clip objects resembling those used in the study.

Stimuli. We created four distinct paper clip objects in MATLAB by randomly selecting eight points in spherical coordinate space that fell within a sphere of unit radius. These points were then joined together in random order to form a closed loop (so that line terminators were not a conspicuously salient feature). Each object was then rotated completely about its vertical axis in steps of 12º, yielding a total of 30 images of each object. The objects were arbitrarily labeled A, B, C, and D and are displayed in Figure 2. Each image subtended approximately 3º of visual angle when presented on screen.

A model of canonicality. The simplified shape of paper clip objects makes it straightforward to formalize a structural model of view canonicality on the basis of foreshortening of object parts. Foreshortening has previously been proposed as a key determiner of canonicality in theories of object representation (Marr, 1982), and paper clip objects lend themselves well to a model that uses foreshortening as a basic descriptor of object attitude. Could we use other proposals as a basis for determining canonical views of a paper clip? Aspect graphs are problematic to compute for paper clips because of their lack of self-occlusion. Similarly, feature-based methods (Denton, Demirci, Abrahamson, Shokoufandeh, & Dickinson, 2004) are also difficult to use, since many features of a paper clip are highly unstable because of the sparse wire-frame structure of the object. Foreshortening is easily applicable to these objects and, more important, provides a straightforward means of investigating the contribution of sequence-based terms in determining view canonicality. Measuring the projected length of segments in the image is an essentially noise-free operation relative to feature tracking, and, likewise, the 3-D segment lengths are veridical. These features are thus easy to compute, and so this model is not complicated by the need to define and robustly detect features within the image. We therefore remain open to the possibility that other methods could usefully describe paper clip images but opt, in the present experiments, to use a model based on segment foreshortening because of these advantages. Crucially, we emphasize that our goal in the present study was not to develop a state-of-the-art model of canonicality that supplants all others. Rather, our goal was to use a simplified set of objects (and a similarly simple model of canonicality) to examine the extent to which viewing sequence plays a role in determining view canonicality for novel objects. We continue by describing our model.

To motivate our model, consider the limiting case of a paper clip object with only one segment. This segment has a true length, and our model is built on the premise that a fundamental component of determining a canonical view of this minimal object is the extent to which the projected length in the image is close to the actual length. An end-on view of the segment would be the worst view possible, whereas any view of the segment oriented in the frontoparallel plane would be best. We can formalize this measure of foreshortening for a single segment by dividing the projected length in the image by the actual length in the world (2-D length/3-D length) to yield a value on the interval of 0 to 1 describing the amount of foreshortening in the view. To extend this measure to our eight-segment objects, we calculate this value for each segment in the object and take the mean over all values to yield a single value reflecting how closely the true lengths of the segments in the entire object are reflected in a particular image.

Therefore, if we have n pairs of 3-D points (each point in the pair denoted as $\{x,y,z\}$ and $\{x',y',z'\}$), the form-based canonicality of a particular view can be calculated as follows:

$$canonicality = \frac{1}{n}\sum_i \frac{\sqrt{(x_i'-x_i)^2+(y_i'-y_i)^2}}{\sqrt{(x_i'-x_i)^2+(y_i'-y_i)^2+(z_i'-z_i)^2}}.$$

This value also ranges between 0 and 1, inclusive. In Figure 3A, we display the canonicality values calculated under this model for all views of our four test objects. To add information regarding the sequence of images to our baseline model, we took the first and second derivatives of each original canonicality function and included these derivatives as additional regressors in a multivariate linear regression model. The method for constructing derivatives is displayed in Figure 3B. Our proposal is that, if the inclusion of these additional sequence-based terms improves the fit of our model to the data, we can tentatively conclude that aspects of object motion (or at least sequence order) influence perceived canonicality. What information do these derivatives carry that the original measure of foreshortening does not? Both derivatives describe how the average foreshortening of the object changes over time. The first derivative is an indicator of whether average foreshortening is increasing or decreasing and tells us the magnitude of that change at each frame. The second derivative similarly indicates the acceleration of that change and is also useful in determining whether or not a particular frame represents a local minimum or maximum of average canonicality. Both of these regressors thus represent basic dynamic properties of the sequence of views presented. Including both of these regressors in the present model is a means of exploring the temporal extent of sequence-based processing of individual views, a process we could continue by adding higher order derivatives. Presently, however,

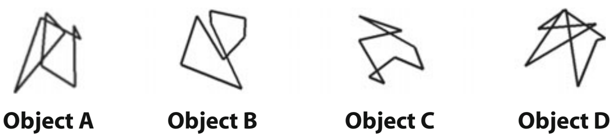

Figure 2. From left to right, Objects A, B, C, and D used in each of our two experiments. Each object has eight unique vertices joined together in a randomly determined order to form a closed loop.

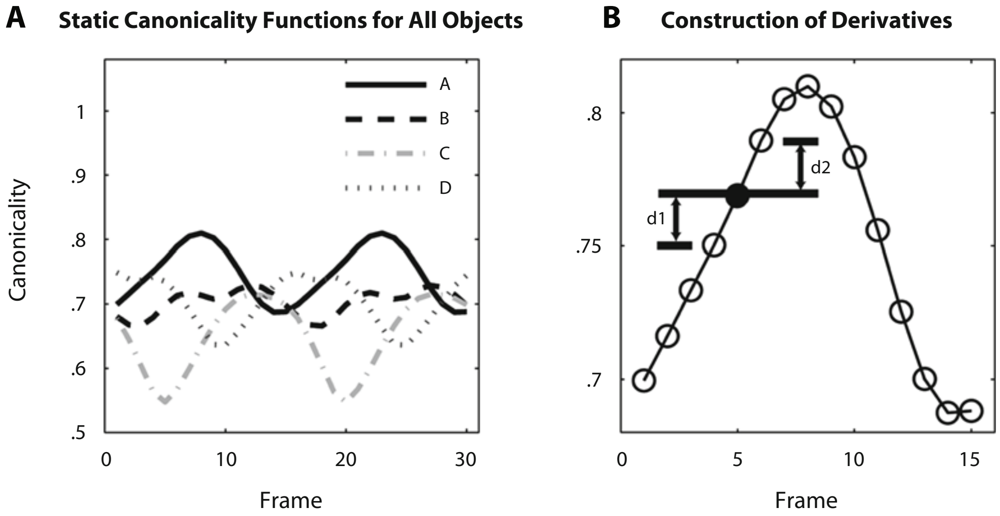

Figure 3. (A) Static-form canonicality functions constructed for each of the four paper clip stimuli used in the experiments using the foreshortening criterion. (B) Augmenting the static model with additional dynamic regressors is done by computing differences in the static canonicality function at each point. The first derivative at the black dot would be the signed difference, d1, whereas the second derivative at the same point would be the signed difference between the values of d1 and d2.

these two derivatives describe straightforward properties of the average foreshortening function as it extends over time (local slope and local curvature), and including derivatives up to the second order is a useful compromise between exploring an intractably large number of regressors and examining more than just first-order sequence information concerning average foreshortening.

