Cognitive psychologists who study perception try to infer (“figure out”) the automatic, unconscious mental processes that let people perceive objects and events so accurately. They use behavioral observations and ideas from many disciplines (linguistics [study of languages], computer science, mathematics, etc.) to try to figure out how people perceive form, motion, depth, etc., how they recognize familiar objects, faces and facial expressions, etc.

Researchers have developed several different kinds of models for form perception. Each has certain positive features, and each has weaknesses, indicating that a broadly useful model has yet to be developed. This exercise will describe briefly a few of these models.

The distinctive feature model of form perception states that perceptual systems respond to a specific feature (or combination of features) in the stimulus array. The higher the level of the perceptual system, the more selective perceptual systems become, requiring more and more complex combinations of features to activate them. This idea is closely related to the physiological feature detector model of form perception described in exercise asgn2s.

Recall from asgn2s that neurons in the visual system respond this way. Neurons in the primary visual cortex respond to an edge or bar of light at a particular angle in the visual world (and therefore on the retina). At each higher level of the visual system, neurons require progressively more complex stimulus patterns: moving bars, moving bars of specific length, and finally complex hand- or face-like patterns. Such neurons appear to code or represent the stimulus pattern that activates them. This means that when such neurons are active we perceive the stimuli that normally activate them.

Q1 Christian is developing a distinctive feature model for recognizing everyday objects, like toasters and chairs.
To do this he tries to describe objects by
A. features, like edges, angles, surfaces, etc., that objects have
B. the neurons in the brain their images activate
C. the features that make up the letters in their names
D. the features that organize objects against the background

The distinctive feature model makes a simple prediction: the more features two letters have in common, the more easily they are confused. This prediction is clearly supported. For example, observers are much more likely to confuse E with B, P, or F than with X, S or Q. The first three letters share three or four features with E; the last three share none.

A related prediction is illustrated in Figure 2-2t. Find the Z in the two columns of six letters, first in the left-hand column, then in the right.
You probably found the Z faster in the left-hand column than in the right. The Z is quite different from the other letters in the left column, but it shares many features with the other letters in the right column. It is the only letter in the left column that has no curved parts; it is like all the other letters in the right column because it has only straight lines. Therefore, the Z stands out more in the left column.
Perceptual models of complex distal objects can be perceptually constructed from a few basic forms.

Letters are a very limited and special group of forms. The world is made up of many complex three-dimensional objects, like automobiles, toasters, coffee cups, suitcases, tables, trees, etc. How do we perceive them? Biederman (1987) has proposed a model which states that all objects can be described as some unique combination of 36 different basic solid geometric forms. Biederman named the mental or perceptual models of these basic forms geons.

Figure 3-2t illustrates a few geons. Each geon can be constructed out of simpler distinctive features. For example, the pyramid can be made out of four 2-dimensional triangles. Geons represent perceptual, not physiological, entities (~things); they may or may not be based on specific neurons in the visual areas. Each geon in an object serves as a high level distinctive feature. Figure 4-2t shows how a table lamp, a suitcase, and a trumpet can be constructed from a few geons.

Q2. If Christian (see preceding question) built his distinctive feature model using geons, he would

A. use the one of 36 different basic geometric solids that it matches best
B. use a set of 36 basic 3-D shapes as the features that describe objects
C. represent an object by the geons the object activates in the brain
D. use features, like edges, angles, surfaces, etc., to represent objects

Object perception does not require that individual features be present in the visual field. Julesz (1965) showed that observers can see a shape created from random dots entirely by stereoscopic depth cues (see asgn2u) in the dots. A somewhat different way to create depth from only binocular cues can be seen HERE.

First the brain/mind has constructed a perceptual model of a distal object. Next, the model must be recognized, which means it must be connected to the perceptual models of other, related objects. Models of perception must do more than describe objects. They must explain how the description can be used to recognize the coded objects. So -- how does the brain/mind recognize something as fitting a particular category of objects, like tables, dogs, cups, etc.? Somehow it matches the perceptual representation of the object with the category to which the stimulus object that produces it belongs. The rest of the exercise describes some models of how people accomplish this task. They are: template, prototype, exemplar, and neural nets or neural networks.

The simplest model, the template model, states that an object is recognized when its perceptual model matches a perceptual model that already exists in the mind. The pre-existing perceptual model in the visual system serves as a template for the mental category it belongs to. In ordinary English, a template is an exact model or pattern of something. A lettering stencil is an example of a template. A stencil contains the outline of each letter in the alphabet, and you trace or paint through the stencil to make accurate duplicates of letters.

Similarly, the template model of form perception states that you have a template for each example of an object category that you have seen: For example, your category "dog" is defined by templates for every mutt, Collie, Labrador, Terrier, German Shepherd, Spaniel, etc., you have ever seen. Furthermore, you have a separate template for each view of each dog. Each of these templates is stored in your memory. Only if something matches one of these templates is it perceived as a member of the category "dog."

The template model has an obvious problem. It would require an enormous amount of memory for all the templates required for the thousands of perceptual categories people have. Furthermore, it would require an extremely fast and efficient search of that enormous memory, because you recognize objects in a split second.

