11

Why We Don't Mind to Be Inconsistent

Jeroen B.J. Smeets and Eli Brenner

Research Institute MOVE, Faculty of Human Movement Sciences, VU University, Amsterdam, The Netherlands

INTRODUCTION

It is frequently assumed that perception involves the creation of a model of our environment. Our senses provide incomplete and noisy information about the objective world. One might think that what we perceive is the situation that best matches the incomplete and noisy information. However, this is not the case. Hermann von Helmholtz (1925) already noted that perception is unconscious inference about the situation that most likely caused the sensory state. This view has recently become very popular and has been formalized in terms of Bayesian inference (Kersten et al., 2004; Knill & Pouget, 2004; Körding & Wolpert, 2006). In this view, the likelihood that you perceive situation X depends on the likelihood that situation X is the cause of the present sensory state, multiplied by the a priori likelihood that situation X occurs. This description of perception as the formation of an internal model of the outside world, which represents the most likely cause of our sensory stimulation, is a very powerful approach to perception, that can explain various phenomena.

One clear prediction of the Bayesian/Helmholtzian approach is that you will never perceive a situation that is physically impossible. This seems a reasonable prediction, but M.C. Escher's drawing in Figure 11.1A shows a clear counterexample. You perceive a situation in which the figures can walk up or down the stairs infinitely while returning to their initial position after each turn. This situation is physically impossible, and thus has an a priori likelihood of exactly zero. Why can we see a situation with a zero a priori likelihood? One might argue that

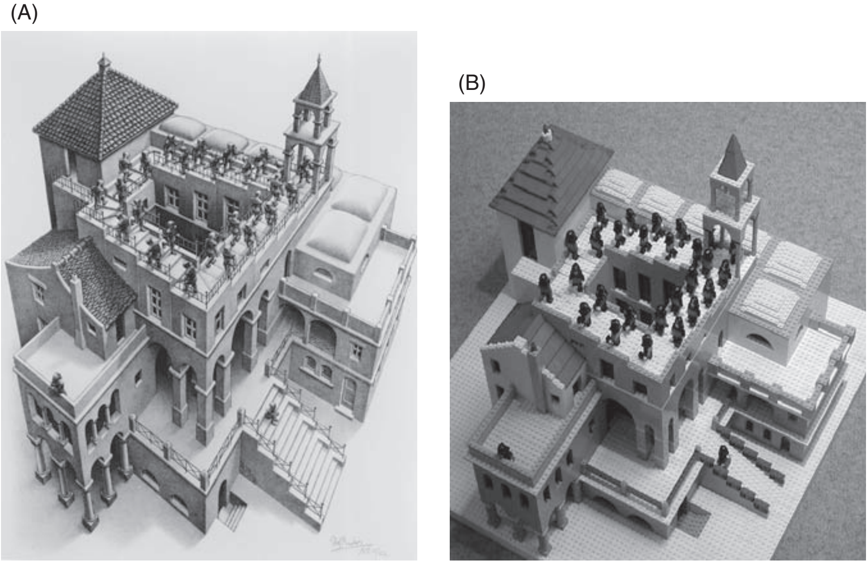


FIGURE 11.1 Perceiving “impossible” constructions. (A) M.C. Escher’s “Ascending and Descending” www.mcescher.com (B) A similar construction has been built in LEGO™. Despite the fact that there is a possible construction leading to this image, we perceive the impossible construction. Details of the construction can be found at the web site: http://www.andrewlipson.com/escher/ascending.html (© Andrew Lipson, reproduced with permission).

perceiving an impossible situation in a picture is not problematic because the a priori chance that something impossible is depicted is not zero at all. It is like looking at a photograph of yourself as a 6 months old baby; you perceive something that you know is not reality anymore, but history. In a similar way, Escher’s drawing might be thought of depicting not reality, but fantasy. One might also argue that because there is no real 3D construction that gives this image, there is no a priori likelihood, so we cannot use Bayesian inference.