Procedure. Each observer provided canonicality ratings for two objects, one after the other, with presentation order balanced across subjects. For each object, the experimental session began with a brief exposure period during which the object under consideration rotated rigidly about its vertical axis. During the exposure period, each object completed eight full revolutions in which each of the 30 images comprising the full sequence of object views was displayed. The frame rate of the monitor was 60 Hz, and each sequence was played at a rate of 7.5 frames per second.

Following this exposure period, observers were asked to provide canonicality ratings for all 30 images that appeared in the training sequence. Subjects were asked to use a 1–7 Likert scale, in which 1 corresponded to a *very poor view* of the object and 7 corresponded to a *very good view*. Observers were instructed that they should rate the images according to their assessment of how "good" an image each object view was. To be more concrete, we suggested that they consider how likely they would be to show each image to a friend if they had to provide examples of what the training object looked like. During the rating task, each image from the training sequence was shown twice in randomized order. Each image remained on screen until the observers provided a rating via the keyboard.

During both the exposure period and the rating task, the observers were seated approximately 50 cm from the monitor in a brightly lit room with no restraints on head position or gaze. Each image subtended approximately 3º of visual angle on screen. All stimulus display and response collection routines were executed with the MATLAB Psychophysics Toolbox for Windows (Brainard, 1997; Pelli, 1997).

Results

For each subject, the ratings provided for each image were averaged, yielding a complete canonicality function over the full set of 30 views per object. In Figure 4, we display the mean canonicality ratings for Objects A, B, C, and D (open circles).

It is evident in Figure 4 that observers were in good agreement regarding their ratings for these objects, leading to strong modes in the data for each stimulus. Specifically, each object appears to have two preferred views that are diametrically opposed on the viewing circle. These images are mirror images of each other because of our use of orthographic projection in rendering the paper clip objects. It is also important to note that the location of these modes is not the same across objects, as we might expect if view canonicality was being driven by generic memory processes, such as primacy and recency effects.

Our next step was to assess the model's goodness of fit for each object's set of ratings, with and without sequence-based terms included. We calculated the correlation coefficient between the actual data (using the mean across observers for each view) and the best fit of the model with and without the sequence-based regressors. To properly compare the model fits of the form-based model and the form+sequence-based model, we used Akaike's information criterion (AIC; Akaike, 1974), a standard metric for comparing the performance of nested models that makes explicit the trade-off between model accuracy and model complexity. The AIC is defined only up to an additive constant, meaning that only relative values for nested models are relevant. Specifically, the preferred model in any case is the one that yields the *smallest* AIC value. The AIC thus does not yield a p value like typical null-hypothesis tests of significance would, nor are models rejected via the value of the AIC. Instead, the number yielded for a particular model is directly compared with the results of competing models. Given the relatively small number of data points to be fit by our model (relative to the number of free parameters), we used a modified version of the AIC designed for smaller data sets (Burnham & Anderson, 2004) to help mitigate the known tendency of the AIC to be overly lenient in permitting additional free param-

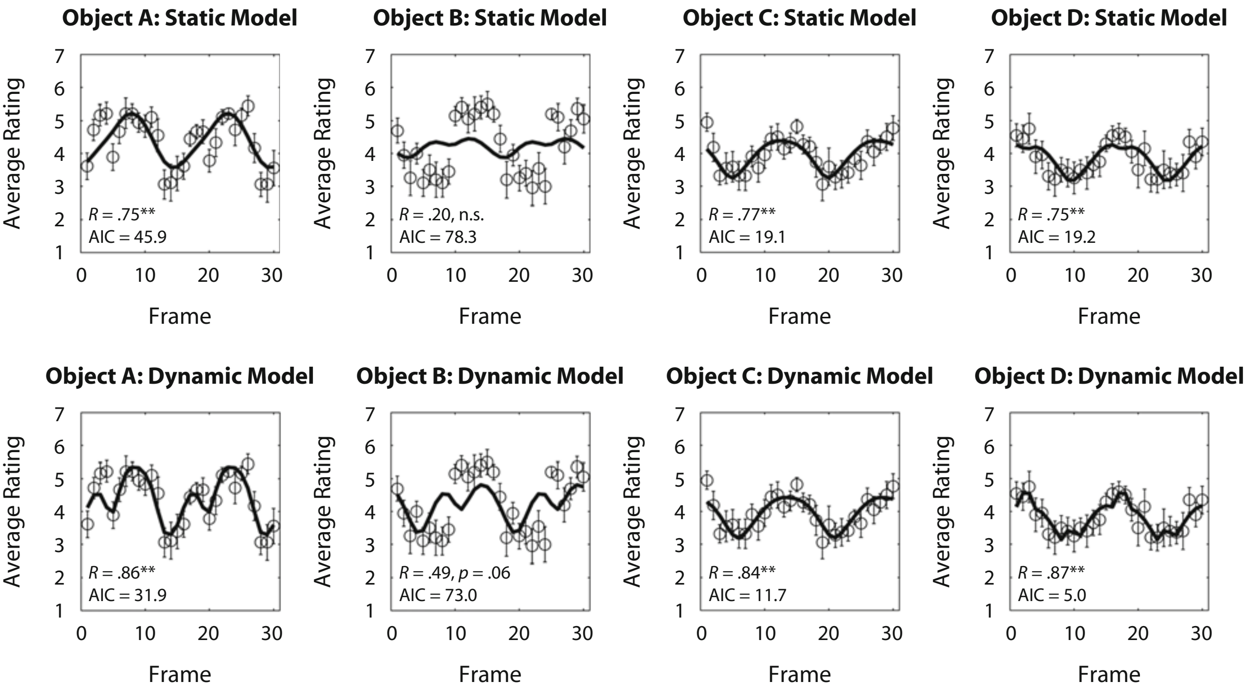

Figure 4. Mean observer ratings across all subjects for Objects A, B, C, and D, replotted along with the best-fit lines obtained from application of the static-form model (top row) and the dynamic model (bottom row). Error bars represent one standard error of the mean. Within each subplot, the correlation coefficient for each model is listed, along with the AIC value. Bigger values of *R* and smaller values of the AIC indicate a better fit to the data. $^{**}p < .01$.