The prototype model avoids this problem. Instead of storing separately each image of each object in a category, this model proposes that the mind computes...
**a single average of all these images.** This average is the "prototype" of that category. To the extent that a new image comes "close to" this prototype, you recognize that image as a member of that category. For example, the prototype model states that every image of every dog you have seen is combined into a single average dog, which is the prototype for "dog." If a new image comes close to that average, you recognize it quickly and accurately as a dog. A Beagle is a "typical" dog, so you recognize one quickly as a dog. A Chihuahua is a tiny dog, usually with little hair. It is very different from the prototype of dog, so you take longer to say "yes, it's a dog"; you might even mistake it for something else. A Siberian huskie may be closer to the prototype of "wolf," so it may be miscategorized (misidentified) as a wolf.

Figure 5-2t. Distribution of faces in a psychological space defined by smile - no smile and normal weight - overweight

![Distribution of faces in a psychological space](image)

The exemplar model of object perception, like the prototype model, combines all the images of a particular category of objects. But it does not compute an average. Instead the exemplar model uses the perceptual model of every example you have ever seen of a category. These perceptual models define the "psychological space," for that category. The "psychological space" has the advantage of not requiring a single average to represent a concept.

For example, we might take a series of photos of women who vary on two dimensions: a) body weight from normal weight to overweight, and b) facial expression from smiling to straight faced. These two dimensions can be used to define a very simplified form of the psychological space for "attractiveness." Most people categorize the normal weight, smiling women as the most attractive and the overweight, straight-faced women as the least attractive.

The overweight smiling and normal weight straight-faced women form a group of intermediately attractive women. An average prototype of the intermediate group is between the two areas of psychological space near the middle of the graph where the dark green dot is; it does not represent either subgroup accurately, whereas the psychological space that these two areas define does describe this intermediate group. Women without bulimic tendencies put more weight on smiling than on body weight in judging attractiveness. Women who do have some bulimic tendencies tend to ignore facial affect in their judgments. [Bulimia is an eating disorder in which people show exaggerated concern about their body shape and eat in binges often followed by purging].

Match the descriptions of models of object perception with the name of the model they fit best.

1. template  
2. prototype  
3. exemplar

Q4A. Objects must be similar to the average of all previous examples of that object that an observer has seen

Q4B. Objects must fit into the "psychological space" defined by all previous examples of that object that an observer has seen

Q4C. Objects must match very closely an object that the observer has seen previously
Computer Models of Perception

The digital computer is often described as an artificial brain. Many people think of it as the model for real brains, both in the way computers operate and in what they can be programmed to do. The idea that computers can model human mental processes is part of artificial intelligence.

Computers have several kinds of "hardware" (the monitor, keyboard, etc.)

- **Input devices**, like keyboards, video cameras, and microphones, which are like human senses.
- **Output devices**, the video screen, printers, and speakers, which are like human response systems (speech, hands).
- **Memory**, in which information can be stored indefinitely
- **A central processing unit** (often abbreviated CPU), which is the active part that carries out basic arithmetic (+, -) and logical (>, <, =, & , or) operations.

Most importantly they have **programs** (or software), which tells the hardware what to do. The software is what makes these machines useful.

The computer's memory, like human memory, stores information for later use. The central processing unit, like human mental processes, takes information from inputs and memory, processes it according to some program (set of instructions). The central processing unit either stores the result of its processing in memory or uses it to produce some sort of output or response.

When computers first became available, some researchers thought that their speed and logical abilities would make it easy to simulate (mimic) human mental processes. For example, they expected to get computers to do what the human visual system can do with a few years of work. The problem of modeling human mental processes turned out to be much more difficult.

Researchers did develop computer programs that recognize objects were developed fairly early. They work something like this: They take the signal from a TV camera and looks for edges or contours. They then classify the contours as lines with different slopes, curves, etc. The program then combines the lines and curves into the outline of an object against a background. It then compares the features of that object with the library of objects in the computer's memory. If a match was made, the program reports the outcome on the video screen. Such a program acts sort of like the template model of form perception, described above.

These early simulations of vision were very limited. They could recognize objects only from a limited set, they took minutes to do what people can do in a split second, and they were easily fooled by changes in background, position, or orientation of the object. Modern machines are much faster, have much more memory, and have much more powerful Central processing units. Nevertheless programs only began to approach what the visual system does so fast and with such apparent ease. In fact, computer engineers now look to the brain for ideas on how to make computers more efficient.

**Real brains work quite differently than do computers.** Computers do billions or trillions operations every second and have a virtually perfect memory. The brain is a million times slower and has a rather imperfect memory. Therefore, perceptual recognition programs failed to simulate visual perception and other human mental functions very well because they are not organized the way the brain is organized. Recent progress in this field is the result of using information about the way visual systems work to the design of the programs.

