Perceptual Salience of Non-accidental Properties

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By

Eric Weismantel

Graduate Program in Psychology

The Ohio State University

2013

Master's Examination Committee:

Professor James Todd “Advisor”

Professor Alex Petrov

Professor Julie Golomb
Abstract

The different types of non-accidental properties have been commonly treated as having equal perceptual salience. In an effort to test this hypothesis, discrimination thresholds were measured for four different types of transformations that altered the Euclidean, affine, projective, or topological structure of two basic 2D objects. The results confirm that there are significant differences in the relative perceptual salience of these transformations such that the topological change was the most salient followed by projective change which was followed by affine change. Changes in Euclidean structure were the most difficult to detect. To determine how well existing metrics account for the observer data, these thresholds were compared to the predictions from pixel distance, pixel angle, correlation, total area, and Hausdorff metrics. The results suggest that all the tested metrics are poor predictors of observer data for the perception of non-accidental properties.
Vita

2006 ................................................. Ankeny High School

2010 .................................................. B.S. Psychology & Philosophy, Central College

Fields of Study

Major Field: Psychology
Table of Contents

Abstract .................................................................................................................. ii
Vita ......................................................................................................................... iii
Fields of Study ..................................................................................................... iii
Table of Contents .................................................................................................. iv
List of Tables ......................................................................................................... v
List of Figures ....................................................................................................... vi
Methods ................................................................................................................ 6
   Observers ............................................................................................................ 6
   Apparatus .......................................................................................................... 6
   Stimuli ................................................................................................................ 7
   Procedure ......................................................................................................... 9
Results .................................................................................................................. 11
Discussion ............................................................................................................ 17
References ............................................................................................................ 21
List of Tables

Table 1: 75% thresholds of stimulus intensity for all four transformation types. These values are given in terms of the Hausdorff metric that is equal to the intensity of change.

Table 2: Correlations of all converted metrics with observer data.
List of Figures

Figure 1: A drawing of a case whose contours possess several types of non-accidental properties (see Biederman 1987) ..........................................................2

Figure 2: The displacement of the short edge is shown on the left and the displacement of the long edge is shown on the right .........................................................5

Figure 3: The two original objects. They consisted of a square or rectangle with curvature added to either the top or bottom and two triangles added to the opposite edge. .............8

Figure 4: Example stimuli of a test shape and the four different transformations. ...........9

Figure 5: An example trial sequence for the curvature condition. Observers were shown a fixation point, a test shape, a mask, the first alternative shape (match to the test shape in this example), a mask, the second alternative shape (a curvature transformation in this example), and a mask. All presentations were for 300 ms. ........................................10

Figure 6: The proportion of hits as a function of stimulus intensity for all four conditions .................................................................12

Figure 7: Graphs of 75% thresholds of stimulus intensity for all four transformation types .................................................................13

Figure 8: Graphs of scaled threshold predictions for all measures based on a 5 pixel change of the vertical scaling condition .........................................................16
Figure 9: A fixed 5 pixel displacement for both the scaling (left) and topological (right) changes overlayed on the same base stimuli (Notice the area of the grey edges compared to the area of the hole on the right).
Literature Review

The question of object invariance has long been an issue of vigorous study in the field of vision research. One class of largely invariant object features that have received significant study are non-accidental properties. These are useful properties for object recognition because they typically remain invariant across different viewpoints. When these properties are present in a 2D image it is likely that the depicted 3D object possesses those properties as well, with the exception of accidental alignments. For example, if lines in a 2D image are parallel, there is a high likelihood that the lines are also parallel in 3D.

A helpful example to demonstrate the viewpoint invariance of non-accidental properties is to think about viewing a coin. When a coin is viewed from almost all orientations, it appears to have curved edges. However, if the coin is rotated so that the observer is looking directly at the thin edge of the coin, it will appear as if the edges are now straight. This is the only accidental alignment that can make the curved edges seem straight. In this example, there are very few viewpoints that make the edges of the coin appear straight, so the property of curvilinearity is highly viewpoint invariant.

There are several types of non-accidental properties that can be present for an object (see Figure 1) For example, the object can have many different types of vertices as determined by its pattern of cotermination, the edges of an object can be straight or...
curved as determined by its smooth continuation (the case shown in Figure 1 below has both straight and curved sections based on the continuation of its line segments), or an object can have different arrangements of lines based on its parallelism (the case shown below has parallelism present for all of its sides).