eters and, hence, overfitting. We note briefly that there are multiple tools that one could potentially use to compare nested models, including leave-one-out cross-validation methods (Browne, 2000) and Bayesian information criterion measures. These different options vary in how conservatively additional free parameters are penalized, and each has distinct strengths and weaknesses (Burnham & Anderson, 2002, 2004). For the number of free parameters and data points involved in the present data set, the modified AIC is a suitable measure that corrects for the tendency toward overfitting exhibited by the unaltered AIC. We revisit the issue of alternative methods of model comparison in the General Discussion section.

In Figure 4, we display the best fits of this model with only the form-based regressor and then with the additional sequence-based regressors added in. The correlation coefficient and AIC values are reported within each subplot for ease of comparison. Also, Tables 1A and 1B list 95% confidence intervals for the coefficients of each regressor in the model.

For three of our four objects, the form-based model provided a good fit to the data. Given the unconstrained and subjective nature of the task, as well as the simplicity of the model, this is surprising. However, for one of our objects (Object B), this model failed to capture any significant portion of the variance in the observed data. Examining Figure 3, we can see that, in our model, this object's form-based canonicality function is very shallow relative to the others, suggesting that there is perhaps not enough vari-

Table 1A
Confidence Intervals (95%) for the Coefficients of All Regressors in the Multiple Regression of the Model Described in the Text on Observers' Ratings of Canonicality (1–7 Scores)

Object	Average Foreshortening	First Derivative	Second Derivative
A	[−2.0, 3.4]	[−4.5, 2.6]	[−13.7, 14.7]
B	**[0.1, 4.6]**	**[−7.6, −0.8]**	[−10.1, 1.3]
C	[−1.1, 0.57]	**[−3.9, −1.2]**	[−6.5, 0.7]
D	[−0.58, 1.7]	**[−4.42, −0.37]**	[−7.5, 0.6]

Note—Intervals that do not cover zero are in boldface. The regressors display varying correlation with one another, depending on the smoothness of the average foreshortening function, which can dramatically affect the regression coefficient in the full model. Table 1B contains intervals for each regressor considered singly.

Table 1B
Confidence Intervals (95%) for the Coefficients of All Regressors Included in the Model Described in the Text on Observers' Ratings of Canonicality (1–7 Scores)

Object	Average Foreshortening	First Derivative	Second Derivative
A	**[8.8, 18.11]**	[−25.8, 6.06]	**[−82.6, −23.4]**
B	[−7.8, 26.1]	**[−62.12, −7.5]**	[−55.3, 29.1]
C	**[4.6, 9.0]**	**[−16.3, −2.7]**	**[−33.7, −9.5]**
D	**[6.4, 12.8]**	[−9.3, 11.7]	**[−32.3, −0.6]**

Note—Intervals that do not cover zero are in boldface. These intervals reflect the results of considering each regressor singly (using only a constant term and one regressor to fit observers' data). This provides a complement to the data in Table 1A, which describes the results of using the full model, which includes regressors that correlate to a varying degree.

ability across views to successfully capture the structure in the behavioral data. In all four cases, the proportion of variance captured by the model increased when we added sequence-based regressors, to the extent that Object B was reasonably well fit by the model. Crucially, the value of the AIC was smaller in each case as well, indicating that the improvement in goodness of fit was large enough to merit the increase in model complexity.

Discussion

Using novel objects and a subjective rating task, we found that observers were consistent in their selection of canonical views. The overall structure of the empirical canonicality functions was generally well fit by a model of average foreshortening across distinct segments but was further improved when information concerning sequence order was included in the model. Perceiving object motion may influence canonicality in this task in two ways: First, motion may provide cues to each segment's true length via the kinetic depth effect (Ullman, 1979; Wallach & O'Connell, 1953), allowing our observers to estimate the 3-D shape information necessary to compute the form-based canonicality function that we have defined. Second, the added benefits of including sequence-based terms in our model as independent regressors suggests that the change in that function over time also contributes to the canonicality of individual views. Thus, the particular sequence observed affects canonicality beyond providing a useful means of extracting 3-D form.

We continued by examining view canonicality for these objects as determined by an objective criterion rather than by subjective ratings. First, we asked whether canonicality judgments, as measured using recall rates, resemble the data obtained from subjective ratings. Second, we examined the roles of object form and sequence order by asking whether either of the two models that we have developed can also fit this behavioral data reasonably well.

EXPERIMENT 2

In this second task, we used recall rates in an *old–new* object recall task as a measure of view canonicality following exposure to the rigidly rotating objects used in Experiment 1. This task provides a potentially more objective means of defining canonicality than do the unconstrained ratings obtained in Experiment 1.

Method

Subjects and Stimuli. Twenty-four volunteers from the MIT community participated in this task. All observers reported normal or corrected-to-normal vision, and none had taken part in Experiment 1. As in Experiment 1, subjects reported having no previous experience with paper clip objects similar to those used in this task. The same four paper clip objects used previously in Experiment 1 were used again here.

Procedure. The testing session was divided into two parts (one per target object), each with four identical blocks. As in Experiment 1, each block began with observers viewing a paper clip rigidly rotating about its vertical axis. Sequences were played at a rate of 7.5 frames per second for eight repetitions of the full image set.