Most computers are **serial processors**. A serial processor has a single central processing unit which carries out a program one step at a time and does only one thing at a time. The brain is different from a serial processing computer in many ways. Two very important differences are these:

1. The brain is a **parallel processor**; different parts work on different parts of a problem at the same time. It is as if the brain had many simultaneously active central processing units.
2. Different parts of the brain affect each other. It is as if each central processing unit's actions affected the actions of many other central processing units.
Q5. A parallel processing computer is probably a better model for the brain that is the serial processor, because both the brain and parallel processors

<table>
<thead>
<tr>
<th>A. work on several parts of a mental task at the same time</th>
<th>B. have interacting parts</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. have single central processing units</td>
<td>D. work very fast, millions of operations per second</td>
</tr>
<tr>
<td>E. A and B are both correct</td>
<td>F. A, B, C, and D are all correct</td>
</tr>
</tbody>
</table>

In the 1980s some researchers began to develop programs that made ordinary computers act like parallel processors. These programs go by various names: parallel distributed processors, connectionist programs, adaptive neural nets, neural networks. I will use this last name.

*Neural networks are made of many "nodes," which are independent but interconnected processing units.* They are modeled after a few basic features of neurons. The interconnections allow each node to affect the activity of many other nodes in the network. The connections from each node increase or decrease the activity of the other nodes it connects to.

A particularly important feature of neural networks is that they *learn* the task they are supposed to do. They are *not programmed*. The typical computer must be told exactly what to do and must be given all its information. For neural nets, all that programmers do is to specify the inputs, the outputs, and the learning algorithm. The machine then learns by trial and error (better, trial and success) how to get from a set of inputs to the correct output.

Neural networks have been developed that learn to recognize objects, such as fruits. One group of researchers (Shallice et al., 1988) built and trained such a network. Then they "damaged" it by randomly disconnecting some percentage of its internal connections. The resulting "brain damage" produces behavior very similar to that observed in visual agnosia (inability to recognize) following damage to the temporal lobes. For example, it might "see" an apricot and call it a peach. Just as some people with visual agnosia respond with the correct category but the wrong word.

A neural network is formed by at least three "layers" of nodes:

1. the input layer, which receives information from the outside world
2. the output layer, which produces the response you measure
3. one or more "hidden" layers, which perform the information processing. The researcher specifies what each input node and each output node stands for, and that's all.

The network is then "trained" to generate the correct response to each set of inputs. Initially the strengths of the connections between nodes are set at random. The researcher gives the network an input (in the example, groups of features that can be put together to form a letter), and the network generates an output. It compares the output with the correct answer. If the network is correct, then it increases the strengths of the connections in the network. If the network is wrong, it decreases the strengths of the connections.

With repeated presentations, the network usually settles into a pattern of connections that (almost always) generates the correct answer. Furthermore, it
can respond correctly to new examples it has never had. The network programs itself by its successes and errors. In contrast, traditional computer models on serial processing computers require the researcher to write out the program in full detail.

Parallel processing systems like neural networks differ from serial processors in another very important way. Serial processors "crash" with almost any "damage." Parallel processors keep working, "sort of," much as the brain does when it is damaged. For example, a neural network was developed to recognize simple objects. It was then "damaged" by randomly disconnecting some percentage of its internal connections. The resulting "brain damage" produces behavior very similar to that observed in visual agnosia (inability to recognize) following damage to the temporal lobes.

Q6. [Mark EACH item True (T) or False (F)] Which of the following are properties of a neural network computer program?

A. Computer programmers write the program that does the computer’s intended tasks

B. It starts with no rules or instructions about how to do its task and learns to change its internal connections to do its task successfully

C. It is made up of interconnected "nodes"

D. It can learn to do recognition tasks for which rules cannot be written in advance

E. When part of it is damaged the rest keeps working, but not as well.

In summary, parallel processing models of the brain, like the neural network model, have several advantages over serial processing models.

1. The brain seems to work much more like a parallel processor than a serial one. The brain does many things at the same time. It continuously monitors the senses, controls responding and internal processing like thinking, etc.

2. Parallel processing greatly increases the overall speed of processing. This is especially important in the brain, where neurons operate one thousand to one million times slower than computer transistors.

3. Parallel processing models seem to behave more like the brain: For example, they have to "learn" (most of) their programs from the training procedure.

4. Parallel processing models do not fail completely ("crash") when partially damaged. Instead, they often show very human-like behavior to such damage, as mentioned above.

5. Like the human brain, neural networks do not require much more time to do complex tasks, like perceiving faces, than to do simple tasks, like perceiving simple geometric forms.

Match the examples with the preceding summary statement it goes with best.

1. 2. 3. 4. 5.

Q7A. A driver can carry on a conversation, accurately steer the car, watch out for other traffic, and chew gum all at the same time.

Q7B. A neuron can respond at most about 1,000/sec, yet the brain can make complex perceptual judgments very fast -- less than one second.

Q7C. Brain damage disturbs some mental functions but often leave most of the rest little affected.

Q7D. People recognize a face about as fast as they recognize a square.

For a more detailed description of neural nets, click HERE or HERE.
After you have finished the exercise, you should understand how the various monocular and binocular "cues" work to give three-dimensional information through the two-dimensional retina.
You can get this effect in Figure 1-2u. With a bit of practice you can learn to relax your eyes, the way they relax as you start to fall asleep, and fuse the two images into one. As your eyes drift apart, you will see two images of the two pencils, one from each eye. Try to move your eyes so that the inner images from each eye superimpose (~are on top of each other) to make a single image. The single image stands out in striking 3-D.

