What determines visual cue reliability?

Robert A. Jacobs

Visual environments contain many cues to properties of an observed scene. To integrate information provided by multiple cues in an efficient manner, observers must assess the degree to which each cue provides reliable versus unreliable information. Two hypotheses are reviewed regarding how observers estimate cue reliabilities, namely that the estimated reliability of a cue is related to the ambiguity of the cue, and that people use correlations among cues to estimate cue reliabilities. Cue reliabilities are shown to be important both for cue combination and for aspects of visual learning.

Visual environments are often rich in information sources. For example, we now know of nearly a dozen different cues to visual depth [1]. This wealth of information poses an interesting problem for human observers: how do observers know whether each visual cue is providing reliable or unreliable information in the current visual context? Determining the relative reliabilities of available visual cues might be difficult for a variety of reasons. Importantly, all visual cues are ambiguous. A visual depth cue, for instance, might be consistent with a range of depth values, sometimes a small range and sometimes a large range (see Box 1). In addition, cue reliabilities can be context sensitive. Stereo cues to depth tend to be reliable when viewing nearby objects but unreliable when viewing distant objects [1,2]. Furthermore, cues in an environment can provide conflicting information. When watching a movie projected on a flat screen, stereo and motion parallax cues indicate that all objects depicted in the movie are at the same depth from the viewer whereas perspective, texture, shading, and other cues indicate that some objects are nearer than others.

In the past, vision scientists often used prisms to alter the characteristics of sensory cues to study the principles underlying observers' assessments of cue reliabilities [3]. However, prisms do not give investigators good control of the visual environment. Recently, advances in computer graphics and virtual reality technologies have provided researchers with new tools for controlling the statistical properties of visual cues in a precise manner. These tools have allowed scientists to address a wider range of questions in a more detailed fashion than was previously possible.

Robert A. Jacobs
Dept of Brain and Cognitive Sciences and the Center for Visual Science, University of Rochester, Rochester, NY 14627, USA.
e-mail: robbie@bcs.rochester.edu

Several recent studies have examined strategies human observers use to assess the relative reliabilities of available cues in a visual environment [2,4–28]. Nearly all of these studies address one or more of the following important questions:

- What criteria do human observers use to distinguish reliable visual cues from unreliable cues?
- How do observers evaluate these criteria so that they can estimate cue reliabilities?
- What do observers do with their estimated cue reliabilities after these estimates have been computed?

Issues regarding visual cue reliabilities have frequently been studied in the context of visual depth perception, and many of the studies reviewed here have examined the perception of the depth extension of objects. It is hoped that conclusions drawn about this case are applicable to other instances of depth perception, such as perception of distance from a viewer to an object or perception of the depth separation between two objects, and to the perception of other properties of visual scenes.

Cue reliabilities and cue ambiguities

A commonly assumed framework for how an observer might go about judging the depth of a visual object defined by multiple cues is the following two-stage process. First, depth estimates based on individual cues are derived. Second, a weighted combination of these estimates is calculated and used as the observer's composite depth percept; the cue weights are based on the relative reliabilities of the cues in the current visual context [7,12]. For example, consider an observer judging the depth of an object defined by motion and texture cues. During stage one, the observer calculates depth estimates based on each individual cue. Let $d_M(m)$ denote the observer's depth-from-motion estimate, and let $d_T(t)$ denote the observer's depth-from-texture estimate. During stage two, the observer combines these estimates into a unified depth percept, denoted $d(m,t)$, using, for instance, a linear cue combination rule:

$$d(m,t) = w_M d_M(m) + w_T d_T(t) \qquad \text{[Eqn 1]}$$

where the linear coefficients for the motion and texture cues, denoted w_M and w_T, are chosen based on the estimated reliabilities of these cues. This article discusses linear cue combination rules for depth because they are commonplace in the vision science literature, they have received a considerable degree of empirical support across a wide variety of experimental conditions [4,5,12,16], and they are both easy to understand and sufficient for illustrating many of our main points. (As an aside, the reader should understand