Following this exposure period, each block continued with a test phase in which object recall was assessed. On each trial of this task, observers were presented with one image in the center of the display, which depicted either an image from the exposure period or an image of an entirely different object (one of the other target objects was used as a distractor, counterbalanced across subjects). The observers' task was to decide as rapidly and accurately as possible whether the image had been in the preceding sequence, using a go/no-go paradigm. If observers believed that the image had been present during exposure, they were to press the space bar as fast as possible. Otherwise, they were to do nothing, and the image would disappear after 3 sec. It was emphasized that response times (RTs) were being recorded and that the subjects should therefore try to respond as quickly as possible without compromising their accuracy. Within a block, each of the 30 target and distractor images was presented twice, for a total of 120 trials per block and 480 trials in the full session. Presentation order was randomized for each subject.

All stimulus display and response collection routines were carried out as described in Experiment 1. Each observer carried out this task for two different objects, with distractor objects in each block and target object order counterbalanced across subjects.

Results

Accuracy. In Figure 5, we display the mean hit rate across frames for each object. The average false alarm rate for all observers across all objects was approximately 4%, which is low enough that we did not correct the hit rate for guessing.

As in Experiment 1, we observed object-specific structure in our canonicality functions, this time as defined by recall rather than by a subjective rating. To what extent are our subjective and objective measures of view canonicality correlated with one another? We measured the correlation coefficient between the mean ratings obtained in Experiment 1 and the mean hit rates across frames obtained in Experiment 2. In general, we found reasonable agreement between the two functions (Object A, $R = .71$, $p < .01$; Object B, $R = .55$, $p < .01$; Object C, $R = .48$, $p < .01$), with the exception of Object D ($R = .28$, $p = .13$). Overall, we concluded from this result that the two measures are related, but not so strongly that these definitions of canonicality should be used interchangeably.

We continued by assessing the goodness of fit for both the form-based model and the form+sequence-based model described in Experiment 1. As before, we were particularly interested in whether the inclusion of sequence-based regressors improved model performance substantially. Figure 5 displays the model fits in each case, along with correlation coefficients and AIC values calculated as described in Experiment 1. Tables 2A and 2B contain 95% confidence intervals for model coefficients obtained from the full multiple regression analysis and also from fitting the model using only one regressor at a time.

Table 2A
Confidence Intervals (95%) for the Coefficients of All Regressors in the Multiple Regression of the Model Described in the Text on Observers' Recall Rates in the *Old–New* Task

Object	Average Foreshortening	First Derivative	Second Derivative
A	**[2.9, 30.1]**	[−17.5, 18.2]	[−51.9, 89.2]
B	[−41.23, 6.9]	**[−97.6, −24.3]**	**[−129.1, −7.7]**
C	**[0.8, 5.6]**	**[−16.2, −5.8]**	**[−33.6, −5.4]**
D	**[10.6, 19.0]**	**[5.2, 20.11]**	**[6.5, 36.3]**

Note—Intervals that do not cover zero are in boldface. Table 2B contains intervals for each regressor considered singly.

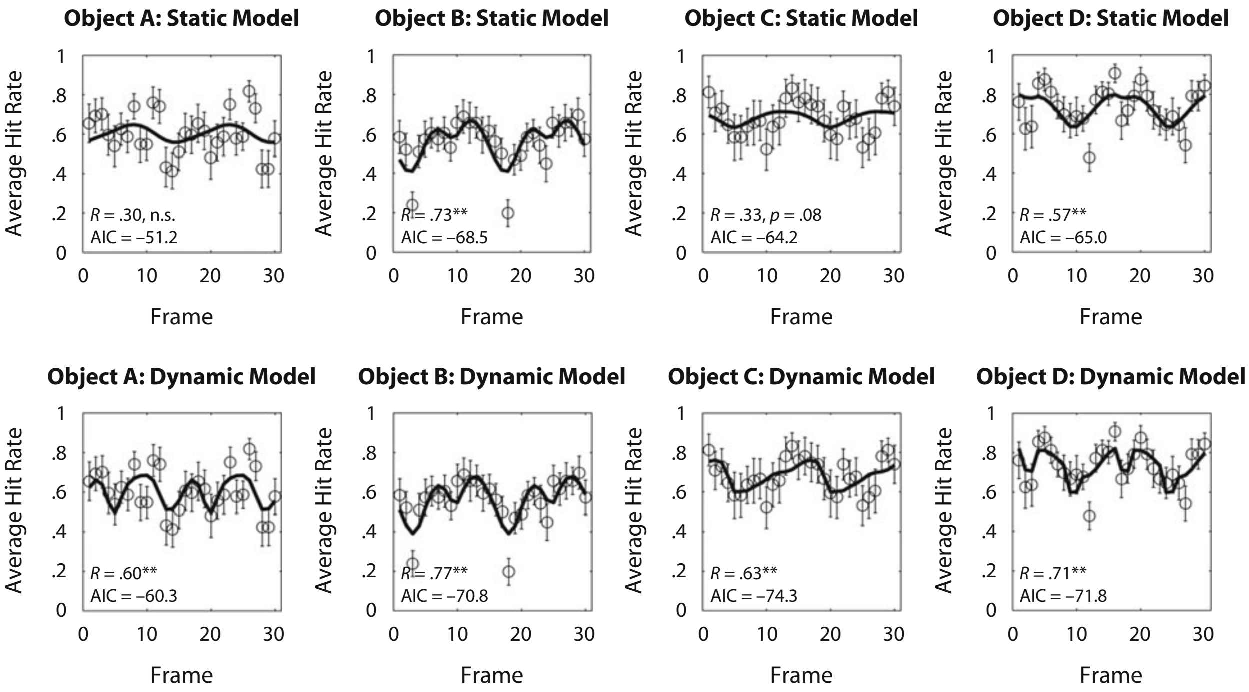

Figure 5. Mean hit rates for Objects A, B, C, and D, replotted along with the best-fit lines obtained from application of the static-form model (top row) and the dynamic model (bottom row). Error bars represent one standard error of the mean. Within each subplot, the correlation coefficient for each model is listed, along with the AIC value. Bigger values of *R* and smaller values (more negative) of the AIC indicate a better fit to the data. $^{**}p < .01$.