Chemists use binocular disparity to make the 3-D structure of complex molecules visible, like the one in Figure 2-2u. Journals print two images of the molecular structure viewed from slightly different angles. Use the technique described above to fuse the two images and see the middle one in 3-D. This one is much easier to fuse than are the pencils in Figure 1-2u. To go to a website about 3-D vision using binocular retinal disparity, click HERE.

Q2. Binocular disparity produces a sense of depth because the images on the retinas of the two eyes ____ as an object gets closer. Hint: How different are the images when the distal stimulus is 50 ft away?
A. become more different   B. become less different   C. blend together more
D. connect together more strongly   E. C and D are both correct

Monocular Cues for Depth

Among the many monocular cues, linear perspective is probably the most familiar. Modern painting began with the discovery of linear perspective by Italian artists in the 15th century. To create linear perspective, the artist chooses a vanishing point on the horizon on the painting, and objects are created so that they refer to this point. Many of you learned to do this in an art class.

Figure 3-2u shows linear perspective using a road and telephone poles converging toward the vanishing point. The building has a side with horizontal edges that would reach the vanishing point if they extended that far.

Linear perspective is easy to find in the real world as well. It is clear in a long hallway, as illustrated in Figure 4-2u. Notice how the lines that form the junction of the ceiling and floor to the two walls converge.

Texture gradient refers to the way the texture of a level surface gets finer and finer the farther away you look. Surfaces textures also provide information about the slope of a surface, because the rate at which the texture decreases depends on the slope of the surface. On a vertical surface the texture does not change. The rate at which the texture gets finer increases as a surface changes from vertical to horizontal.

A long hallway is a good example again. The floor and ceiling tiles and the cement blocks that form the walls appear to get smaller, the farther down the hall they are. This is illustrated in the Figure 6-2u. In contrast, the texture spacing of the concrete block on the vertical wall at the end of the hall appears about equal. Thus, a horizontal surface produces...
intervals in a texture that appear to get finer and more closely spaced as it recedes from you. A vertical surface produces intervals in a texture that appear about equal.

The cylinders in Figure 6-2u also illustrates these effects. They show that the cues for depth produce the expected effect on perception of size (see asgn2v). The cylinders look as if they are in a row down the hallway. The image of each cylinder makes a smaller image on your retina than does the one before, following the rules of linear perspective and texture gradient. Nevertheless, most people see them as the same size but farther and farther back.

Most surfaces have textures, which create perceptual depth gradients. A ploughed field is a good example. The photograph of a gravel roadway in Figure 7-2u illustrates another.

Other monocular depth cues include:

1. **Elevation.** The closer something is to the horizon, the farther away it looks.
2. **Superposition.** An object placed over another object appears nearer.
3. **Aerial perspective (haze).** Haze makes things look farther away. When you approach the Rocky Mountains in Colorado across the great plains, they look closer than they are, because the clear, dry air there has little haze.
4. **Shading.** A surface can appear curved inward or outward depending on the location of shading, as Figure 8-2u illustrates. Light is assumed to come from above, so shading at the bottom signals a bulge out of the surface, toward you, whereas shading at the top signals a dent into the surface, away from you (bottom panel of Figure 8-2u). Shading in the middle signals a depression; shading at the sides signals a bulge (top panel).

Match the following depth cues with their names below:

1. linear perspective  
2. superposition  
3. texture gradient  
4. haze (aerial perspective)  
5. elevation

- **Q3A.** parallel lines receding toward the horizon appear to meet
- **Q3B.** images closer to horizon line in the visual field look farther away
- **Q3C.** fuzzier images look farther away
- **Q3D.** an object partially covering another object looks closer
- **Q3E.** spacing between elements: farther elements on a horizontal surface appear closer together

Motion parallax provides strong and reliable information about depth. It is the only one that tells whether a scene is really three-dimensional or a picture, because it cannot be put into a picture. Most monocular cues are available to a stationary observer (one who does not move). An artist can reproduce them to produce a strikingly realistic impression of depth in a painting. A picture can be so realistic that you can have trouble telling whether it's a 2-D picture or the real thing in 3-D. But one of the most powerful cues for depth cannot be reproduced in a still picture. This cue is motion parallax, depends on the observer's motion relative to the visual world. It provides powerful, unambiguous information about distance, which cannot be reproduced in a still picture.

Imagine you are in a car driving past an open field. Imagine fixating (~look straight at) on a tree in the middle of that field. The weeds and fence close to you by the roadside appear to move fast in the direction opposite to
your motion. Things farther from you up to the tree appear to move slower, the farther they are from you. The tree itself appears motionless, because you keep looking straight at it. Things past the tree look as if they are moving with you. The farther they are beyond the tree, the faster they seem to move.

Another way to see motion parallax very clearly is to look at the trees as you walk through a wooded place. Trunks and branches that are close to you appear to move faster than trunks and branches that are farther away. As you move, things close to you appear to move rapidly in the opposite direction. The farther away an object is the slower it appears to move. The place in space you fixate on appears stationary; objects beyond it appear to move with you. As you walk, he trees, branches, leaves, etc. appear to flow toward you, spreading out as they come. TV shows like StarTrek show a spreading flow of stars or other objects to create the feeling of moving rapidly forward.

