Object Constancy refers to our ability to recognize different two-dimensional images as representations of a particular three-dimensional object (SOGI, in this case).

The object in A can be recognized over:
- C – translation, as in movement of the object or gaze
- D – changes in size (distance from observer)
- E – changes in lighting (from upper left to upper right)
- F – shifts in picture-plane orientation
- G – shifts in depth-plane orientation
- H & I – viewpoint shifts caused by SOGI turning in depth or observer moving around the object.

- Note that changes in viewpoint might cause parts of the object to become visible or to occlude other parts of the object.
  - The star-shaped side of SOGI shown in H occludes more of the figure than in A.
- To a computer, all of the images in figure 1 are just collections of numbers that specify the intensity of each small dot, or pixel.
  - Teaching a computer to distinguish between the figures in A and B would be relatively trivial because all that would be required would be point-by-point comparisons of pixels, but teaching the computer that C through I are also the same figure as A would be much more difficult.
  - Pixel-by-pixel comparisons would also find A to be different from images in C through I.

Template Matching

- Before considering 3-D figures, let us examine possible representations of two-dimensional patterns.
- Template theories propose that patterns are not really analyzed at all—templates are holistic entities that are compared to input patterns to determine amount of overlap.
- Template matching works well in pattern recognition machines that read letters and numbers in standardized, constrained contexts (scanners that read your account number off from checks, machines that read postal zip codes off from envelopes).

Template Matching: Strengths

- There is abundant physiological support that simple features (lines and edges of particular orientations) are represented in the nervous system with template-like receptive fields in the visual cortex.
- They are amazingly reliable. If the to-be-encoded stimulus is present, it’s template will become active.

Template Matching: Weaknesses

- The difficulty with template matching as a model for perception is that contexts are rarely constrained.
- For instance, slight deviations in shape, size, orientation, would prevent template matchers from reading even the limited number of letters (26) in English
  - They are not inherently view invariant. For every different possible view, there would have to be a different template (replication). As such, template representations are uneconomical.
Template Matching: More Weaknesses

- Normalization with regard to size, shape, and orientation is one possible way around the problem, but individuals can read written messages that contain gaps in letters and variations in writing instruments, so the number of normalizations would be enormous.
- Even with replication and normalization, it would be difficult to represent the third dimension (depth) with template matching (since the retina is a two-dimensional receptor array).
- Standard templates contain no information about whole-part relations—the only two levels of representation are “whole entity” and pixel (or receptor).

Feature Comparison Models

- For several decades, the most popular class of shape representation was **feature lists**: a symbolic description consisting of a limited simple set of attributes.
- According to this view, perceived shape is defined by the set of features that an object possesses.
  - The number of shared features can represent similarity/dissimilarity between two shapes.

**Feature detection** models assume that perceptual systems detect the presence or absence of particular features (as binary variables, 0 or 1).

**Feature analysis** models also analyze displays down into component features, but they allow gradations between 0 and 1.

Gibson (1969)

- One of the best-known feature theories is Eleanor Gibson’s (1969) account of perceptual learning of the alphabet by children.
- She asserted that perceptual learning occurred through the discovery of distinctive features between letters.
- Children first confronted with an “E” and an “F” may not be aware of how the two differ.
- The distinctive feature is the lower horizontal line (present in the E but not in the F).

Egeland (1975) showed that pre-kindergartners could be taught to distinguish between confusable letters (R-P, Y-V, G-C, Q-O, M-N, K-X) when the distinctive features were brought to their attention (by highlighting the distinctive feature in red).

- During training the distinctive feature was gradually changed back to black to match the rest of the letter.
- Another group of children was trained with black letters exclusively; they received only correct/incorrect feedback about their choices (“point to the R”).
- The group trained to learn the distinctive features performed better immediately after training and one week later (delayed test), even though the features were not highlighted during testing.
• Another example of how distinctive features can be used to enhance recognition is in face-recognition tasks.
• Distinctive features can be enhanced to create caricatures using a computer program.
• The program takes photographs and exaggerates the distinctive features of the face.
  - For example, a face of someone with large ears and a small nose would be converted into a caricature of someone with really large ears and a very tiny nose.
• When students were shown line drawings of their acquaintances, they identified their friends quicker when shown the caricatures as opposed to the accurate line drawing.