![Symmetrical Object Diagram](image_url)

Figure 1: A drawing of a case whose contours possess several types of non-accidental properties (see Biederman 1987)

There is a large body of evidence that supports the idea that non-accidental properties are salient enough that they can be used to make object recognition decisions (e.g., Biederman, 1987; Biederman & Ju, 1988; Biederman & Cooper, 1991). Non-accidental properties are especially useful for object recognition because they remain relatively stable over different vantage points. Studies have consistently found that observers are more sensitive to changes in non-accidental properties than changes in metric properties (e.g, Biederman & Barr, 1999; Vogels, Biederman, Bar, & Lorincz, 2001). Specifically, Biederman & Barr (1999) had observers perform a same/different
task for novel 3D objects at different orientations. In some trials, an object’s metric properties were manipulated (e.g., a scaling change) while in other trials an object’s non-accidental properties were manipulated (e.g., a skewing change). The metric property and non-accidental property differences were scaled to make them equally detectable when the objects were at the same orientation. Reaction times were higher and accuracy was lower when the objects had a metric property difference as compared to trials in which the difference was in a non-accidental property.

Most prior discussions of these properties have treated them as a homogeneous set. These earlier studies have included several different types of non-accidental change but have not examined whether they are all perceptually equivalent. However, there are mathematical reasons to suspect a different treatment is needed as nonaccidental properties are not all equally stable over different types of object transformations (Chen, 1983, 2005). Chen hypothesized that the relative stability of object properties under different types of transformations is systematically related to their relative discriminability. The more transformations under which an object property remains invariant, the easier it will be to detect a change in that property. For example, the distinction between parallel and non-parallel contours remains invariant over affine transformations (e.g, stretching); the distinction between straight and curved contours remains invariant over all affine and projective transformations (e.g, skewing); and the topological structure (e.g., the pattern of connectivity) remains invariant over all affine and projective transformations.
A goal of the current experiment is to expand upon earlier work that examines whether the discriminability of non-accidental properties varies systematically with their relative stability under change. As a test of Chen’s hypothesis, Todd, Chen, & Norman (1998) had observers perform a match to sample task where a 3D wire-frame object’s Euclidean, affine, or topological structure were manipulated. All manipulations kept the 2D topology the same while changing the orientation of one of the 4 line segments of the wire-frame object. The metric difference of the three types of manipulations was equivalent. The topological changes had the highest accuracy and the lowest reaction times followed by affine changes that had lower accuracy and longer reaction times. Euclidean changes had the lowest accuracy and the highest reaction times.

This work provided support for Chen’s ideas, but there is an important obstacle that needs to be overcome to make valid conclusions from studies such as this that use different types of object transformations. In order to compare discrimination performance for different types of transformations, it is necessary to measure each type of shape change with a common metric. This is typically achieved by describing images as vectors in a high dimensional space, as defined by pixel intensities or Gabor wavelets (e.g. Biederman & Bar, 1999; Vogels et al, 2001). Differences between images can then be measured using a vector dot product, calculating the Euclidean distance between vector endpoints, or correlating the two images. To better understand the nature of these measures it is useful to consider the two possible distortions of a rectangle shown below (see Figure 2), one that involves the displacement of a long edge and another that involves the displacement of a shorter edge. Note that the displacement of the long edge
affects a greater number of pixels (or Gabor wavelets) than an identical displacement of the short edge, and it would therefore produce a larger shape change using any of the standard metrics described above. An alternative metric called the Hausdorff measure (Moran 1946) scales shape changes based on the magnitude of the maximally displaced point on an object. This metric differs from the pixel and gabor metrics mentioned above because it requires point to point correspondence. According to the Hausdorff metric both of the shape changes shown below are equally large.

Figure 2: The displacement of the short edge is shown on the left and the displacement of the long edge is shown on the right

In the present study, we examined the systematic differences in the detection of non-accidental properties by measuring the relative discriminability of four types of structure modifications (metric, affine, projective, and topological). Based on Chen’s hypothesis and the earlier work of Todd, Chen, & Norman (1998), we expected to find significant differences in the relative perceptual salience of different types of transformations. Specifically, we hypothesized that changes in topological properties will be the easiest to detect followed by projective properties that will be easier to detect than affine properties that will be easier to detect than metric properties. These
discriminability levels were compared to the existing computational models to see how well they accounted for the observer data.