Unfortunately, the reasoning above is not correct, as there is a possible 3D construction that leads to an image as in Escher’s drawing: Andrew Lipson built it as a *LEGO*™ construction and photographed it (Figure 11.1B). Everybody perceives this picture as depicting the same impossible 3D situation as Escher’s original. However, as the a priori likelihood of an impossible object is zero, Bayesian inference would predict that this percept can never occur. Moreover, the a priori likelihood that the actual LEGO construction depicted here exists is definitely not zero, so Bayesian inference would predict that you perceive the actual construction. Many other “impossible” drawings (such as the Penrose triangle; Gregory, 1968) give the same retinal image as real objects. The question we will address in the remainder of this chapter is how we can understand the perception of the picture in Figure 11.1B within the Bayesian framework.

DETECTING ATTRIBUTES

We will start by sketching some textbook knowledge about the neural basis of visual perception. Textbooks start with the receptors at the level of the retina: the rods and three types of cones that respond to the incoming light. Subsequently, still at the level of the retina, information from various receptors is combined to improve the sensitivity for a certain aspect of information. For instance, lateral inhibition increases the sensitivity for local luminance differences at the expense of losing information about the absolute luminance level. This mechanism (very useful for detecting object edges) is commonly regarded as being on the basis of various illusions, such as Mach bands and the Cornsweet illusion. In these explanations, the luminance difference is caused by an erroneous integration of the luminance differences obtained by the edge detectors (Land & McCann, 1971; Arend, 1973). Why would the brain first take a spatial derivative and subsequently integrate it? It is an efficient way of coding, making it possible to transmit information about differences in reflectance that are orders of magnitude smaller than the variations in illumination. Despite the limited bandwidth of the optic nerve, we can recognize objects and other animals both in the sunlight and in the dark.

To explain the illusions using the image coding based on edge detectors, one implicitly assumes that the brain makes systematic errors in the differentiation and/or subsequent integration (the integration of a perfect derivative wouldn't yield any error). Thus the essence of the explanation is not the use of edge detectors but the systematic errors that are made. What is the reason for making such systematic errors? These illusions are the consequence of an ambiguity in images: Are differences in lightness due to differences in illumination or due to differences in the surface reflectance?

In most images, there are various other cues for the illuminant and surface properties. For instance, the illumination is likely to vary much more with position for curved surfaces than for flat surfaces (Figure 11.2). By varying the presence of such other cues, it has been shown that the perception of the equiluminant territories flanking the Cornsweet edge varies according to whether these regions are more likely to be equally illuminated surfaces having different material properties or unequally illuminated surfaces with the same properties (Knill & Kersten, 1991; Purves et al., 1999). The illusion is thus not the consequence of low-level processing errors, but a percept that is optimal from a Bayesian perspective. In a similar way, the presence of Mach bands can be explained in terms of the likelihood of photometric highlights near contrast edges (Lotto et al., 1999).

There is an interesting difference between the attributes luminance and local-luminance gradient. Luminance itself is very sensitive for naturally occurring slow variations of illumination over a smooth surface, whereas such variations are negligible at the scale of the edge detectors. So the prior information needed to reliably determine luminance itself is not useful for determining edges on the basis of local luminance gradients. Although the Bayesian approach yields optimal estimates for both attributes (luminance and luminance gradient), it has the