Advantages of Feature Analysis
• Feature List approaches solve the whole-part problem by including the parts as features of the whole. For instance, the features of a head might be “having two eyes,” “having a nose,” “having a mouth,” etc.
• Features also seem capable of solving the 3-D problem, since features can refer to 3-D qualities (spherical, pyramidal), just as easily as 2-D features (circular, triangular).

Disadvantages of Feature Analysis
• It is not obvious how various features would be “computed.” It is one thing to say that it takes place, but quite another to specify the mechanism.
• For many complex figures, it is difficult (if not impossible) to specify a defining list of features. So far, feature lists have only been developed for relatively simple stimuli (e.g., letters and numbers).

One of the earliest (and most colorful) of the feature theories is Pandemonium, which was introduced by Selfridge (1959). Pandemonium consists of four separate layers: each layer is composed of “demons” that are specialized for different tasks. For instance, there might be a cognitive demon for each of the 26 letters in the English alphabet, each responsible for recognition of particular letters.

Disadvantage of Pandemonium
• It is entirely data-driven, yet perception has many examples of “context effects,” which are handled better by conceptually driven, top-down processes.
• Data-driven processing refers to processing that is driven by the stimulus pattern—there is no manner in which context effects can be easily built into the model.
• In Pandemonium, the patterns to be recognized come from the image demons at the bottom, which are then passed on to higher and higher level demons.
Examples of Context Effects:
Top-Down Processing

- One of the better phenomena for demonstrating contextual effects is the word-superiority effect.
- Logically, one might conceive of the letters that compose words to be independent units of text, each one identified separately from the others.
- One might also suppose that words are read on the basis of the letters composing them.
- The fact that letters can be more quickly and accurately identified when they are imbedded in meaningful words than in meaningless letter strings has been demonstrated since 1886 (Cattell).

Cattell’s Original Demonstration of the Word Superiority Effect (1886)
- He compared the number of letters that subjects could report from 10-ms exposures to English words vs. non-words. His task was akin to:
  - HOW MANY LETTERS CAN YOU REPORT NOW?
  - HWO NMYA RSTELTE NCA OYU RPTERO NWO?

- Instead of differences in the number of letters identified, it could well be the number of letters remembered that differed. Providing the letters in words allows for convenient “chunking.”

Modern Word Superiority Effect
(Reicher, 1969; Wheeler, 1970)
- The effect, as it is now studied, is that single letters can be identified more quickly and with a higher level of accuracy when they are imbedded in real words.

Word-Superiority Effect

<p>| Precuing refers to telling (orally) the participants the two possible letters in advance of the stimulus. |</p>
<table>
<thead>
<tr>
<th>Word Condition</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>Precue</td>
</tr>
<tr>
<td></td>
<td>74%</td>
</tr>
<tr>
<td>Nonword Condition</td>
<td></td>
</tr>
<tr>
<td>ORWD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>58%</td>
</tr>
<tr>
<td>Letter Condition</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>59%</td>
</tr>
<tr>
<td>Target Display (50 ms)</td>
<td>Mask and Response Alternatives</td>
</tr>
</tbody>
</table>

McClelland and Rumelhart: Interactive Activation Model

Each letter can be represented by a subset of 12 possible segments. Therefore, there are 12 feature nodes at each of the 4 letter positions for a total of 48 feature nodes.
Interactive Activation Model (IAM): Feature Level

- The feature level consists of 12 features for each of 4 letter positions: a total of 48 feature nodes.
- Projection from the feature level to the letter level is entirely bottom-up. There is no feedback from higher levels.

IAM: Letter Level

- The letter level consists of 26 letters at 4 possible positions: a total of 104 letter nodes.
- The letter nodes all get excitatory input from all feature nodes that represent segments contained by the letter and inhibitory inputs from all feature nodes that are not present in the letter.
- As such, the node representing a horizontal line at the top excites the letter nodes for A, B, C, D, E, F, G, I, O, P, Q, R, S, T and Z. It inhibits the representations of H, J, K, L, M, N, U, V, W, X, and Y.
- When all segments for a particular letter present, the node for that letter will be highly active.

IAM (Continued)

- Other letter nodes will be active for a given position, depending on the number of shared features.
- For instance, the segments that comprise the McClelland and Rumelhart A will also activate the H, since the only difference between an A and an H is the presence of the top horizontal for the A but not the H.
- To sharpen the representation of letters, each letter inhibits the other 25 letter representations (this implements the winner-take-all rule).