  PII: S1364-6613(02)01948-4

Box 1. Visual cues are ambiguous

The images that fall on our retinas are two-dimensional, yet we perceive the visual world in three dimensions. When thinking about this remarkable accomplishment, we should consider why it is that deriving three-dimensional perceptions from two-dimensional images might be a difficult computational problem for the brain to solve. This question is commonly answered with reference to the 'inverse optics problem': given a two-dimensional image, the observer needs to determine the three-dimensional scene from which the image is a projection. This is a difficult problem because every two-dimensional image is consistent with an infinite number of three-dimensional scenes.

Another way of approaching the inverse optics problem is to consider cue ambiguity. Every visual cue is ambiguous. There are many reasons underlying this ambiguity, including physical factors, such as atmospheric or optical blurring, and biological factors, such as noise inherent to human nervous systems. Therefore, there is no 'correct' interpretation of a cue. Consider an observer viewing a coffee-cup. The visual environment provides many cues to the shape and depth of this cup. Now consider one particular cue, such as a shading cue, to the depth of the coffee-cup (this depth might be defined, for instance, as the distance from the point on the cup closest to the observer to the point furthest away). The horizontal axis of the graph in Fig. I gives possible values of the depth, and the vertical axis gives the conditional probability of a depth conditioned on the value of the cue. This probability distribution is not a delta function; that is, the cue is not consistent with one, and only one, depth value. Rather, the cue is consistent, to a lesser or greater degree, with a range of depth values. If the variance of the probability distribution is relatively small, then the observer might believe that the cue is reliable because it specifies the cup's depth as lying within a narrow range. Consequently, the observer should make extensive use of the information provided by this cue. If, however, the variance is relatively large, then the observer might believe that the cue is unreliable because it is consistent with many possible depths. In this case, the observer might ignore the information provided by this cue, or at least discount the information provided by this cue relative to the information provided by other, more reliable, cues. An observer that follows the logic outlined here would be acting in accordance with a mathematical model known as a Kalman filter, as discussed in the text.

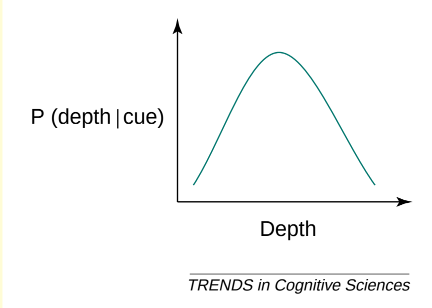

Fig. I. The visual environment typically provides many cues to the depth of an object. Consider one particular cue. The horizontal axis of this graph gives possible values of the object depth, and the vertical axis gives the conditional probability of each possible depth conditioned on the value of the cue. Note that the cue is ambiguous with respect to object depth because it is consistent, to a lesser or greater degree, with many possible depths.

that results inconsistent with a linear rule have also appeared in the literature [6,21,25], that information from multiple cues cannot always be directly averaged because these cues provide different types of information [see the discussion of cue promotion in Landy *et al.* [12]], and that alternative models of how observers combine information from multiple cues have been proposed [6,17,21,29].)

An important hypothesis concerning visual cue reliabilities is that the estimated reliability of a cue should be related to the ambiguity of the cue, such that highly ambiguous cues are regarded as unreliable and less ambiguous cues are regarded as reliable. This idea is nicely illustrated by a mathematical model known as a Kalman filter, which is an instance of a maximum likelihood estimator. According to one version of a Kalman filter, a cue is reliable if the distribution of inferences based on that cue has a relatively small variance; otherwise the cue is regarded as unreliable. In addition, more reliable cues are assigned a larger weight in a linear cue combination rule, and less reliable cues are assigned a smaller weight (see Box 1 and Fig. 1). Continuing with our example from above, let d denote a possible depth of a visual object, and let m and t denote the values of the motion and texture cues. In addition, let d^*_m denote the optimal estimate of visual depth based solely on the motion cue [this is the depth d that maximizes the probability of a depth value given the motion cue, $P(d|m)$], let d^*_t denote the optimal depth estimate based solely on the texture cue [the depth d that maximizes $P(d|t)$], and let d^* denote the optimal depth estimate based on both motion and texture cues [the depth d that maximizes $P(d|m,t)$]. Given certain mathematical assumptions, Yuille and Bülthoff [29] used Bayes' rule to show the following result:

$$d^* = w_m d^*_m + w_t d^*_t \qquad \text{[Eqn 2]}$$

where

$$w_m = \frac{\frac{1}{\sigma_m^2}}{\frac{1}{\sigma_m^2} + \frac{1}{\sigma_t^2}} \quad \text{and} \quad w_t = \frac{\frac{1}{\sigma_t^2}}{\frac{1}{\sigma_m^2} + \frac{1}{\sigma_t^2}} \qquad \text{[Eqn 3]}$$

and σ_m^2 and σ_t^2 are the variances of the distributions $P(d|m)$ and $P(d|t)$ respectively. This version of the Kalman filter (Eqns 2 and 3) has several appealing properties. First, the optimal estimate of depth based on both motion and texture cues is a linear combination of the optimal estimates based on the individual cues. Second, the linear coefficients, the weights w_m and w_t, are non-negative and sum to one. Finally, the weight on a cue, such as the motion weight w_m, is large when the cue is relatively reliable (the variance σ_m^2 is smaller than the variance σ_t^2), and small when the cue is relatively unreliable (σ_m^2 is larger than σ_t^2).

Equations 2 and 3 specify the statistically optimal cue combination rule given certain mathematical assumptions. They do not, however, tell us how human observers combine information from multiple visual cues. Do human observers use optimal cue combination strategies? Several researchers have addressed this question and found that the optimal cue combination rule predicts qualitatively [11,12,16,26] and quantitatively [19,27] observers' responses in a wide variety of circumstances.

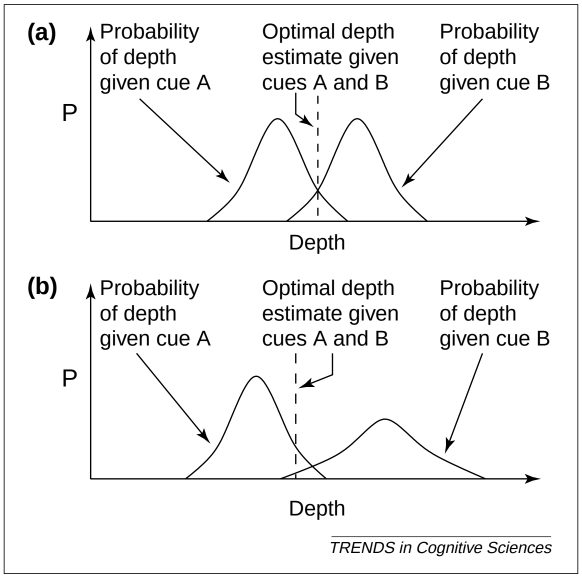

Fig. 1. Consider a Kalman filter estimating the depth of an object based on two cues, labeled A and B. The horizontal axis of each graph in this figure represents possible depth values, and the vertical axis represents the probability of a depth value. In each graph is shown the probability of depth given cue A, the probability of depth given cue B, and the filter's optimal depth estimate given both cues. The optimal depth estimate based on both cues is a weighted average of the means of the distributions based on single cues. If the distribution of depth given one cue is equal to the distribution given the other cue, then the weights used in the average are equal and the optimal estimate is halfway between the two means (top graph). If, however, the depth distribution given cue A has a smaller variance than the distribution given cue B, then the mean based on cue A is assigned a larger weight than the mean based on cue B and the optimal estimate is closer to the mean based on cue A (bottom graph).