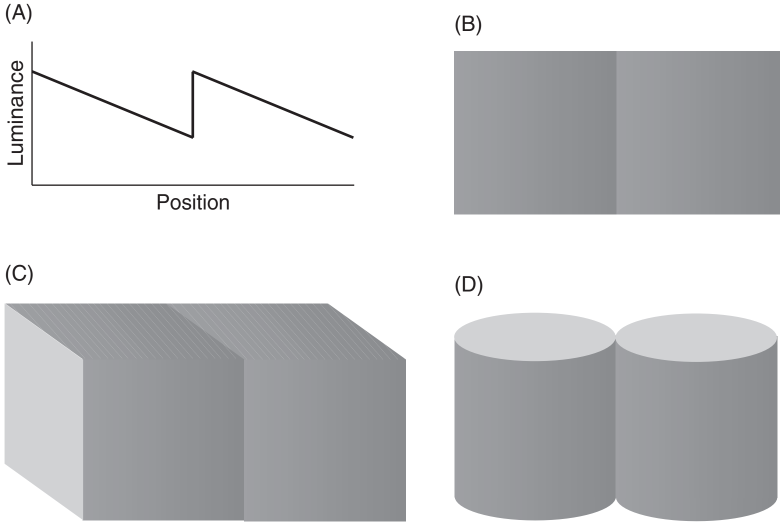


FIGURE 11.2 The Cornsweet illusion: if two surfaces with equal luminance gradients are presented next to each other, one perceives the two as having unequal brightness (A, B). This illusion depends on the interpretation of the scene: it is strong if the surfaces seem to be flat (rendered as part of cubes), (C) than if the surfaces that seem curved (rendered as part of cylinders), (D). For the cylinders, the luminance gradients and luminance step are assumed to be caused by differences in illumination of the surfaces (due to the varying orientation relative to the assumed light source left above), whereas for the cubes, a constant luminance gradient due to the illumination is assumed, with in addition a step in reflectance at the border of the two cubes.

side effect that the perceived luminance gradients might differ from the gradient of perceived luminance if the actual situation is not very likely. In other words, Bayesian perception might be inconsistent. Should this bother us?

SPATIAL PERCEPTION

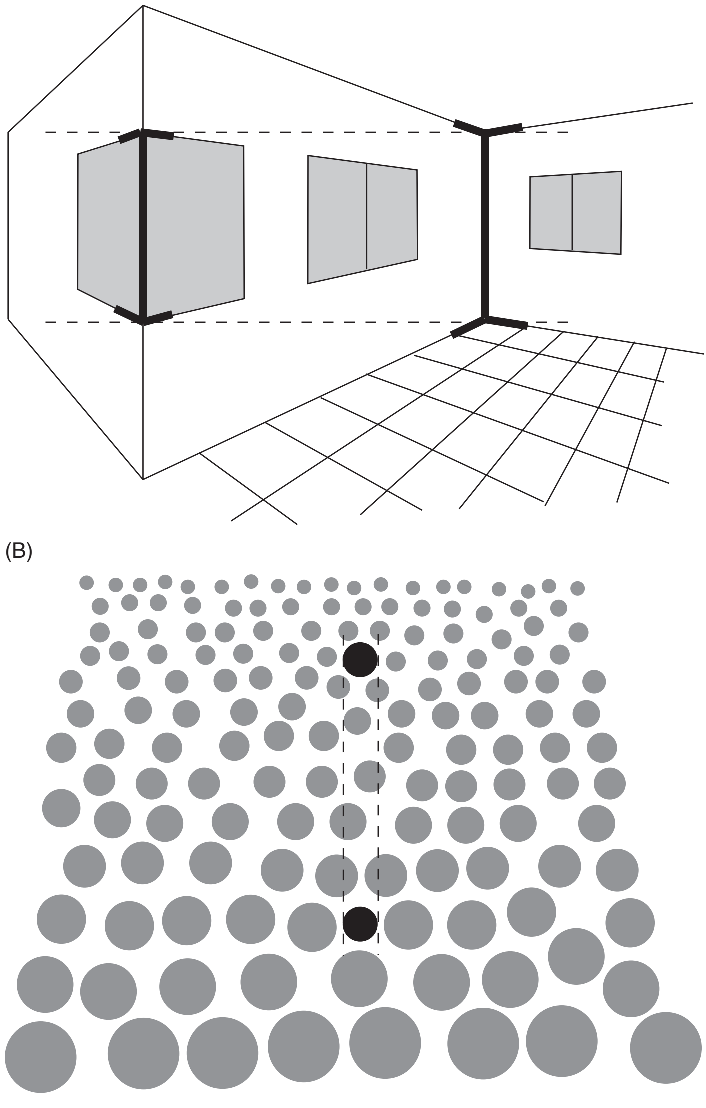
This inconsistency between the perception of an attribute and the perception of related attributes, such as difference measures and derivatives, is also the basis of many illusions in the spatial domain. The absolute measure of interest here is position, and the related difference measures are distance, length, and velocity. We are notoriously imprecise in determining the absolute position of objects in space. The resolution is about 0.5° (van Beers et al., 1998), presumably a combination of a limited resolution of eye orientation of about 0.15° (Smeets & Brenner, 1994; Brenner & Smeets, 2000) and that of head orientation. The visual acuity of a person with normal vision is one minute of arc, which is about 10 times as precise as our perception of location. The reason is that visual acuity is only determined by the properties of the retina, and is therefore independent of the low resolution of information about the orientation of the eye.