The series of photos in Figure 9-2u illustrates motion parallax as you walk down the aisle of a large class room. It shows the changes you see while fixating on the red plus in the middle of the far wall. The red dashed line marks one row of seats. Note that the near seats shift position a lot, whereas the seats close to the far wall don’t. Note also that the red dashed line swings from downward right diagonal to a downward left diagonal. This provides very clear information about depth.

To see a short video illustrating optic flow, click HERE. On the small screen that appears, click and hold the + button to move in. Click and hold the - button to move out. Notice how all parts of the image spread out as you move inward and how they shrink back toward each other as you move outward.

Q4. As you are riding on a country road, you fixate on a tree in the middle of a field. The barn at the far edge of the field looks as if it moves in the same direction as you are. The fence posts by the roadside look as if they move in the direction opposite to your motion. This describes __, which provides information about __.
A. phi phenomenon; motion in the absence of true motion
B. motion parallax; backwards masking of movement
C. motion parallax; depth (distance)
D. motion texture; surface orientation
E. motion adaptation; memory for direction of motion

Motion in Perception

Motion contributes importantly to many other aspects of perception, because the changing relations among the elements of a stimulus array uniquely define what the moving object is. Again, the relation among elements is crucial, as Gestalt psychology first clearly pointed out.

Motion helps pick out a target from among many "distractors." A camouflaged animal is difficult to pick out until it moves. The small movement against the background is quite effective in catching your attention, even out of the corner of your eye. The same effect appears in the laboratory. Finding a single letter T among a lot of F’s is greatly speeded if the T moves. Finally, motion of a few dots demonstrates the Gestalt law of common fate. A few dots moving in one direction stand out against a background of many other dots moving in a different direction. They are grouped together as a figure against a background of the many dots, as diagramed in Figure 10-2u.
Motion alone often provides enough information for recognizing individuals and objects. You can often recognize a friend just from the way his/her body moves while walking. You can recognize a person's gender from his/her walk. You can even perceive a great deal about what a person is doing just from his/her movement. Research has shown that movement is really the cue you use by putting tiny lights onto various parts of the body and videotaping those lights in complete darkness. All an observer sees is the movement of the lights. Observers accurately judge the gender of the person, what the person is doing, even how heavy a weight a person is lifting just from the motion of the lights relative to each other.

Motion can induce perception of 3-dimensional form. In one example, which you can see below, you see a pattern of dots that look random, except for a concentration inside two vertical lines. When the dot pattern is set in motion, a rotating vertical cylinder pops out. When the motion stops the cylinder collapses and disappears. The cylinder appears because the motion of the dots changes their position relative to each other. To see this demonstration, click HERE.

You can see another demonstration of effect of motion for producing 3D effects, by clicking HERE. When the new screen appears, click on B15H23 at the right of the screen. Then put the cursor on the pattern that appears in the box, depress and hold the left mouse button, and move the cursor within the box. The pattern snaps into 3D as long as it is moving. As soon as it stops, it snaps back into 2D.

Q5. Visual motion
A. attracts attention  B. can organize a group of elements that move together into a figure  
C. can create 3-dimensional effect which disappears when the motion stops 
D. alone can let observers tell what people are doing  E. A, B, C and D are all correct

The simplest theory of motion perception states that the adequate stimulus for motion is stimulation of neighboring points on the retina in sequence. The Phi phenomenon described in Exercise asgn2r on Gestalt laws shows that this theory is inadequate. The Phi phenomenon creates a strong sense of motion without any motion of the stimulus. It is induced by flashing two quite widely separated lights one after the other, as illustrated in Figure 11-2u. Observers perceive motion from one light to the other, not two separate, stationary light flashes.

You can see something like the Phi phenomenon with a digital watch that displays seconds. As the seconds change, the segments that turn off seem to move to the segments that turn on. The effect is quite clear when only one segment changes, as from 2 to 3. Movies and TV are based on the Phi phenomenon. They are produced by a sequence of still images presented at 24 frames a second.

Other perceptions of motion indicate that an object does not have to move to be perceived as moving. You may have seen an example of this when looking at the moon behind a thin layer of clouds on a windy night. Although the clouds are moving and the moon is stationary, the moon looks as if it were moving in a direction opposite to the cloud movement. The clouds may even appear stationary. This illusion of motion suggests an alternate model: motion is perceived relative to a background.

Two simple tests of this model have already been described in the exercise on Gestalt psychology: the relative motion effect, illustrated in Figure 12-2u, and the rim light-hub light illustrations in Exercise asgn2q. See (Ramachadran, 1986) for more information about the perception of relative motion.
In the relative motion effect, observers in a dark room see a luminous dot inside a luminous rectangle. If the rectangle moves in one direction, the dot appears to move in the other, but the rectangle appears not to move. This shows that the dot is perceived relative to the rectangle.

Q6. The phi phenomenon is an example of perception of motion in the absence of real motion. It
A. shows that we perceive motion when neighboring points on the retina are stimulated in sequence
B. shows that stimuli are perceived in relation to other stimuli and to the background, not in isolation
C. is the basis of motion pictures and television
D. A and C are both correct
E. B and C are both correct

One of the most remarkable achievements of perceptual systems is the stability of the perceptual world. This fact is called perceptual constancy. Under ordinary conditions, perception of distal objects (cars, tree, tables, books, etc.) remains quite accurate and stable, despite large variations in the proximal stimulus (the pattern of light rays reaching the eyes). Every time you move or every time what you are looking at moves, the image it forms on the retina changes. Yet the object remains the same object. In hearing different speakers saying the same word produce different sound patterns, yet listeners hear the same word.