Observers seem to judge the reliability of a cue as being inversely proportional to the variance of the distribution of inferences based on that cue. It is not known, however, how observers estimate this variance. A possibility that might be applicable to many visual environments is that observers evaluate a cue at different points in time. To investigate this possibility, Triesch *et al*. [28] conducted an experiment in which the consistency of cues was manipulated over time. For example, on a particular trial the color of an object might have changed rapidly, meaning that color is a poor predictor for an object's identity. By contrast, an object's shape either might not have changed or it changed infrequently, meaning that shape is a better predictor of identity in this situation. They found that observers tended to assign a large weight to information provided by a cue that did not change its value in the recent past, and assign small weights to cues that recently changed their values. Moreover, their judgments of cue variances seemed to require the temporal integration of information over a 1-second time window.

A second possibility, discussed by Ernst and Banks [27], is that a neural representation of a scene property might encode the ambiguity of a sensory cue. For example, consider an observer viewing an object defined by a stereo cue. Presumably, the activities of neurons in the observer's visual cortex form a neural population code that represents an estimate of the object's depth and also the uncertainty in this estimate. If the population code is shaped like a normal distribution, such that the mean and variance of this distribution represent the depth estimate and the uncertainty in this estimate, respectively, then the nervous system could implement a Kalman filter in a direct manner. The product of two normal-shaped population codes, based on two sensory cues such as stereo and texture cues to depth, is also a normal-shaped population code. The mean and variance of this new code represent the optimal depth estimate based on both cues and the uncertainty in this estimate.

Cue reliabilities and cue correlations

A second important hypothesis regarding observers' assessments of cue reliabilities is that these assessments are based on cue correlations. That is, a cue is regarded as reliable if the inferences based on that cue are consistent with the inferences based on other cues in the environment. Otherwise, the cue is regarded as unreliable. This hypothesis assumes that consistency among cues is unlikely to occur by accident. Instead, it is more probable that this consistency arises because the values of different cues are determined by the same underlying property of the environment. Using cue correlations, there are at least two ways that observers can adapt their visual perceptions so as to make them more veridical. Observers can adapt their cue combination strategies by increasing the cue weights associated with reliable cues and decreasing the weights associated with unreliable cues. In addition, observers can adapt their interpretations of individual unreliable cues so that these interpretations are more consistent with those based on reliable cues. This might occur, for example, if texture indicates a value of an object's depth and is judged to be an unreliable cue, whereas motion indicates a different depth value and is regarded as a reliable cue. An observer in this circumstance can adapt his or her depth-from-texture estimates so that they are more consistent with his or her depth-from-motion estimates.

This view is perhaps most closely associated with the theorizing of Hans Wallach [30]. He believed that in every perceptual domain, such as depth or shape perception, there is one primary source of information, usable innately and not modifiable by experience. Other cues are acquired later, through correlation with the innate process. For example, Wallach considered the phenomenon of induced motion, an illusion in which a small stationary object is perceived as moving when it is surrounded by a larger moving object (e.g. when a large cloud passes in front of the moon, it often seems as if the moon is moving through the cloud). He conjectured that this illusion can be accounted for as follows. Suppose that image displacement is the primary cue to visual motion. Observers have learned about other cues to visual motion because these cues are correlated with

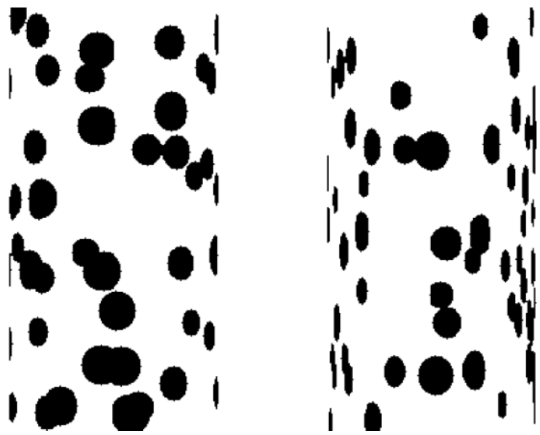

Fig. 2. Two examples of elliptical cylinders (i.e. cylinders whose horizontal cross-sections are ellipses) defined by a texture cue. The cylinder on the right should appear to extend in depth more than the cylinder on the left.