**Methods**

The goal of this experiment was twofold. First we wanted to measure the differences in perceptual sensitivity for four different types of transformations. Second, we wanted to determine how well existing computational models accounted for the perceptual sensitivity data. We derived 75% thresholds for four different types of transformations using the method of constant stimuli with a 2AFC match to sample task. We compared perceptual sensitivity by computing the total area of change, pixel distance, correlation, and the maximum displacement (Hausdorff metric).

**Observers**

Four observers affiliated with The Ohio State University participated in the study; including the two authors and two who were naïve observers. All had normal or corrected-to-normal vision.

**Apparatus**

This study used 4 simple objects that were created by Autodesk 3DS Max Design 2011. They were presented on a 21” Dell CRT monitor at a resolution of 1280 x 1024 connected to a Dell Dimension 8300 computer. The viewing distance was 65 inches. The response collection and display timing were controlled with MatLab.
Stimuli

The objects consisted of a square or rectangle with curvature added to either the top or bottom and two triangles added to the opposite edge (See Figure 3). Four different transformations were performed on the objects to create our stimuli; A scaling transformation that either extended or shrunk the objects along the horizontal axis, a curvature transformation that added curvature to both sides, a skewing transformation that affected the parallelism of the sides, and a topological transformation that added a vertical white rectangle of varying intensity to separate the objects in two (see Figure 4). The two original objects were distorted gradually in order to create 24 versions of each base object with a different magnitude of change along the horizontal axis. These transformations produced 700 stimuli. The two test shapes and 700 transformed shapes were also flipped both horizontally and vertically to create four different versions of each test shape and each transformed shape.

When generating our stimuli we modified the intensity of each transformation in terms of the Hausdorff metric for ease of calculation. By generating the stimuli based on the Hausdorff measure, we were able to have a common currency by which to compare the observer results. The intensity of the transformation for a stimulus is the same value as the maximum displacement for that stimulus (e.g., if a stimulus has a metric change intensity of four pixels, then the maximum displacement value for that stimuli is 4 pixels). The topological condition was unique because it had intensities that were less than 1. Pilot testing showed that observers were at ceiling for any topological change of more than 1 pixel. In order to get psychometric functions for this condition, we needed a
range of accuracies, so we used subpixel displacements for all topological stimuli shown to observers. In order to get psychometric functions for this condition, we needed to use subpixel displacements. The basic idea behind this technique is to mimic the optical effect of a higher resolution display by manipulating the intensities of individual pixels. For example, one pixel at 1/4 the maximum intensity would be optically equivalent to one pixel at full intensity on a display with four times the display resolution. The standardized scale also gave us a way to standardize the other metrics so that we were able to compare them against each other.

Figure 3: The two original objects. They consisted of a square or rectangle with curvature added to either the top or bottom and two triangles added to the opposite edge.
Procedure

Observers viewed the displays binocularly in a room with controlled lighting. Participants performed a match to sample task where they were asked to decide which object matched the previously presented sample. Subjects were not informed prior to the task regarding which kind of changes the stimuli may go through. Feedback was shown after each trial. The participants saw some stimuli multiple times in different pairings.

Each trial began with a fixation point presented at the center of the screen. This was followed by the test shape at a random point on the screen for 300 ms and a mask for 300 ms. The test shape and mask were followed by the first alternative shape for 300 ms on the left side of the screen, a mask for 300 ms, and the second alternative shape for 300

Figure 4: Example stimuli of a test shape and the four different transformations.
ms on the right side of the screen. The test and alternative shapes were all rotated between 10 and 35 degrees. The trial concluded with another mask for 300 ms. Observers were asked to respond as quickly as possible to indicate which alternative shape matched the test shape (See Figure 5).

![Figure 5: An example trial sequence for the curvature condition. Observers were shown a fixation point, a test shape, a mask, the first alterative shape (match to the test shape in this example), a mask, the second alterative shape (a curvature transformation in this example), and a mask. All presentations were for 300 ms.](image)

Before completing experimental blocks, each observer ran two practice sessions of 256 trials each. They were informed before these sessions about the four types of change each shape could undergo. This information was provided because we wanted the observers to be aware of what they should be looking for in each image to more easily compare the salience of these different types of change. After completing the practice sessions, each observer participated in 5 test sessions of 256 trials with each session being separated by at least 30 minutes to avoid observer fatigue. Observers received simple feedback after each trial in the form of an onscreen message indicating whether the response they chose was correct or incorrect. The sessions were divided into 4 blocks
with each block having each condition appear twice. The order of the conditions was counterbalanced between blocks.