As in the case of the subjective ratings obtained in Experiment 1, the form-based model provided a good fit to the data, and the inclusion of the sequence-based terms improved the model's fit substantially. Note that the AIC is defined only up to a constant, and so the *more negative* AIC values in Tables 2A and 2B indicate a better fit, taking into account the number of parameters in each model.

Discussion

Observers' memory for target frames observed during the training sequence correlated reasonably with their subjective evaluation of canonicality. Canonicality, as defined here by observers' recall rates, also appeared to incorporate information regarding sequence order, as evidenced by the improvement in goodness of fit accompanying the inclusion of our sequence-based terms. This finding suggests that the particular directed sequence of views used during training has consequences for view selection beyond the form-based properties inherent to each image. We conclude both by discussing our results in the context of existing data regarding object motion and recognition and by elaborating on how our model might be extended to include more complex objects.

Table 2B
Confidence Intervals (95%) for the Coefficients of All Regressors Included in the Model Described in the Text on Observers' Recall Rates in Our Old–New Task

Object	Average Foreshortening	First Derivative	Second Derivative
A	[−0.19, 1.66]	[−3.55, 0.86]	[−7.13, 2.7]
B	**[2.61, 5.55]**	**[−7.67, −0.83]**	**[−10.5, −0.88]**
C	**[0.01, 1.03]**	**[−3.05, −0.96]**	[−3.3, 1.63]
D	**[0.64, 2.23]**	**[−3.94, −0.11]**	**[−6.53, −0.36]**

Note—Intervals that do not cover zero are in boldface. These intervals reflect the results of considering each regressor singly (using only a constant term and one regressor to fit observers' data). This provides a complement to the data in Table 1, which describes the results of using the full model, which includes regressors that correlate to a varying degree.

GENERAL DISCUSSION

Across three tasks with the same four novel paper clip objects, we examined the factors governing appearance encoding for a novel moving object in several ways. In Experiments 1 and 2, we attempted to fit data from subjective and objective view canonicality tasks to a structural model that could either include or exclude terms relevant to the extended temporal structure of the input. In both cases, we found that including sequence-based terms improved the model's ability to fit the data, suggesting that the order of appearances presented to the observer makes a distinct contribution to canonicality judgments. Given these results, we begin our discussion by addressing the questions that motivated this study, discussing possible topics for further investigation along the way.

Are Consistent Canonical Views Evident in the Behavioral Data?

In general, we have observed evidence for privileged views in all of our tasks. Subjective ratings from Experiment 1 and recall rates from Experiment 2 both showed significant fluctuations across target frames that were object specific. These results suggest that canonicality (as defined in each task) was not determined by general-purpose memory processes, such as the primacy and recency of views, but was directly related to differences in object appearance across target frames. As a result, we conclude that privileged views, or *keyframes*, emerge naturally following brief exposure to a novel moving stimulus and can successfully be determined by either subjective or objective measures. Furthermore, we confirmed that these distinct definitions of canonicality are correlated, indicating that view preferences are fairly robust and generalize, to a limited degree, across testing procedure.

Can We Approximate the Behavioral Data With a Purely Form-Based Model?

Our form-based model of canonicality depends on a drastic simplification of object shape, but it worked surprisingly well in both tasks. There are certainly more factors to consider (such as image-based features and the primacy and recency effects that we have so far discounted), but this simple model provides a useful starting point for further investigation. Could we scale this model up to work with more complex natural objects? The stick-figure nature of our model is an obvious limitation to carrying out a meaningful analysis of complex object appearance, but we briefly discuss two distinct ways in which our model could be extended to consider a wider range of objects.

One possibility for adapting our existing model to be applicable to arbitrary 3-D objects is to reduce objects to a skeleton of vertices defined by salient image patches. *Interest points* could be defined on the first image of an object by standard techniques for identifying salient image regions in complex scenes (e.g., center-surround or Forstner operators). Image patches could then be tracked from frame to frame, allowing for the computation of projected segment length between the centroids of paired image patches. Even without data concerning the true length of the virtual segment joining two patches in 3-D space, the change in projected length can still be computed. This model, using only changes in projection, rather than ratios of projected length to true length, is a fine analogue of our paper clip model in concept, but there are several issues that may plague its application. First, it is unclear how we should cope with self-occlusion, since features can be lost during rotation. Wallraven and Bülthoff (2001) suggested an interesting computational approach that tracks micropatterns within the larger image and registers a pattern's disappearance from view as part of a criterion for computing keyframes within a larger sequence. This method offers a potentially useful technique for quantifying canonicality in terms of local changes in feature visibility. A second challenge facing such a model is coping with the uncertainty in the position of a particular feature because of changes in appearance during rotation or smoothness of object form. This uncertainty, which could be substantial, depending on the surface properties and geometry of the object, will add noise to our estimates of foreshortening. The result may be a fragile model that breaks easily when subjected to a real object. That said, an implementation of a model similar to the one described here has been tested for object recognition (Grabner & Bischof, 2005). The goal of their model is to achieve a useful set of local image features that can be used for recognition. To ensure that the features they select are robust, their model learns a set of image patches by tracking local features during an extended spatiotemporal sequence and retaining the most stable (and therefore trackable) features. The result of their procedure is a useful vocabulary of local features for static recognition, but their methods for tracking micropatterns during an extended sequence may be directly applicable to constructing a model of canonicality for natural objects. Another means by which a model of canonicality could be built upon a wire-frame representation of a natural object is to use recently developed methods for determining the skeleton of an object's structure (Feldman & Singh, 2006). These methods are reminiscent of medial-axis representations of object geometry (Blum, 1967) and use sophisticated statistical inference procedures to generate complex branching structures from arbitrary object images. The resulting skeleton could be used as the basis of a foreshortening model, obviating the need to track features across successive frames.