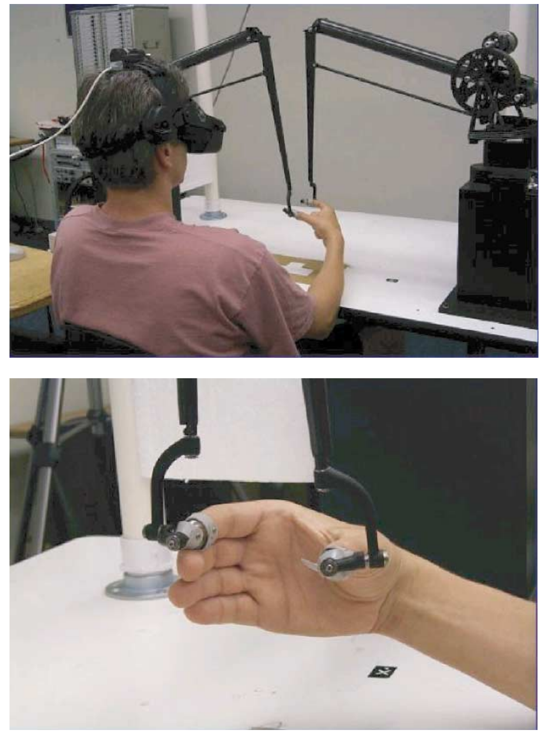

Fig. 3. Top panel: the experimental apparatus used by Atkins *et al.* [24] consisted of virtual reality goggles and two PHANToM™ 3D Touch interfaces that were attached by two fingerholders to an observer's thumb and index fingers. This apparatus allowed observers to interact physically with virtual objects viewed via the goggles in a natural way using a wide range of movements (e.g. grasping, moving, or throwing objects). The 3D Touch interfaces generated force fields that created haptic sensations (e.g. weight, hardness, friction) appropriate to the motor interactions with the object displayed in the goggles. Bottom panel: an observer is grasping a (virtual) elliptical cylinder. Because of the forces generated by the 3D Touch interfaces, he cannot close his fingers any further, and so he feels as if he is grasping the rigid surfaces of the cylinder.

image displacement. In particular, configurational change, the changing configuration in the vicinity of a moving object caused by the object's changing position relative to other image contents, tends to start and stop simultaneously with image displacement and perceived motion. Observers have learned, therefore, that it too can be regarded as a cue to visual motion. Consequently, the configurational change that occurs when a large cloud passes in front of the stationary moon leads observers to misperceive the moon as moving through the cloud.

It is worth noting that the use of cue correlations to estimate the reliabilities of visual cues is not limited to correlations among visual cues. Instead observers might correlate visual cues with those occurring in other sensory modalities. It has often been speculated that people learn to visually perceive the world by comparing their visual percepts with percepts obtained during motor interactions with the environment [for example, observers can correlate visual and haptic (touch) percepts of the depth of their coffee-cup when they view and grasp their coffee-cup]. Historically, this idea might have been proposed first by Bishop George Berkeley [31]. Berkeley speculated that visual perception of depth results from associations between visual cues and sensations of touch and motor movement. A frequently cited quote from his book is that 'touch educates vision.' Piaget [32] used similar ideas to explain how children learn to interpret and attach meaning to retinal images based on their motor interactions with physical objects.