Motion is determined by motion detectors that compare the activation of two areas in the visual field that are separated by a distance (span), with a characteristic time delay at which the activity of the two areas is compared. The smallest size ("span") of motion detectors is in the order of the retinal resolution; about two minutes of arc (van Doorn & Koenderink, 1982). This is much smaller than the resolution at which position can be determined. The reason for the high precision is that motion detectors do not differentiate egocentric position, but determine the motion relative to other retinal input (Smeets & Brenner, 1994). This way of calculating motion has the consequence that motion perception can be inconsistent: you can see an object moving without changing position. This can be perceived in the Duncker illusion (motion perception of a stationary object induced by motion of the background (Duncker, 1929)) or in the aftereffect after prolonged exposure to motion.

Binocular vision (the perception of spatial layout based on the difference in information between the two eyes) is regarded as normal when stereoacuity is better than one minute of arc. How is this achieved in a situation in which each of the eyes has a precision that is not better than 10 minutes of arc? The story is again in the information that is used: the threshold for stereoacuity is based on relative disparity, the differences between the images of the two eyes, irrespective of the location of these images on the retina. Perceiving spatial layout is based on a difference measure that is insensitive to the least precise information available for egocentric localization. This means that shapes and relative positions can be determined very accurately. However, when we need to localize an object in depth relative to ourselves (instead of relative to other visual items), we have to rely on information about eye orientation, so that the precision is limited by the 0.15° resolution of eye orientation (Brenner & Smeets, 2000).

A similar reasoning holds for the perception of the size of objects. If the perception of size would be based on the difference between the judged egocentric positions of the object's edges, than the precision would be rather limited. Fortunately, there is a much better solution. If one bases ones size judgment on the retinal size, scaled by an estimate of distance, precision can be enhanced, as long as one can get a reliable judgment of distance. This is possible only if one does not limit oneself to extra-retinal information, but uses all available visual cues like perspective, familiar size, and texture gradients in a Bayesian way (Gregory, 1968). The Müller-Lyer illusion has an interpretation in terms of depth perception (Figure 11.3A). As can be seen in the same figure, this interpretation affects the perceived length of the line, but not other aspects of space, such as the perceived orientation of the dashed lines connecting their endpoints. The Ebbinghaus illusion can also be interpreted in terms of perspective (Figure 11.3B), and again this affects the perceived size of the disks, but not other spatial aspect in the figure, such as the parallelity of the dashed lines connecting the edges of the black disks. Along the same lines, it has been shown that retinal and extra-retinal information used to judge an object's size, shape, and egocentric distance are combined in a way that yields the most likely value for each of these attributes independently, ignoring any resulting inconsistency between the attributes (Brenner & Van Damme, 1999).