IAM: Word Level

- The word level consists of a lexicon of more than 1000 4-letter words.
- Each word receives excitatory input from the letter level for its constituent letters at the four positions and inhibitory inputs to the other 25 letters at each of the four positions.
- As such, the letter A in the first position sends excitatory inputs to the word nodes for ABLE and ACTS, whereas B sends inhibition to the representations of both words. The node for A in the first position inhibits the word-level nodes of BACK and GAVE.
- All word nodes are mutually inhibitory, so it is a winner-take-all network.
IAM: Accounting for the Word Superiority Effect

- It is the excitatory feedback from the word level to the letter level that is responsible for the word-superiority effect!
- Inhibitory feedback from the word level to the letter level is not proposed.

IAM: Accounting for the Pseudoword Superiority Effect

- The pseudoword superiority effect is probably due to the fact that pronounceable non-words will have considerable activation at the word level.
- Furthermore, that activation will feed back to the letter level:
  - Even though MAVE may have no word node, it will partially activate the word nodes for MATE, MAKE and CAVE, with these partial activations feeding back to the letter level.
- Since no word in the lexicon is active, there is no “winner” to suppress activity.

Structural Models

- Structural descriptions are representations that contain explicit information about parts and about relationships between parts.
- They are generally depicted as a network in which nodes represent the whole object and its various parts and labeled links (or arcs) between nodes represent the relations between parts.

Shape Primitives

- Structural descriptions differ most widely in terms of the shape primitives that they possess.
  - Shape primitives are the fundamental elements or building blocks of a representation.
  - Marr has proposed that generalized cylinders make good 3-D shape primitives.
    - A standard cylinder is constructed by taking a circle of some diameter as its base and sweeping it along a straight axis some distance (the only variables are diameter and length).
    - A generalized cylinder allows the shape of the base to be something other than a circle (any closed 2-D figure), the axis that the base is swept through to be something other than straight, and the base to vary in size and shape as it is swept through its arc!
Illustration of five variables in constructing generalized cylinders. The central cube (A) can be modified to construct the 8 other geons shown by changing just one of five parameters: curvature of cross-sectional edges (B), cross-sectional symmetry (C and D), curvature of sweeping axis (E), diameter of sweeping rule (or cross-sectional size) (F and G), and aspect ratio (length of sweeping axis to the length of the largest dimension of the cross-sectional area) (H and I).

**Recognition by Components**

- Biederman (1987) used these generalized cylinders as his building blocks for representing objects ("geons," short for geometric ions).
- RBC (Biederman, 1987) proposes that objects are represented by a finite number (about 36) of shape primitives (called geons).
- These can be combined in different ways (different structural descriptions) to yield an infinite number of objects.
  - Since the structural description is included, RBC is a structural model.

- Biederman, Ju, and Clapper (1985) studied the perception of briefly (100 ms) presented partial objects that lacked some of their components.
- As more components (geons) are presented, RBC would predict better performance, since there would be a greater number of diagnostic matches to the object’s representation in memory.
  - The stimuli were line drawings of 36 common objects.
  - The total number of components that composed the objects (level of complexity) was varied at 2, 3, 6 and 9 (9 instances each).
  - Subjects received a list of item names prior to testing, but this probably had little effect on error rates or reaction times.
- It is important that object recognition be robust to occlusion, noise, and rotation in depth, so a demonstration that we can name incomplete objects would add credence to RBC theory.

The authors showed that error rates were low and reaction times were fast for objects that consisted of as few as 4 (of 9 components). There was some improvement as the number of geons increased, consistent with RBC.
• If recognition is based upon edge-based geons, then color, brightness, and texture should contribute little to recognition.
  – Biederman and Ju (1986) found equivalent reaction times and error rates for line drawings of objects vs. professional color photographs of the same objects.
  – This included results for objects for which color might be a major diagnostic attribute (bananas).

Recoverable vs. Non-recoverable Objects?

• Biederman (1987) suggested that the vertices contain information about the relations between geons, and he predicted that removing those line segments would be particularly harmful to recognition (indeed, parsing the figure down into component geons becomes difficult).
The line drawings in the middle column have line-segment information removed from them. The objects are easier to identify (recoverable) if the segment information is removed from between the vertices (as in the middle column) than if it is removed from the vertices themselves (unrecoverable, as in the right-most columns).