The question of whether or not observers assess the reliabilities of visual cues based on consistencies between visual and haptic percepts has been addressed in a direct and detailed manner in recent studies [22,24]. For example, Atkins *et al*. [24] used a virtual reality environment that allowed observers to view and grasp elliptical cylinders. Visually, the cylinders were defined by texture and motion cues (the texture cue is illustrated in Fig. 2). A haptic percept of the depth of a cylinder was obtained when an observer grasped the cylinder along the depth axis with his or her thumb and index fingers (see Fig. 3). On each trial, observers viewed and grasped a cylinder, and judged whether their visual and haptic percepts of the cylinder's depth were the same or different (no feedback was given to observers regarding the correctness of their judgments). In the texture-relevant experimental condition, the haptic and visual texture cues indicated the same cylinder depth, whereas the visual motion cue indicated a cylinder depth that was uncorrelated with the depth indicated by the haptic cue. Observers tended to adapt their visual cue combination strategies by increasing the weight assigned to the visual texture cue and decreasing the weight assigned to the motion cue. This result suggests that observers (unconsciously) noticed that the haptic and texture cues were correlated whereas the haptic and motion

Questions for future research

- Cue variances and cue correlations are two statistical measures that a visual system can use to estimate how informative a cue is about the true state of an environment. What is the mechanism(s) for estimating cue reliabilities? A general purpose mechanism that computes statistical measures in a large number of domains and a domain-specific mechanism that is limited to computing properties of the visual world have been proposed [20]. What are the similarities and differences between the mechanisms estimating visual cue reliabilities and related statistical mechanisms operating in other domains?
- Estimating cue reliabilities is likely to be complex owing to context dependencies. For example, people rely on depth-from-stereo information more than depth-from-motion information when viewing nearby objects but not when viewing distant objects [1,2]. What are the exact conditions in which each visual cue is highly informative versus less informative about the true state of the environment?
- Wallach [30] speculated that there is one primary source of information in every perceptual domain that is usable innately and not modifiable by experience, and that other cues are acquired later through correlation with the innate process. If so, then how do we identify which perceptual cues are usable innately and which are learned on the basis of experience? Developmental studies might shed light on this issue.
- From a neuroscientific perspective, how do correlations among visual cues and cues from other sensory modalities help observers estimate the reliabilities of available visual cues? Research is finding evidence for brain regions that are not specialized for individual sensory domains but instead are multimodal [34]. Could these multimodal regions subserve the correlation of visual cues with cues from other senses for the purpose of estimating visual cue reliabilities? If so, how?

cues were uncorrelated and, thus, they regarded the texture cue as reliable and the motion cue as unreliable. In the motion-relevant condition, by contrast, the cylinder depths indicated by haptic and motion cues were identical, whereas the texture cue indicated a cylinder depth that was uncorrelated with the depth indicated by the haptic cue. In this condition, observers tended to adapt their visual cue combination strategies by increasing the weight assigned to motion and decreasing the weight assigned to texture, suggesting that they regarded motion as relatively more reliable than texture. Overall, these results suggest that observers can correlate visual and haptic cues to assess the relative reliabilities of the visual cues available in the environment.

This study, among others [20,22], suggests that observers use cue correlations to determine the reliabilities of available visual cues, and then adapt their visual cue combination strategies to emphasize information provided by reliable cues. An alternative way in which cue correlations can play a role in visual learning is in the adaptation of observers' interpretations of individual cues. This form of adaptation is known as cue recalibration; when it occurs on the basis of cue correlations it is also known as recalibration by pairing. The study of this form of visual learning was commonplace in the 1960s and 1970s [33]. Recently, however, cue recalibration has been examined again.

Adams *et al.* [23], for example, asked observers to wear a horizontal magnifier in front of one eye for several days. Initially observers perceived large three-dimensional distortions owing to the magnifier, although the magnitude of these perceived distortions diminished during the course of the experiment. Before, during, and after wearing the magnifier, observers were tested on a set of slant perception tasks in which they judged the slant of a planar surface defined by binocular disparity and/or texture cues. Their judgments with texture-defined surfaces did not change during the course of the experiment, indicating that they did not recalibrate their perceived slant from texture. In addition, when surfaces were defined by both texture and disparity cues, observers' cue weights in a linear cue combination rule did not change during the experiment, indicating that changes in perceived slant were not due to an adaptation of these weights. It was found, however, that observers' judgments with disparity-defined surfaces changed dramatically during the course of the experiment. In conjunction with the results from a control experiment, this finding suggests that observers recalibrated their interpretation of binocular disparities in response to the presence of a horizontal magnifier by adapting the mapping between disparity and perceived slant.