Two kinds of processes operate to achieve perceptual stability in the face of stimulus variation. One is familiarity or expectancy, based on experience. The other is "higher-order" stability (Gibson, 1950, 1965) in the varying patterns of stimulation, which perceptual systems automatically extract.

"Higher-order" refers to relations between "elements" of a stimulus array, as opposed to the elements themselves. "Higher-order stability" refers to relations among elements of a stimulus array that remain unchanged or that change predictably as the stimulus array changes.

For example, as you walk, the objects you see (cars, sidewalk, trees, signs, etc.) appear to flow toward you, spreading out as they come. TV shows like StarTrek show reverse this process with a spreading flow of stars or other objects to create the feeling of moving rapidly forward. The things you see appear to get bigger and spread apart. Nevertheless, the distances between the corners of a box, the parts of a tree, etc. remain in a fixed relation to each other. The combination of the increasing size and the fixed relation among parts provides reliable information about the shape and motion of an object.

Higher-order stability plays an important role in perception. For example, recall the role of motion in making a pattern appear in two or three dimensions (asn2u; click HERE to go to the demonstration.) Motion can make a pattern look three-dimensional because the relationship between points in the pattern changes in a stable way if the form a three-dimensional object.

Lightness or Brightness Constancy

The simplest example to understand is brightness constancy (more accurately, lightness constancy). A white piece of paper in a room lit with a 60W lamp reflects considerably less light than does a black lump of coal outside on a bright, sunny day. Yet the paper looks white, and the coal black.

Brightness constancy depends on extracting the percent of light (albedo) that the coal and the paper reflect. This can be done by comparing the amount of light from the coal and from the paper to the average amount of light from their surroundings. The paper reflects more light than its surrounding; the coal reflects less. Thus, the total amount of light from white paper (or black coal) can vary greatly, but the relation of light from white paper or black coal to light from the surroundings remains unchanged. This is an example of a higher order stability. So the visual system compares light from each object to the light from its neighbors. This comparison gives an accurate estimate of the per cent of light that each object reflects.
**Q1.** Brightness (or lightness) constancy is based on a higher order stability because

A. it depends on the relation between the amount of light reflected from a surface to the amount of light reflected from surrounding areas.
B. a surface looks light if it reflects more light than do surfaces around it.
C. a surface looks dark if it reflects less light than do surfaces around it.
D. A, B, and C are all correct

**Size Constancy**

_The farther away an object is the smaller image it makes in your eye,_ as Figure 1-2v explains.

! The blue arrow and the red arrow are the same height, but the red arrow is twice as far away from the eye.
! So the image of the red arrow on the back of the eye (retina) is 1/2 as large.
! The green arrow is 1/2 the size of the blue and red arrows. It is located at the same distance from the eye as is the blue arrow.
! So it makes an image on the retina that is 1/2 the blue arrow's image, but the same size as the red arrow's image.

**Figure 1-2v.** The relation between retinal image size, object size, and distance of object.

If an observer cannot tell the distance of the arrows:
! the blue arrow appears perceptually twice the size of the red arrow and the green arrow.
! the green arrow and red arrow appear perceptually the same size.

If adequate depth information is available in the visual stimulus, then perception of size becomes accurate to about 100 ft. _The visual system combines information from image size and image distance to achieve perceptual size constancy._

_Figure 2-2v shows the effect of depth cues on the perception of size._ Most observers perceive the four blue cylinders as about the same size, though the size of the images gets progressively smaller from left to right. The linear perspective and texture gradient in the diagram provide strong depth cues. The visual perceptual system combines the distance and image size cues to produce the perception of (approximately) constant cylinder size.

**Figure 2-2v.** The depth cues from the texture gradient and linear perspective make the blue cylinders look equal in size but at different distances.

**Figure 3-2v.** The moon illusion is based on perception of the moon up in the sky as closer than when near the horizon.

_The moon illusion is an example of a failure of perceptual constancy._ Although its image size remains constant, the full moon looks much larger when it is just above the horizon when it rises or sets. The main reason for this illusion is the way people perceive the sky.

For most people, the sky is like the inside of a shallow bowl. The sky right above you appears the closest, and its perceived distance increases the farther to one side you look. The distance is greatest at and near the horizon. If you see the sky this way, the overhead moon should appear much smaller than the moon near the horizon, because it seems closer. The actual image size is the same, but the image on the horizon, which appears farther away, also looks bigger. Figure 3-2v summarizes this explanation of the moon illusion.
Q2. James is 6 ft tall. When you see him 9 ft away, his image on your retina is about 2/3 inch. As he walks away from you, his image shrinks. Your perceptual size constancy fails if
A. if the image on your retina gets smaller
B. if you cannot tell how far away he is
C. if his distance from you changes more than his actual height changes
D. if his distance from you changes less than his actual height changes
E. if his distance from you changes much as his actual height changes

Brightness constancy and size constancy depend on higher order stability: a constant ratio between the object and its surroundings. For brightness constancy, the ratio of light from the object and its surroundings remains constant. For size constancy, the ratio of image size and perceived distance remains constant. Other constancies depend on predictable changes in the relation between parts of the visual world.