- Biederman and Blickle (1985) found that removing segments near the vertices most harmed object-naming performance (longer RTs and higher error rates).
- When amount of deletion is varied systematically, greater deletion has the biggest effect (65% deletion is associated with error rates and RTs that are higher for some specified exposure duration).
• Although the geons are volumetric representations, they are encoded by the features of their two-dimensional retinal images (lines, edges, and vertices). The properties from which geons are to be recognized are their “nonaccidental features.”

• These are aspects of the image structure that, if present, mean that it is very likely that they also exist in the object.
  – For example, if there is a straight line in the image (collinearity), the visual system assumes that the edge producing the line in the three-dimensional world is also straight.
  – The visual system ignores the possibility that the property in the image could have arisen from the highly unlikely “accidental” alignment of the eye with a smoothly curved edge (accidental viewpoints).

• As such, geon identity is viewpoint invariant (constant across viewpoints).
• Smoothly curved elements in the image (curvilinearity) are implied to arise from smoothly curved features in the three-dimensional world.
• If an image is symmetrical (symmetry), we also assume that symmetry exists in the object.
• When edges in the image are parallel or coterminate, we assume that the edges in the real-world object also are parallel or coterminate, respectively.

CONTRASTS IN NONACCIDENTAL PROPERTIES

1. Smooth Continuation
   - Straight
   - Curved
   - Yes (vertex type)
   - No
   - "L" - Fork Arrow

2. Cotermination
   - Curved end in 3-space
   - Curved end in 3-space
   - Yes (with bias in depth)
   - No

3. Parallelism

Biederman and Gerhardstein (1993)
• In this experiment, there were 48 line drawings of 24 objects (2 exemplars each).
• Each was created at 3 views differing in depth rotation by 67.5°.
• In the “priming block,” each object (1 exemplar) was presented either at 0 or 135°.
• In the second block, the degree of rotation in depth was varied, as was whether or not the participant received the same exemplar.
• The participants were to name (aloud) the object.
They find no effect of rotation in depth on object recognition.

- In a subsequent experiment, Biederman and Gerhardstein wanted to examine whether or not novel shapes would show the same viewpoint invariance.
- The stimulus set consisted of 10 5-component figures, with each drawn at 3 views separated by 45° in depth.
- In one of the views, parts that were not visible in the central view become exposed, while the other rotation presents the same 5 geons (the geon structural descriptions stays constant).

A sequential matching task was used in which the first presentation was randomly chosen to be from any of the 3 views, and the second (750 ms later) was always of the view shown in B. As such, presentation of A-B would be presentations in which no geons changed, while C-B would be trials upon which the geon structural description changed. Participants only responded when the object depicted was the same in the two intervals (different views are considered as same trials; there were also trials in which different objects were presented in the two intervals).

The findings show that RT and error rates only increased when the geon structural description changed across the two intervals. Only when the parts changed (for the same object) were the error rates and reaction times higher as function of angular disparity!

Tarr, Williams, Haywood and Gauthier (1998): Are geons really view-invariant?

- This study sought to determine whether or not recognition of geons themselves was truly viewpoint invariant.
- In particular, they were concerned that Biederman and Gerhardtstein, in their first experiment, had used highly familiar objects that were probably learned from several different viewpoints.

Experiment 1a-e utilize a sequential matching task in which two images are presented sequentially and the observer must decide whether or not they represent the same geon where “same” responses apply to images that are rotated 0, 45, and 90° (different trials were not used in the analysis).

Experiments 2a-c utilized a match-to-sample task in which the observer ran blocks with one geon shown at 0° followed by 12 trials consisting of three different orientations of the same geon (0°, 45°, and 90°) along with 9 other geons interspersed. Participants pressed a key when the same geon was shown.

Experiment 3 was a simple naming experiment in which verbal labels were learned to the 0° representations but then tested with all 30 of the above figures.
For all nine experiments, there are profound effects of viewpoint—reaction times are longer as the viewpoint difference increases. It is generally the case that viewpoint effects are largest early during the experiment. As participants run more blocks, viewpoint effects diminish.

- Tarr and his students generally believe that our visual system represents multiple, orientation-specific views of the same object.
- Without specific training, the most likely representation to be used is the “standard” or “canonical” view.
- With additional experience, other views are learned, and novel views are compared to the one they match most closely.
- One then observes view-point dependences that are relative to a learned view instead of the canonical view (Tarr, 1995).

View-based explanation: Image must be mentally rotated to match stored, canonical view.

Structural-description explanation: Positional relations change with rotation (above>beside), causing an imperfect match with stored representation.

Task: Name the object as quickly as possible