An alternative method for specifying canonicality for arbitrary 3-D objects is to discard the requirement that segments be constructed from the original image of a target object. Describing a real object in terms of a set of line segments ignores the fact that most natural objects have surfaces, which are surely more relevant to view canonicality and recognition than an artificial skeleton imposed on the object for computational convenience would be. Indeed, Farah and Hammond (1988) observed drastic changes in observer behavior when paper clip objects were transformed into *potato-chip* objects by interpolating a surface between the line segments of the paper clip in a manner similar to the way soap film forms on twisted wire. We suggest that a better model of canonicality (that retains the logic of our paper clip model) can be defined for arbitrary objects in terms of surfaces rather than line segments. Specifically, we submit that, just as a canonical view of a paper clip can be defined as one in which the segments are oriented, on average, as much in the frontoparallel plane as possible, so too should a canonical view of a real object be one in which as much of the object's *surface* is oriented in the frontoparallel plane as possible. For an arbitrary 3-D object, this would mean that a canonicality function could be computed by determining the extent to which visible surface normals are oriented orthogonal to the picture plane, but this computation is bound to be difficult to achieve practically. However, a potentially useful proxy of this measure is to simply count the number of pixels in the image. As the average frontoparallel orientation of an object's surface increases, its overall image size should increase as well (although there are numerous exceptions to this rule, including highly stellated objects or objects with deep concavities). This very crude heuristic obviously does not reflect the rich percepts that we know accompany 3-D object perception, but it does offer a tractable means of extending our paradigm to include arbitrary 3-D shapes. We continue by discussing our results concerning canonicality and the sequence of views presented during exposure.

Does Sequence Order Influence Canonicality?

Having examined the evidence for a role of sequence order in canonicality judgments across three tasks, we conclude that the sequence of observed appearances does play a role in defining the keyframes of a novel moving object. Our modeling results from Experiments 1 and 2 revealed that, in every case, the inclusion of sequence-based regressors improved the proportion of variance captured by our models, occasionally succeeding where the form-based model has failed. The increase in goodness of fit was also validated using the modified AIC, demonstrating that the increase in model performance was worth the additional complexity introduced by adding regressors. Although the role of sequence order appears to be complex and object specific, our results suggest that this information is indeed a relevant factor in determining view canonicality. This result is broadly consistent with those of previous work that demonstrates that observing coherent object motion benefits object memory and recognition (Lawson, Humphreys, & Watson, 1994; Liu, 2007). A more complex issue is how our results relate to models that use temporal association between distinct object views to facilitate generalization across views via binding over time (Foldiak, 1991; Wallis, 1998). Given that these models suggest that temporal association promotes generalization across images, the continued observation of a moving object should eventually flatten out the behavioral canonicality function by increasing observers' tendency to bind images together into a common representation. Our proposal, by contrast, would be that temporal proximity might continue to be relevant to differentiating between distinct views and establishing canonicality. Further work could explore these two predictions via more explicit consideration of learning effects over the course of extended viewing of a moving object.

There are also a few caveats that we must add regarding the contribution of sequence order to canonicality judgments as observed here. The first of these is that it is clear from our analysis that sequence-based information makes a relatively small contribution to the overall canonicality function relative to information regarding 3-D shape. The form-based terms that we compute using our model generally do a good job of explaining the data, with the sequence-based regressors capturing a relatively small additional portion of the total variance. Viewing sequence thus matters but most likely matters a good deal less than does 3-D form. This conclusion is in line with previous results suggesting that pure spatial continuity between images is enough to induce the kind of image binding described in temporal association experiments (Perry, Rolls, & Stringer, 2006) and also with the finding that previous experience with specific object rotations is not necessary for robust extrapolation across views (Wang, Obama, Yamashita, Sugihara, & Tanaka, 2005). In general, it may be the case that the sequence of views seen by the observer makes a small enough contribution to the processes governing canonicality, recognition, or generalization across views that removing it from the equation does not cause the visual system to break in a profound way (Harman & Humphrey, 1999). To revisit a point made earlier, we also note that the modified AIC—although it was developed to counteract a tendency toward overfitting evident in the original AIC—may be more lenient than other measures of model suitability in its penalty for additional parameters. Given both the comparatively small amounts of variance captured by these terms and the possibility that the modified AIC may not be as conservative as other techniques for assessing nested models, we suggest that our results distinctly argue for motion as a secondary determiner of view canonicality.

There is also a second caveat regarding our analysis of sequence order that we have yet to directly address. Briefly, our results do not allow us to make a strong statement about whether it is sequence order or true temporal factors that influence canonicality. This is because the derivatives that we computed to build our sequence-based regressors were not strictly derivatives in time but, rather, derivatives across the sequence of forms that we presented during training. This means that our modeling thus far cannot distinguish between a temporal mechanism, as opposed to what one might call a *second-order* form mechanism, or a mechanism that is blind to time but integrates appearance across discrete changes in appearance. Although it may seem most natural to assume that an observer would construct functions like our sequence-based regressors online during sequence viewing, it could still be the case that the underlying representation is not based on the individual sequence observed during training. For example, these terms might really be computed by an offline process that generates a large family of local form comparisons after using object motion to extract a good estimate of 3-D shape. Alternatively, it may be that the only relevant factor in our tasks is the ordered sequence of images presented to the observer rather than the actual percept of a moving object.

Distinguishing between these possibilities is very difficult experimentally, since preserving the sequence of images in a dynamic stimulus while destroying the percept of object motion generally introduces new, nontrivial confounds, such as masks or blank intervals. Although we cannot definitively distinguish between these proposals with the data presented here, further investigation of temporal factors, such as the rate of sequence playback or the inclusion of nonuniform velocity, could help tease apart these factors. Given the obvious ecological relevance of object motion, however, these results certainly speak to aspects of form encoding that are engaged during natural visual experience with a moving object, even if the underlying mechanism is not truly motion based.

Other Issues Regarding Canonicality in General

Having discussed the three specific questions regarding view canonicality that we targeted in the present study, we continue by discussing additional issues relevant to the present discussion. The first of these concerns the definition of canonicality. In our experiments, we measured canonicality in two distinct ways: In Experiment 1, we asked observers to rate previously seen views on a 1–7 scale, and, in Experiment 2, we used recall rates for individual frames. These two measures (and variations of them) have been used by many of the previous studies mentioned in our introduction but demand very different things from the observer. In particular, Experiment 2 raises the issue of whether canonicality is a function of an object considered solely in terms of its own structure or if, instead, canonicality is a function of how discriminable two or more objects (and their constituent object views) are from one another. To put it another way, a view of a paper clip that is highly canonical, given a particular distractor, may be very poor if an alternative distractor has a very similar view. One can imagine highly pathological conditions like this, some of which are realizable in the context of paper clip objects. Our model—which uses only information about the target object to define canonicality—cannot make strong predictions about the ability to discriminate between particular targets and distractors, which is potentially a relevant factor to consider when using a recall task to measure canonicality behaviorally. In our case, using a varied set of distractors balanced across subjects helps ensure the robustness of the obtained canonicality values. Also, as we have already observed, our two distinct behavioral measures of canonicality appear to be related, despite the potential for task-related differences such as the one that we have just described. In general, however, this concern is a key issue for subsequent consideration. We suggest that future researchers should focus on using a range of methods to assess view canonicality, and direct investigation of the respective roles of interobject discriminability and intrinsic object structure would also be very useful.