For example, shape constancy depends on the predictable changes in the relations among an object’s features as it moves in space. Consider a cube that rotates from its left to its right. An edge of the cube appears on the left side, gets larger, and moves faster and faster toward the middle. When it reaches the middle, the changes go in the exact reverse, until the edge disappears at the right. This transformation can be derived from a sine wave (basic trigonometric function generated by the rotation of a circle). The perceptual system appears to extract this stable pattern of change and uses it to represent the stability or constancy of the cube.

The same kind of predictable transformation underlies other kinds of constancies. As you walk or drive, the texture of the visual world expands predictably toward you from the point toward which you are moving. For example, as you walk down a hall as illustrated in Figure 4-2v, the end of the hall casts a bigger and bigger image on the retina of your eye.

But the ratio of width to height (width/height) for the hall remains unchanged (blue lines and numbers in Figure 4-2v). In addition, the change in the hall’s width and height is completely determined. The surface texture of the walls, ceiling, and floor expands around you, as the red arrows in Figure 4-2v illustrate. The visual perceptual system uses these stable or predictable “higher-order” features to adjust for the changes in the stimulus array on the retina and maintain perceptual constancy.

This predictable transformation is the basis of motion parallax, the very strong and reliable monocular depth cue, described in the exercise on depth perception. As a sound source moves from your left to your right, the time and intensity difference of the sound reaching your ear changes predictably. The visual and auditory perceptual systems appear to extract these predictable changes to create perceptual constancies.

Q3. The car in front of you makes a right turn. Its image on your retina changes a lot, but you still perceive the car as a solid object that is turning, not changing its shape. You can do this because
A. you know cars are solid
B. your mind keeps the original image and ignores the changing image
C. the edges, corners, surfaces of the car change predictably in relation to each other
D. you rotate the image of the car in your head
Motion by itself is a very powerful cue for perception of objects and events. You can get motion independent of form simply by videotaping something in complete darkness, except for tiny lights on edges, bends, corners, etc. All you see in the resulting video is a pattern of moving lights, yet you can recognize a lot from these limited stimuli.

Such lights fastened to people’s joints provide a remarkable amount of information. Just from the moving lights you can recognize that it is a person, the person’s gender, even what a person is doing. You can recognize that a person is lifting something. You can even estimate with considerable accuracy how heavy the object is, because people lift differently depending on how heavy something is.

These results are important because they show that the proximal stimulus reaching the eye contains lots of information about the distal objects. This information is available in the predictable higher-order relations among parts of the stimulus array. Body motion contains such higher-order information because each part of the body changes position predictably in relation to other parts. These changes are predictable because the body parts can make only a specific set of position changes. The perceptual ability to detect these higher-order features presumably depends on the motion-sensitive system in the visual association area of the parietal cortex. This system must be able to extract the relation among the different moving parts of the body.

Q4. The data about information available from motion alone shows
A. the mind/brain knows how people move
B. the visual system extracts information from predictable changes in relations among parts of a stimulus array
C. guessing is important in perception
D. the mind/brain contains templates of movement as well as form

The motions described above result from stimulus patterns moving across a stationary retina. If the eyes move and the visual world remains stationary, you get an equivalent movement of the retinal stimulation. But the perception is quite different. Instead of perceiving a moving visual world, the world appears stationary as the eyes scan different parts of it.

The brain corrects for the motion of the image on the retina produced when the eye moves. The correction appears to be based on the motor output to the muscles that move the eyes. If these muscles are paralyzed, and the observer tries to move them, the eyes don’t move, but the visual world appears to move. The signals that correct for eye movements are still applied, though the eye muscles cannot move the eyes, so the visual world appears to move instead.

You can also move the eyes passively. Put your finger on your eyelid just below the end of your eyebrow and push gently in and out on your eyeball. The visual world appears to wobble, because you are moving the eye passively. No corrective signal is applied to the information in the brain, so the visual world appears to move as the eyeball moves. A similar effect occurs when you watch a video made with the camera moving quickly, especially in unusual or unexpected directions.

Q5. When you move your eyes to look at different parts of the visual field, the visual world does not appear to move. Yet when your eyes are moved artificially, the world appears to move. The fact indicates that
A. the movement on the retina under the two conditions cannot be the same
B. the world is not perceptually stable
C. the image on the retina is not the source of perceiving motion of things in the visual field
D. the brain must correct for image motion when it actively moves the eyes
E. B and C are both correct

Eye movements are required for accurate perception. One kind of eye movement is called a saccade. In a saccade, the eye makes a rapid jump from looking at one place to another. For example, when you read, your eyes make a saccade, usually to the second or third word to the right. After a saccade, they fixate...
Impossible 3-dimensional figures depend on having to combine information from several fixations. Look at the larger drawing in Figure 5-2v. At first it looks like a 3-dimensional object. Further inspection shows that is an “impossible” figure. It cannot represent a real object. If the same figure is small, you can take it in with one fixation, and you see that it is a rather odd flat drawing.