Results

Logistic regressions were computed for each condition for each individual observer using SigmaPlot. Based on these regressions, 75% thresholds were computed. The data for all 4 observers was strongly correlated (r’s > .95, p < .01). Table 1 gives a summary of these combined data using the average for all observers in terms of the Hausdorff metric. Figure 6 shows the regression curves for one of the participants.

<table>
<thead>
<tr>
<th></th>
<th>Avg</th>
<th>Vert</th>
<th>Horiz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td>6.92</td>
<td>7.58</td>
<td></td>
</tr>
<tr>
<td>Parallel</td>
<td>4.66</td>
<td>4.76</td>
<td></td>
</tr>
<tr>
<td>Curvature</td>
<td>1.73</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td>Topological</td>
<td>0.20</td>
<td>0.22</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: 75% thresholds of stimulus intensity for all four transformation types. These values are given in terms of the Hausdorff metric that is equal to the intensity of change.
Figure 6: The proportion of hits as a function of stimulus intensity for all four conditions

Figure 7 shows the graphs of the average 75% thresholds. From these graphs it is clear that the topological change was the most salient and the metric change was the least salient. The threshold values for the scale condition were the largest followed by the values for the parallel condition followed by the values for the curvature condition followed by the values for the topological condition. There is no significant difference between the vertical and the horizontal conditions.
For the second question of interest, we needed to compare the predictions of common shape metrics to the actual observer data. In the next step of data analysis, we calculated three different types of shape metrics based on the threshold data. We chose these metrics because they are all commonly used metrics found widely in the literature. We used Matlab to create a program that maximally aligned the unmodified stimuli with the modified stimuli. To compute the total area of change, we summed the displacement of all shifted points between a test stimulus and a transformed stimulus.

Another common metric is to calculate the pixel distance between two images. For this calculation, we started by considering each image as a vector in high dimensional space, where each individual pixel defined a dimension, and the intensity of the pixel defined a specific position along that dimension. To calculate the Euclidean (pixel)
distance between the corresponding vector endpoints, we used this equation where $p$ is the point on the test stimulus and $q$ is the shifted point on the transformed stimulus and $d$ is the distance between these two points:

$$d(p, q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$$

The third common shape metric that we used is the correlation of the two images. To calculate the correlation between the two images, we calculated the correlation coefficient of the matrix of a test stimulus and a matrix of a transformed stimulus. For the fourth metric, pixel angle, we took the ArcCos of this correlation to calculate the angle between the two vectors.

To determine how well the metrics capture observer judgments, we needed to find a way to calculate the predictions of each measure on a standardized scale. We computed standardized predictions by calculating how much Hausdorff change is required for all of the conditions to have the same amount of difference (in terms of the total area, correlation, and pixel distance metrics) as a fixed 5 pixel displacement of the vertical scaling stimuli (the condition with the highest thresholds). For example, to approximate how much Hausdorff change is needed for the horizontal scaling stimuli to have the same amount of change in total area as the fixed 5 pixel displacement of the vertical scaling stimuli, we calculated that a fixed 5 pixel displacement of the vertical scaling condition had a total area change of 3420 pixels. The total area of change metric was then computed for all of the horizontal scaling stimuli. These values were used to approximate that according the Hausdorff metric a displacement of 8.05 pixels was needed to have a total area change of 3420 pixels. The scaling of all metrics according to the Hausdorff
metric gave us the predicted maximum displacement for the total area, correlation, and pixel distance measures. If observer’s judgments had been based on the Hausdorff metric, the thresholds would have been the same in all conditions.

Figure 8 shows the graphs of these predictions converted to a log scale. From the graphs, it is evident that the predicted ordering of the different types of change for the total area, pixel distance, and correlation metrics should go from scale changes as the easiest, parallel as more difficult than scale, curvature as more difficult than parallel, and topological change as the most difficult. Those 3 graphs also show a difference between the vertical and horizontal conditions where the vertical condition should be easier than the horizontal condition. The Hausdorff measure predicts that all measures will be the same difficulty and there will be no difference between the horizontal and vertical conditions.