Concluding remarks

Visual environments often contain many cues to properties of an observed scene. To integrate information provided by multiple cues in an efficient manner, observers must assess the degree to which each cue provides reliable versus unreliable information. This article has reviewed two hypotheses regarding how observers might estimate cue reliabilities. One hypothesis is that the estimated reliability of a cue is related to the ambiguity of the cue, such that highly ambiguous cues are regarded as unreliable and less ambiguous cues are regarded as reliable. This idea underlies the operations of a mathematical model known as a Kalman filter. A second hypothesis is that people use cue correlations to estimate cue reliabilities. A cue that is correlated with other cues in the environment is regarded as reliable, whereas a cue that is uncorrelated with other cues is regarded as unreliable. Based on these estimates of cue reliabilities, people adapt their cue combination rules so as to place more weight on information derived from reliable cues. They also recalibrate their interpretations of unreliable cues so that these interpretations are more consistent with those based on reliable cues. Consequently, cue reliabilities are important both for cue combination and for aspects of visual learning.

Acknowledgements
I thank M. Hayhoe and D. Knill for commenting on an earlier version of this manuscript. This work was supported by NIH research grant R01-EY13149.

References

1 Cutting, J.E. and Vishton, P.M. (1995) Perceiving layout and knowing distances: the integration, relative potency, and contextual use of different information about depth. In *Perception of Space and Motion* (Epstein, W. and Rogers, S., eds), pp. 69–117, Academic Press