Q6. If the image of an object takes up more than a small area on the retina, then the eyes must make ____ to perceive it.
A. several templates B. several figures against the ground
C. several fixations D. binocular images disparate
E. unconscious inferences

Bottom Up and Top Down Processing

The preceding description of perceptual constancy is based on **bottom up** processing. **Bottom up refers to the idea that the senses extract the information needed for perception from the stimulus array that a sense organ like the eye gets.** The name comes from the view that the sense organs (eyes, ears, etc.) are at the **bottom** of their perceptual systems. So bottom up processing operates on information as it comes into the senses -- at the "bottom."

**The alternate idea, called **top down processing**, refers to the idea that experience and expectations are essential factors in perception.** Memories and expectations operate at the top of the perceptual system, perhaps at the level of recognition, for which association cortex is essential. Therefore, it acts at the top of the perceptual system, starting with the receptors at the bottom and ending in the sensory association areas at the top.

Everyday experience tells you that expectancies strongly affect perception. I often miss seeing something -- say, the sugar bowl -- that is moved even a foot or less from its usual place. Expectations influence even people who are supposed to be good, objective observers, like scientists. I read somewhere that geologists have a saying: "I wouldn't have seen it, if I hadn't believed it." [Read it again to be sure you did not turn the meaning around to fit the more common saying.]

A very good example of top down processing is the effect of **context** (~background information already existing in the mind). **People use context a lot to figure out meaning.** You learned much of your vocabulary by figuring out what a new word means from the context in which it is used. The Cloze test uses the context effect to measure reading ability. In this test, every fifth word is replaced by a blank. Readers can fill in a large fraction of the blanks correctly from the context of the words that come before. Good readers do this much better than poor readers. To see a copy of this test, click HERE.

Context effects appear in many places. For example, when people read, they often read right over typographical errors. The text that precedes the error puts much context information in the mind, so the reader knows what to expect. So s/he automatically and unconsciously corrects the incorrect letter(s) in the print.

Figure 6-2v illustrates two examples of context. In the left drawing, you interpret the pattern in the middle as a B when you look top to bottom, because it forms part of the familiar sequence A . . B . . C. When you look from left to right, you interpret it as part of the sequence 12 . . 13 . . 14. Find the typographical error in the triangle on the right side of Figure 6-2v. These examples show that people have expectancies about what words will appear. **These expectancies often override the actual stimulus to the eyes.** So readers often miss a typographical error, or they can guide
Q7. We often perceive objects and events based on what we expect to see rather than on the stimulus information our senses receive. This process
A. depends on perception during a single fixation
B. is based on using context and top down processing
C. uses prototypes rather than templates
D. uses serial rather than parallel processing

In 1949, Ames described several illusions that he believed reflected top down processing. His particular idea was that perceptions depend on the perceiver's expectations about the world. The distorted room, illustrated in Figure 7-2v, is the best known of these. The observer looks through a small hole (or a camera lens) to eliminate binocular and motion cues for depth. S/he sees what looks like an ordinary room with two people, one in each far corner. The person in the far left corner looks about the size of the person in the far right corner.

As the two people change places, the one going from the left to the right corner appears to get bigger, and the one going from right to left appears to get smaller. Of course, the room is actually distorted. As the top part of Figure 8-2v shows, the left corner is actually twice as far from the observer as is the right corner.

But the room is so painted with perspective and texture cues that the left and right corners look equally far away. The lower part of the drawing shows the real shape of one side wall. Notice that the far (upper) end is larger than the near end.

Ames believed that this illusion was based on our expectancies about rooms. People assume rooms are rectangular based on our extensive experience. This is a top down explanation, because it is based on what people know (have in their heads) about rooms.

This explanation has two main problems.
1. For some unexplained reason, our extensive experience with people remaining the same size doesn't seem to have any effect on this perception.
2. The room was built to give false information about the distance to the left corner. Accurate information about distance is essential for size constancy. Therefore, observers should make errors based on false distance information.

The person in the far left corner makes a smaller image on the observer's retina. Nevertheless, she or he appears as close as is the person in the right corner.

Because the condition for size constancy (accurate information about distance) is violated, the person in the left corner should appear smaller.

The bottom up alternative states that the room was designed to provide false stimulus information about...
the shape of the room. The stimulus information tells us that the far corners of the room were equally far away, when in fact the left corner was twice as far as the right corner. Therefore, a person in the far corner makes an image in the viewer's eye that is about 1/2 as big as the image of the person in the near corner.

Because we perceive size by combining image size with distance, we perceive the person walking from the left corner to the right corner to grow. Her image grows as she comes closer to us, but the false depth cues tell us falsely that she remains the same distance away. Because the image size grows, but the perceived distance from us remains constant, we perceive the woman to grow in size (when actually she is coming closer).

Q8. A top down explanation of the distorted room illusion states that it depends on ____. A bottom up explanation of the distorted room illusion states that it depends on ______.
A. distinctive features of the room; prototypes of the room
B. expectancies; false information about distance
C. prototypes of the room; distinctive features of the room
D. false information about distance; expectancies
E. abnormal activity in temporal cortex; abnormal activity in the retina

To go to a web site showing some visual illusions, click HERE.