The converted metric values were correlated with the average observer threshold values. The pixel count ($r = -.901$), correlation ($-.826$), pixel angle ($r = -.790$), and pixel distance ($r = -.790$) measures all showed significant negative correlations ($p < .05$) (see table 3). Based on this analysis, it is evident that the relative discriminability of the different types of shape change is the opposite of what would be predicted by the pixel count, correlation, and pixel distance shape difference metrics (i.e., the changes that produce the highest shape difference measures resulted in the highest discrimination thresholds). The Hausdorff metric is also not a good fit for the data as the thresholds are not all the same for the different types of transformations.
Figure 8: Graphs of scaled threshold predictions for all measures based on a 5 pixel change of the vertical scaling condition (converted to log scale)

Correlations ($r$)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total area of change</td>
<td>-0.901</td>
</tr>
<tr>
<td>Euclidean distance</td>
<td>-0.79</td>
</tr>
<tr>
<td>Pixel angle</td>
<td>-0.79</td>
</tr>
<tr>
<td>Correlation</td>
<td>-0.826</td>
</tr>
</tbody>
</table>

Table 2: Correlations of all converted metrics with observer data
Discussion

The present research had two major goals; first we wanted to determine the relative perceptual salience of non-accidental object properties. Second, we sought to compare common shape metrics to evaluate how well they fit observer data. To address the first goal, we measured discrimination thresholds for four different types of transformations that altered the topological, projective, affine, or Euclidean structure of a simple 2d object. These thresholds in terms of the Hausdorff metric were measured for two different versions of the object, one that had more width (horizontal condition) and the other that had more height (vertical condition). The data supported the hypothesis of Chen (1983) regarding the relationship between perceptual stability and discriminability. The topological change was the most salient followed by projective, affine, and metric change. Regarding the horizontal/vertical distinction, the general trend was that the vertical change was not significantly more salient than the horizontal change.

For the second goal, we analyzed how well the common metrics were able to fit the observer data. We computed correlations of all measures converted to the Hausdorff metric with the observer data. Our analysis showed that all three of the standard metrics (total area of change, pixel distance, and correlation,) were extremely poor fits for the data because they had strong negative correlations. The Hausdorff was a better fit than the three standard metrics, but it was still poor because it predicted that all the conditions would have the same thresholds.

A brief analysis of the metrics demonstrates possible points of problems for all the metrics. The three standard pixel metrics face problems because they all produce
larger values for scaling changes than for topological changes as they merely calculate
the pixels that have changed between the images. According to these metrics, the scaling
change should be the most salient out of the four types of changes in this study because it
has the greatest amount of pixels that are changed.

As a further example of this, Figure 9 shows a fixed 5 pixel displacement for a
scaling change compared to a fixed 5 pixel displacement for a topological change. In the
scaling change, the entire edge of the object will scale outward which causes a large pixel
difference. In the topological change, only the center of the object will change which
causes a much smaller pixel difference. Based on these three metrics, the scaling change
should be much more salient than the topological change because there has been a
substantially larger change in pixels. The observer data directly contradicts these
predictions with the scaling change being the least salient out of the four types of change.
In fact, the strong negative correlations show that the observer data is the exact opposite
of what is predicted by the metrics which makes these metrics virtually useless when it
comes to scaling image difference involving non-accidental properties. The standard
metrics also predict a horizontal/vertical distinction that isn’t found in the observer data.
In the present study all stimuli in the vertical condition had a 1/3 larger change than those
in the horizontal condition according to the total area, pixel distance, and correlation
metrics. The observer data doesn’t show a consistent pattern of the vertical change
condition being easier than the horizontal condition. Because these three metrics are
unable to capture the distinctions found in the observer data, they will be very poor
predictors of the relative perceptual salience of different types of change.
The Hausdorff metric is better able to capture the distinctions found in the observer data. Because the Hausdorff metric is focused on the maximally displaced point, the actual amount of changed pixels between the horizontal and vertical images is irrelevant to this metric. It predicts that there will be no difference between the horizontal and vertical conditions. This prediction is supported by the observer data. However, the Hausdorff metric also has a large problem that keeps it from being a useful metric. In the example above, it would predict that a 5 pixel scaling change will be equally as salient as a 5 pixel topological change. This prediction is not aligned with observer data, but it is a much closer prediction than the other four measures that predict a relationship in the wrong direction (scaling being more salient than topological).

This brief analysis of these metrics shows where both of the metrics face problems when predicting the perceptual salience of non-accidental properties. The failure of the three standard metrics is particularly problematic as these measures are widely used for scaling image differences in many domains of cognitive science. The
basic assumptions of these metrics need further study to determine if their efficiency outweighs some glaring issues. Future research is also needed to devise and compare alternative metrics that might be better able to capture the strength of the Hausdorff metric in correctly predicting that a scaling change will not be more salient than affine, projective, and topological changes while avoiding treating all of the non-accidental properties in the same manner.
References


Chen, L. (1983). What are the units of figure perceptual representation?. University of California, Irvine, School of Social Sciences.