(A)



(B)

FIGURE 11.3 Examples of inconsistencies in illusions of depth. (A) The Müller-Lyer illusion and other perspective cues make the thick line on the left look smaller than the one on the right. At the same time, the horizontal dashed lines connecting the endpoints seem to be parallel (which they are). (B) The two black disks are equal in size, but the upper one seems to be larger due to the smaller surrounding gray disks. This version of the Ebbinghaus illusion only affects size, not other aspects of space. For instance, the dashed lines look parallel (which they are), which is inconsistent with the apparent difference in size of the two disks.

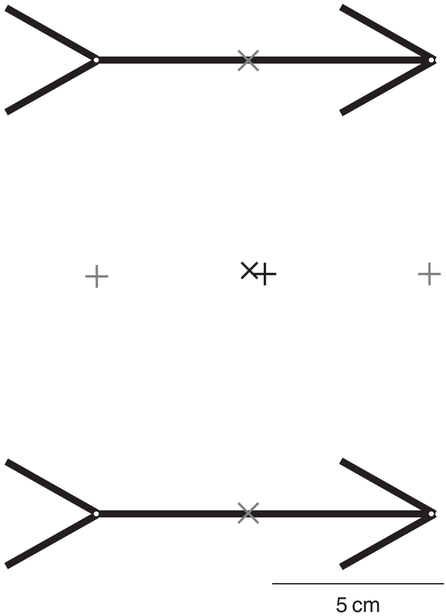


FIGURE 11.4 The center between the four white dots at the endpoints of the Judd-figures is determined in two ways. First, the pairs on each Judd-figure are bisected (gray x's), and then the midpoint between these two points (black x) is determined. Secondly, the midpoint between each vertical pair is found (gray +'s), and then the midpoint between these two points is determined (black +). Although the two ways should yield the same result (according to any affine geometry), the outcome is systematically different.

The inconsistency in the examples discussed earlier is between different attributes. The expected relationships between attributes are not present. Two line segments of different length are connected by parallel lines (Figure 11.3A), and the lines connecting sides of differently sized disks seem parallel (Figure 11.3B). This clearly defies Euclidian geometry. Is there another (non-linear) geometry that can describe human perception? A simple experiment shows that this is very unlikely. In a pencil-and-paper task, we asked subjects to judge the center of four white dots with two Judd-figures attached to them (Figure 11.4). We instructed them to perform the judgment in two ways. On the first sheet of paper, they were asked to bisect the horizontal distances (and thus the Judd-figures) first (gray "x"), and subsequently bisected the vertical distance between the resulting positions (black "x"). On the second sheet of paper, they started with two vertical bisections (gray "+"), and subsequently bisect the horizontal distance between the resulting positions (parallel to the Judd-figures, black "+"). The resulting center differed systematically between the two variants of the tasks. Such a result is not a simple consequence of a non-Euclidian (but nevertheless affine) space. A similar task has been performed to study the spatial deformation in 3D space (Todd et al., 2001). In that experiment, the center between four positions was systematically misjudged, but this misjudgment was independent of the order of bisections. Geometric illusions are thus essentially different from normal perceptual misjudgments.

INCONSISTENT ACTION

It might seem quite disturbing that there can be inconsistencies between attributes within perception. Is this inconsistency not far from optimal? It probably would if the purpose of perception would be to create an internal representation of the outside world. But the purpose of perception is to let an organism survive, for instance by allowing him to find food or to flee for a predator. One might think that organisms combine all information to make the best plan for a movement. For instance, if an animal wants to catch a running prey, it could combine all information about position and motion to predict the time and the position of interception. Is this how animals act? Experiments on human interception show that this is not the case. By using motion illusions, we showed that position, direction of motion and a priori estimate of speed are used to direct the hand, whereas speed information is used to time the action (Smeets & Brenner, 1995a, b; Brenner et al., 2002a).