A second issue that we should raise about the present study is the extent to which our model can be used as a heuristic tool for making predictions about novel objects. For now, we confine this discussion to the domain of paper clip objects, but it is also a relevant concern for more general applications of our model. There are two kinds of predictions that we consider: First, if an observer has seen a limited number of object views, can our model predict the canonicality (or recognizability) of unseen views? Second, if an observer has provided ratings for a limited number of object views, can our model predict the canonicality of unrated views? The key distinction between these two scenarios is whether or not task demands require the observer to extrapolate new views from limited information. We argue that the first scenario, in which observers would be required to do this, is not an appropriate framework within which to apply our model. Although we could use the limited data acquired from the observers' training views to fit the parameters of our model, the model is not designed to account for the observers' ability (or inability) to generate novel appearances from a sparse training set. This task likely requires abilities beyond computing how foreshortened an object is and comparing that value to nearby ones, potentially including robust structure-from-motion estimation, mental rotation, and subsequent projection. These additional demands—or whatever counterparts might be actually employed by the visual system—make it unlikely that our model would be a useful tool for modeling behavior in this scenario. We expect that our model's predictions would likely be far too optimistic, especially considering the relatively poor generalization performance exhibited by observers attempting to extrapolate to new views of a novel paper clip (Farah & Hammond, 1988). The second scenario, by contrast, offers a scenario highly amenable to predictive use of the model. By offering observers the chance to see all object views, measures of canonicality are not contaminated by the additional need to generate novel appearances. Given enough ratings to constrain the free parameters of our model (2–4, depending on the number of regressors included), our model can easily be used to make quantitative predictions of view canonicality. We have not conducted such an analysis in the present study (our focus being to describe the role of sequence order rather than to test predictive capacity), but our model can be adapted in this manner.

Future Work

A direct means of testing whether or not sequence information is truly being put to use during view selection is to alter the sequence of images in some way that preserves the observer's ability to recover 3-D form. If performance is affected by this transformation of the directed sequence of views, we can infer that the directed viewing sequence presented to the observer is relevant to the task (rather than simply the 3-D information computed via object motion). In previous work, we imposed a local scrambling operation on our paper clip sequences that has the qualities described above and used this manipulation to look for a direct influence of sequence order on image recall (Balas, 2007). Preliminary results indicate that this local-scrambling manipulation does have an effect on object recognition, suggesting that object motion does more than support the recovery of shape. As we stated earlier, however, any manipulation of a coherent spatiotemporal stimulus introduces profound low-level changes that may overshadow or interfere with the processes that we wish to isolate. Further efforts to refine sequence-scrambling designs are needed to make it possible for clear conclusions to be drawn regarding the role of perceived motion as opposed to a time-blind process that may operate only over sequence order.

A related issue that our model has not addressed is exploring the scale (in sequence or in time) of the effects of sequence order on canonicality judgments. The sequence-based terms in our model extend over only two or three frames from any given point in the sequence, leaving it unclear whether longer-range processing in time has any effect. To the best of our knowledge, this remains an open question in dynamic object perception more generally. The local-scrambling manipulation that we described above may be an interesting way to proceed, since it permits disruption of the ordered sequence of views at arbitrary scales. Similar manipulations have been used in recent studies (Vuong & Schultz, 2008) but have not been used to explore this question regarding the scale of temporal effects on object perception. Also, systematic exploration of the role of speed in object perception and recognition may help to define the natural scale for appearance integration over time and to disentangle temporal factors from sequence-based factors.

Finally, there is a wide range of factors that may influence canonicality that we have explicitly ignored here. Some of these have already been named, such as the influence of top-down knowledge concerning the salience of particular surface features of the object (faces or text) or the function of familiar objects (an upside-down chair is bad for sitting, and thus a "bad" view, despite potentially having the same geometric properties as a right-side-up chair under our model). At present, we cannot speak to the relative weighting given to the purely bottom-up factors that we have investigated here and these more complex cognitive aspects of view canonicality. An additional interesting question that we have not examined is the role of active versus passive experience of an object's appearance. The effects of active exploration on object recognition have been examined previously (Harman, Humphrey, & Goodale, 1999; James, Humphrey, & Goodale, 2001; James et al., 2002), but, to our knowledge, neither the properties of the sequences defined by observers during object exploration nor the effect of active manipulation on canonicality judgments have been determined. Investigating how these processes interact would be a highly valuable contribution to our understanding of object perception.

Overall, our key contribution is to demonstrate the utility of a modeling approach for studying view canonicality, complementing previous work with a computational analysis of how sequence order affects view selection. Our results are in accord with those of existing research, suggesting a relevant (but secondary) role for object motion in determining which views of a novel object are preferred. The present results raise many interesting and important questions for future work.

CONCLUSIONS

Our results suggest that adult observers do extract prototypical views, or keyframes, from sequences of novel moving objects after very little exposure. Subjective and objective measures of view canonicality are generally in agreement, and a simple form-based model of foreshortening provides a surprisingly good fit to the data. This model is substantially enhanced by the inclusion of dynamic information concerning sequence order, however, suggesting that observed object motion may provide information for canonicality judgments beyond a robust estimate of form.

AUTHOR NOTE

B.J.B. was supported in this work by an NDSEG fellowship. We offer special thanks to Jim DiCarlo, Eric Grimson, and Maggie Shiffrar for their comments on an early version of the manuscript. Correspondence concerning this article should be addressed to B. J. Balas (e-mail: bjbalas@gmail.com).