2 Johnston, E.B. *et al.* (1994) Integration of motion and stereopsis cues. *Vision Res.* 34, 2259–2275

3 Welch, R.B. (1978) *Perceptual Modification: Adapting to Altered Sensory Environments*, Academic Press

4 Dosher, B.A. *et al.* (1986) Tradeoffs between stereopsis and proximity luminance covariance as

determinants of perceived 3D structure. *Vision Res.* 26, 973–990
5 Bruno, N. and Cutting, J.E. (1988) Minimodularity and the perception of layout. *J. Exp. Psychol. Gen.* 117, 161–170
6 Bülthoff, H.H. and Mallot, H.A. (1988) Integration of depth modules: stereo and shading. *J. Opt. Soc. Am. A* 5, 1749–1758
7 Maloney, L.T. and Landy, M.S. (1989) A statistical framework for robust fusion of depth information. *Vis. Commun. Image Process. IV* (*Proc. SPIE*) 1199, 1154–1163
8 Rogers, B.J. and Collett, T.S. (1989) The appearance of surfaces specified by motion parallax and binocular disparity. *Q. J. Exp. Psychol. A* 41, 697–717
9 Johnston, E.B. *et al.* (1993) Integration of depth modules: stereopsis and texture. *Vision Res.* 33, 813–826
10 Nawrot, M. and Blake, R. (1993) On the perceptual identity of dynamic stereopsis and kinetic depth. *Vision Res.* 33, 1561–1571
11 Young, M.J. *et al.* (1993) A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Res.* 33, 2685–2696
12 Landy, M.S. *et al.* (1995) Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res.* 35, 389–412
13 Frisby, J.P. *et al.* (1996) Stereo and texture cue integration in the perception of planar and curved large real surfaces. In *Attention and Performance 16: Information Integration in Perception and Communication* (Inui, T. and McClelland, J.L., eds), pp. 77–92, MIT Press
14 Tittle, J.S. *et al.* (1997) The perception of scale-dependent and scale-independent surface structure from binocular disparity, texture, and shading. *Perception* 26, 147–166
15 Turner, J. *et al.* (1997) The relationship between binocular disparity and motion parallax in surface detection. *Percept. Psychophys.* 59, 370–380
16 Knill, D.C. (1998) Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Res.* 38, 2635–2656
17 Poom, L. and Boerjesson, E. (1999) Perceptual depth synthesis in the visual system as revealed by selective adaptation. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 504–517
18 Backus, B.T. and Banks, M.S. (1999) Estimator reliability and distance scaling in stereoscopic slant perception. *Perception* 28, 217–242
19 Jacobs, R.A. (1999) Optimal integration of texture and motion cues to depth. *Vision Res.* 39, 3621–3629
20 Jacobs, R.A. and Fine, I. (1999) Experience-dependent integration of texture and motion cues to depth. *Vision Res.* 39, 4062–4075
21 Porrill, J. *et al.* (1999) Robust and optimal use of information in stereo vision. *Nature* 397, 63–66
22 Ernst, M.O. *et al.* (2000) Touch can change visual slant perception. *Nat. Neurosci.* 3, 69–73
23 Adams, W.J. *et al.* (2001) Adaptation to three-dimensional distortions in human vision. *Nat. Neurosci.* 4, 1063–1064
24 Atkins, J.E. *et al.* (2001) Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Res.* 41, 449–461
25 Saunders, J.A. and Knill, D.C. (2001) Perception of 3D surface orientation from skew symmetry. *Vision Res.* 41, 3163–3183
26 Landy, M.S. and Kojima, H. (2001) Ideal cue combination for localizing texture-defined edges. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 18, 2307–2320
27 Ernst, M.O. and Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433
28 Triesch, J. *et al.* (2002) Fast temporal dynamics of visual cue integration. *Perception* 31, 421–434
29 Yuille, A.L. and Bülthoff, H.H. (1996) Bayesian decision theory and psychophysics. In *Perception as Bayesian Inference* (Knill, D.C. and Richards, W., eds), pp. 123–161, Cambridge University Press
30 Wallach, H. (1985) Learned stimulation in space and motion perception. *Am. Psychol.* 40, 399–404
31 Berkeley, G. (1963) An essay towards a new theory of vision. In *Works on Vision* (Turbayne, C.M., ed.), Bobbs-Merrill, Indianapolis
32 Piaget, J. (1952) *The Origins of Intelligence in Children*, International Universities Press
33 Epstein, W. (1975) Recalibration by pairing: a process of perceptual learning. *Perception* 4, 59–72
34 Amedi, A. *et al.* (2001) Visuo-haptic object-related activation in the ventral visual pathway. *Nat. Neurosci.* 4, 324–330

The brain circuitry of syntactic comprehension

Edith Kaan and Tamara Y. Swaab

Syntactic comprehension is a fundamental aspect of human language, and has distinct properties from other aspects of language (e.g. semantics). In this article, we aim to identify if there is a specific locus of syntax in the brain by reviewing imaging studies on syntactic processing. We conclude that results from neuroimaging support evidence from neuropsychology that syntactic processing does not recruit one specific area. Instead a network of areas including Broca's area and anterior, middle and superior areas of the temporal lobes is involved. However, none of these areas appears to be syntax specific.

Reading or hearing a sentence such as '*The little old man knocked out the giant wrestler*' demonstrates the crucial role of syntax in normal language understanding. Identifying who did what to whom enables humans to understand the unlikely scenario that is described here. Thus, syntactic information helps us combine the words we hear or read in a particular way such that we can extract the meaning of sentences (see Box 1). Many regard syntax as a cognitive module that is separable from other more general cognitive processes such as memory and attention [1] and whose properties can be distinguished from semantic-conceptual information ('meaning') [2]. In this tradition, some theories of sentence processing propose a separate syntactic processing mechanism that is insensitive to nonsyntactic information [3]. However, alternative views exist [4,5]. Given these competing views of syntax, one can ask whether there is neurological evidence in favor of a syntactic processing module [1]; that is, is there a specific area in the brain that is specialized for syntax alone?

Evidence from brain lesions

Research on the relationship between brain and language dates back to the mid- to late-1800s when Paul Broca and Karl Wernicke linked specific lesions in the brain to specific language deficits known as aphasia. Broca identified patients with problems in

  PII: S1364-6613(02)01947-2