Inconsistency in our actions can even be observed for the simple task of moving our arm from point A to point B. This inconsistency is caused by the fact that although knowing the target position is enough to move your hand to the target, this information is not enough to move along a straight line (which is what we normally tend to do). To move along a straight line to a target, we need to know in what direction to start our movement. This initial movement direction is often not correct (de Graaf et al., 1991; Brenner et al., 2002b), can be adapted independently of the location of the endpoint (Wolpert et al., 1995), and is easily influenced by illusions (Smeets & Brenner, 2004).

So, not only our conscious perception is inconsistent, but the use of spatial information in our actions is just as inconsistent. This is not surprising, as consistency is not important to find food or to flee for a predator. It is of utmost importance to have fast access to relevant information, such as the velocity of the predator and its position. Although it might be that further processing and combining can improve information, the animal would already be caught before the final percept was completed.

COMBINING INFORMATION

But even without the temporal constraints, it may not be useful to try to make all attributes consistent. As argued earlier, the brain uses all available information to make the best estimate for an attribute. By making attributes consistent, one has to change the values from these optimal values, which is—by definition—sub-optimal. This is similar to an issue in the cue-combination literature. It is well established that if two cues are in conflict, this conflict is not resolved: the cues remain in conflict, although this conflict might not be noted explicitly (Hillis et al., 2002; Muller et al., 2007).

The same holds for combination between senses. To know where our hand is, we have visual and proprioceptive information. We normally don't realize that we have these two sources, because we only have a single idea of where our hand is. But when closing our eyes, we realize that we still know where our hand is. The interesting aspect is that it is easy to induce conflicts between the senses, for instance by wearing wedge prisms. We don't perceive the discrepancy, only realize that we make errors, and adapt our behavior accordingly (van Beers et al., 1999).

But discrepancy between proprioception and vision is not restricted to experimental manipulations. If you put subjects in the dark, and let them make ample back-and-forth movements with their hands between visual targets without seeing their hand, they start making errors. These errors are not accidental: the same errors reoccur on repetition of the experiment the next day (Smeets et al., 2006). So our senses are not calibrated. The reason for this lack of calibration is similar to the reason why the inconsistencies are not resolved: if the combination of senses yields the most reliable estimate, recalibration can only reduce the reliability of the information. It is for instance not clear which of the modalities would need to be recalibrated. Is this lack of calibration problematic? Again, it is not. When controlling our hand movement, we don't use a single sense but use the optimal combination of all senses. And this does not only hold for our hand but also for any possible target we want to reach for with that hand. This means that the conflict might be present between attributes and between senses but that these conflicts do not interfere with our performance.

CONSCIOUS PERCEPTION

We made our argument in terms of the information needed for controlling our actions. We reached a radically different conclusion than for instance Goodale (2001), who claims that "accurate metrical information about an object" is needed to guide one's action. In our view, the same erroneous and inconsistent perceptual information can be used in both perception and controlling movements (Smeets et al., 2002). There is however one fundamental difference: whereas timely information is essential in the control of movements, our consciousness has ample time to reconsider information. Whereas control of action needs to rely on the fast feed forward processing of information, our cognition can wait until the information processing is recurrent (Lamme & Roelfsema, 2000). The consequence that our conscious percept is based on further processed information than used to control our actions, but this should not be taken as evidence for independent processing.

We started this chapter by discussing the limitations of the Bayesian approach, and argued that the inconsistent precepts are not very Bayesian. We continued by showing that inconsistency is very widespread in perception, and is a consequence of optimally determining information about each attribute of the world around us. These optimal estimates are enough to guide our actions. An exact calibration of perception is not needed for controlling ones actions. What we need are transformation rules between perceptual attributes and aspects of an action. That is what we learn very quickly while practicing an action.

REFERENCES

Arend, L. E. (1973). Spatial differential and integral operations in human vision: Implications of stabilized retinal image fading. *Psychological Review, 80,* 374–395.

Brenner, E. & Smeets, J. B. J. (2000). Comparing extra-retinal information about distance and direction. *Vision Research, 40,* 1649–1651.