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Balas, B. J. (2007). *Learning about dynamic objects and recognizing static form*. Unpublished doctoral dissertation, Massachussetts Institute of Technology, Cambridge.

Balas, B. J., & Sinha, P. (2008). Observed object motion induces increased generalization and sensitivity. *Perception*, **37**, 1160-1174.

Blanz, V., Tarr, M. J., & Bülthoff, H. H. (1999). What object attributes determine canonical views? *Perception*, **28**, 575-600.

Blum, H. (1967). A transformation for extracting new descriptors of form. In W. Whaten-Dunn (Ed.), *Models for the perception of speech and visual form* (pp. 362-380). Cambridge, MA: MIT Press.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436.

Bricolo, E., Poggio, T., & Logothetis, N. (1997). *3D object recognition: A model of view-tuned neurons.* In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems 9* (pp. 41-47). Cambridge, MA: MIT Press.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, **44**, 108-132.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*, **89**, 60-64.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical–theoretic approach* (2nd ed.). New York: Springer.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, **33**, 261-304.

Cox, D. D., Meier, P., Oertelt, N., & DiCarlo, J. J. (2005). "Breaking" position-invariant object recognition. *Nature Neuroscience*, **8**, 1145-1147.

Denton, T., Demirci, M. F., Abrahamson, J., Shokoufandeh, A., & Dickinson, S. (2004). Selecting canonical views for view-based 3-D object recognition. *Proceedings of the 17th International Conference on Pattern Recognition* (pp. 273-276). Philadelphia: Drexel University Press.

Edelman, S. (1999). *Representation and recognition in vision*. Cambridge, MA: MIT Press.

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, **32**, 2385-2400.

Farah, M. J., & Hammond, K. M. (1988). Mental rotation and orientation-invariant object recognition: Dissociable processes. *Cognition*, **29**, 29-46.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, **103**, 18014-18019.

Foldiak, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, **3**, 194-200.

Grabner, M., & Bischof, H. (2005). Extracting object representations from local feature trajectories. In D. Chetverikov, L. Czuni, & M. Vincze (Eds.), *Proceedings of the Joint Hungarian–Austrian Conference on Image Processing and Pattern Recognition* (Vol. 192, pp. 265-272). Vienna: Austrian Computer Society.

Harman, K. L., & Humphrey, G. K. (1999). Encoding "regular" and "random" sequences of views of novel three-dimensional objects. *Perception*, **28**, 601-615.

Harman, K. L., Humphrey, G. K., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology*, **9**, 1315-1318.

Humphrey, G. K., & Jolicœur, P. (1993). Visual object identification: Some effects of image foreshortening, monocular depth cues, and visual field on object identification. *Quarterly Journal of Experimental Psychology*, **46A**, 137-159.

James, K. H., Humphrey, G. K., & Goodale, M. A. (2001). Manipulating and recognizing virtual objects: Where the action is. *Canadian Journal of Experimental Psychology*, **55**, 113-122.

James, K. H., Humphrey, G. K., Vilis, T., Corrie, B., Baddour, R., & Goodale, M. A. (2002). "Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, & Computers*, **34**, 383-390.

Koenderink, J. J., & van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics*, **32**, 211-216.

Lawson, R., & Humphreys, G. W. (1996). View specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 395-416.

Lawson, R., Humphreys, G. W., & Watson, D. G. (1994). Object recognition under sequential viewing conditions: Evidence for viewpoint-specific recognition procedures. *Perception*, **23**, 595-614.

Liu, T. (2007). Learning sequences of views of objects: The effect of spatiotemporal coherence on object recognition. *Perception*, **36**, 1320-1333.

Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology*, **4**, 401-414.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, **5**, 552-563.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

Newell, F. N., Wallraven, C., & Huber, S. (2004). The role of characteristic motion in object categorization. *Journal of Vision*, **4**(2, Art. 5), 118-129. doi:10.1167/4.2.5

Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 135-151). Hillsdale, NJ: Erlbaum.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.

Perret, D. I., Harries, M. H., & Looker, S. (1992). Use of preferential inspection to define the viewing sphere and characteristic views of an arbitrary machined tool part. *Perception*, **21**, 497-515.

Perry, G., Rolls, E. T., & Stringer, S. M. (2006). Spatial vs temporal continuity in view invariant visual object recognition learning. *Vision Research*, **46**, 3994-4006.

Peters, G., Zitova, B., & von der Malsburg, C. (2002). How to measure the pose robustness of object views. *Image & Vision Computing*, **20**, 249-256.

Stone, J. V. (1998). Object recognition using spatiotemporal signatures. *Vision Research*, **38**, 947-951.

Stone, J. V. (1999). Object recognition: View-specificity and motion specificity. *Vision Research*, **39**, 4032-4044.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review*, **2**, 55-82.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1494-1505.

Tarr, M. J., & Kriegman, D. J. (2001). What defines a view? *Vision Research*, **41**, 1981-2004.

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London B*, **203**, 405-426.

Vuong, Q. C., & Schultz, J. (2008). Dynamic objects are more than the sum of their views: Behavioral and neural signatures of depth rotation in object recognition [Abstract]. *Journal of Vision*, **8**(6, Abstract 39), 39a. doi:10.1167/8.6.39

Vuong, Q. C., & Tarr, M. J. (2004). Rotation direction affects object recognition. *Vision Research*, **44**, 1717-1730.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, **45**, 205-217.

Wallis, G. (1998). Spatio-temporal influences at the neural level of object recognition. *Network*, **9**, 265-278.

Wallis, G. (2002). The role of object motion in forging long-term representations of objects. *Visual Cognition*, **9**, 233-247.

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences*, **98**, 4800-4804.

Wallraven, C., & Bülthoff, H. H. (2001, December). *Automatic acquisition of exemplar-based representations for recognition from image sequences.* Paper presented at the IEEE Computer Society's Computer Vision and Pattern Recognition Conference, Kauai.

Wang, G., Obama, S., Yamashita, W., Sugihara, T., & Tanaka, K. (2005). Prior experience of rotation is not required for recognizing objects seen from different angles. *Nature Neuroscience*, **8**, 1568-1575.

(Manuscript received July 28, 2008;
revision accepted for publication November 8, 2008.)