Brenner, E. & Van Damme, W. J. M. (1999). Perceived distance, shape and size. *Vision Research, 39,* 975–986.

Brenner, E., De Lussanet, M. H. E., & Smeets, J. B. J. (2002a). Independent control of acceleration and direction of the hand when hitting moving targets. *Spatial Vision, 15,* 129–140.

Brenner, E., Smeets, J. B. J., & Remijnse-Tamerius, H. C. (2002b). Curvature in hand movements as a result of visual misjudgements of direction. *Spatial Vision, 15,* 393–414.

de Graaf, J. B., Sittig, A. C., & Denier van der Gon, J. J. (1991). Misdirections in slow goal-directed arm movements and pointer-setting tasks. *Experimental Brain Research, 84,* 434–438.

Duncker, K. (1929). Über induzierte Bewegung (Ein Beitrag zur Theorie optisch wahrgenommener Bewegung). *Psychologische Forschung, 12,* 180–259.

Goodale, M. A. (2001). Different spaces and different times for perception and action. *Progress in Brain Research, 134,* 313–331.

Gregory, R. L. (1968). Perceptual illusions and brain models. *Proceedings of the Royal Society London B, 171,* 279–296.

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science, 298,* 1627–1630.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55,* 271–304.

Knill, D. C. & Kersten, D. (1991). Apparent surface curvature affects lightness perception. *Nature, 351,* 228–230.

Knill, D. C. & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27,* 712–719.

Körding, K. P. & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences, 10,* 319–326.

Lamme, V. A. F. & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571–579.

Land, E. H. & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America, 61,* 1.

Lotto, R. B., Williams, S. M., & Purves, D. (1999). An empirical basis for Mach bands. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 5239–5244.

Muller, C. M. P., Brenner, E., & Smeets, J. B. J. (2007). Living up to optimal expectations. *Journal of Vision, 7*(3), 1–10.

Purves, D., Shimpi, A., & Lotto, R. B. (1999). An empirical explanation of the Cornsweet effect. *Journal of Neuroscience, 19,* 8542–8551.

Smeets, J. B. J. & Brenner, E. (1994). The difference between the perception of absolute and relative motion: A reaction time study. *Vision Research, 34,* 191–195.

Smeets, J. B. J. & Brenner, E. (1995a). Perception and action are based on the same visual information: Distinction between position and velocity. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 19–31.

Smeets, J. B. J. & Brenner, E. (1995b). Prediction of a moving target's position in fast goal-directed action. *Biological Cybernetics, 73,* 519–528.

Smeets, J. B. J. & Brenner, E. (2004). Curved movement paths and the Hering illusion: Positions or directions? *Visual Cognition, 11,* 255–274.

Smeets, J. B. J., Brenner, E., de Grave, D. D. J., & Cuijpers, R. H. (2002). Illusions in action: Consequences of inconsistent processing of spatial attributes. *Experimental Brain Research, 147,* 135–144.

Smeets, J. B. J., van den Dobbelsteen, J. J., de Grave D. D. J., van Beers, R. J., & Brenner, E. (2006). Sensory integration does not lead to sensory calibration, *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 18781–18786

Todd, J. T., Oomes, A. H. J., Koenderink, J. J., & Kappers, A. M. L. (2001). On the affine structure of perceptual space. *Psychological Science, 12,* 191–196.

van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1998). The precision of propioceptive position sense. *Experimental Brain Research, 122,* 367–377.

van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology, 81,* 1355–1364.

van Doorn, A. J. & Koenderink, J. J. (1982). Spatial properties of the visual detectability of moving spatial white noise. *Experimental Brain Research, 45,* 189–195.

von Helmholtz, H. (1925). Treatise on Physiological Optics, Vol. III, The Perceptions of Vision. Birmingam, Alabama: The Optical Society of America.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). Are arm trajectories planned in kinematic or dynamic coordinates – an adaptation study. *Experimental Brain Research, 103,* 460–470.