Perceiving motion: relativity, illusions and the nature of perception

Alexander H. Wertheim

An extended version of Von Holst and Mittelstaedt's model is used to illustrate that the visual perception of motion can be described in terms that are compatible with the relative nature of the physical concept of motion. The model concerns how we perceive object motion, not only during eye movements, but also during head and ego motion of the observer, and with the inclusion of a noise factor. Various rather counterintuitive illusions deriving from this view are described and some philosophical consequences as to how we perceive reality are discussed. (*Netherlands Journal of Psychology*, 64: 119-125.)

Keywords: illusion; motion; perception; vision

Motion, stationarity, velocity and the direction of motion are relative concepts. Physicists only discuss motion as relative to the frame of reference which defines that particular motion. We may drive at 100 miles an hour (relative to the road), and at the same time remain stationary (relative to another car driving next to us at the same speed). If the other car moves faster than we do, we move backward relative to it, and when it moves slower we move forward relative to it, which illustrates that motion is a symmetrical concept. A moving relative to B physically means the same as B moving relative to A.

In perception, however, this does not seem to be the case. Here it appears as if motion is an object property and symmetry does not seem to make any sense. We see cars move relative to the road, never a road relative to a car. It is as if our brain transforms the relational concept of motion into an object property. The easiest way out of this dilemma is to claim that concepts of physics do not necessarily apply to the psychology of perception (e.g. Gibson 1979). But in recent years that claim is heard less and less often, the current general consensus being that scientific concepts should not be contradictive, within, nor between different fields of science. If that nevertheless happens (e.g., relativity theory *vs* quantum mechanics) it indicates an unsolved problem, usually thought of as a challenge, not as a solution.

With respect to the relativity of motion, we are confronted with exactly such a problem. The aim of this paper is to show how this problem can be resolved and which empirical and philosophical consequences the solution has.

Utrecht University
Correspondence to: Alexander H. Wertheim, Utrecht University, Heidelberglaan 2, NL 3584 CS Utrecht, e-mail: a.h.wertheim@uu.nl

Received 11 December 2007; revision accepted 7 July 2008.

Percepts of object motion: extending Von Holst and Mittelstaedt's model

The solution derives from the realisation that retinal image movements do not necessarily generate percepts of object motion. Take, for example, the case of an airplane passing high over a tree top. If we fix our gaze on the tree top, the image of the tree does not move across the retinas and we see the tree as stationary. However, if we follow the plane with our eyes, its image also remains (more or less) fixed on our retinas. But even so the plane is seen to move. Conversely, when we look at the tree, the image of the plane moves across the retina and the plane is seen to move. But when again we now move our eyes to follow the plane, the image of the tree moves on the retinas, but the tree is nevertheless seen as stationary. It thus follows that retinal image motion cannot really be a stimulus for perceived motion, nor can the absence of retinal image motion be a stimulus for perceived stationarity. How can we explain that?

Helmholtz (1910) most famously proposed a solution to this problem while considering the question of why the world remains stationary in perception during eye movements that cause its image to move on the retinas. He assumed that the effort to move our eyes somehow makes the brain cancel image movements caused by eye movements. This idea was later formalised and made symmetrical in Von Holst and Mittelstaedt's well-known model (Von Holst & Mittelstaedt 1950; Von Holst 1954, Mittelstaedt, 1990). Their model holds that to see motion at least two neural signals are necessary, signals that can be conceptualised as velocity vectors. One signal, here to be termed the 'retinal signal', encodes how the image moves on the retinas. The other signal, which is reminiscent of Helmholtz's effort to move the eyes, encodes how the eyes move in their orbits. It presumably consists of a neural copy of the efferent impulses to the eye musculature. Hence, it has become known as the 'efference copy' signal. The model states that the brain subtracts these signals from each other, and that it is the difference between them, not the signals themselves, which serves as the stimulus for seeing motion. Neither the retinal signal, nor the efference copy, have any perceptual significance in themselves. Nor do their magnitudes matter. It is their vectorial difference that makes us see motion. Percepts of object motion and velocity correspond to the magnitude of this difference. When that difference is zero, the perceived velocity of the object that creates the retinal signal is also zero and we perceive the object as stationary. We see motion only if the difference is not zero, and we then see the object move with a velocity that corresponds to the magnitude of that difference.

Accordingly, in the present example the plane is never seen as stationary, because the difference between the two signals is never zero: either we follow the plane with the eyes, in which case the retinal signal is zero and the efference copy is not, or we keep our eyes fixed on the tree, which means that the efference copy signal is zero and the retinal signal of the plane is not. And the tree, like Helmholtz's world, is always seen as stationary because here the difference between the two signals is always zero, either because the two signals themselves are both zero (when we fixate the tree with our eyes), or because they are both non-zero but equal in size (when the retinal image motion of the tree results from our eye movement).

Although Von Holst and Mittelstaedt's model is widely accepted in the literature on motion perception, it is too simple and incomplete.

To start with: two neural signals are not just equal or different. A neural signal consists of a firing pattern, i.e. it is inherently noisy, because it is always very slightly irregular over time (see e.g. Wertheim et al., 1985). And we know that to be detectable a difference between two signals has to be larger than their noise level. In psychophysical terms this is known as the Just Noticeable Difference (JND) between the two signals. Hence, to perceive motion the difference between the retinal signal and the efference copy signal must be larger than the JND between them. The JND, in other words, defines the threshold for seeing motion.

The JND is a factor that was not taken into account by Von Holst and Mittelstaedt (see e.g. Wallach & Kravitz, 1965 and MacKay 1973, who were the first to note this deficiency). How big, then, is this JND? The answer comes from Weber's law, which states that the JND is always a fixed percentage (known as Weber's fraction) of the magnitude of the signals involved. This provides a way to measure the magnitude of the JND and at the same time to test Von Holst and Mittelstaedt's model.

Imagine a small fixation mark that smoothly moves, at a fixed velocity, from left to right across a large screen. On the screen we also project a large visual background pattern, e.g. a grating, which does not move. Subjects are then asked to track the fixation mark with their eyes. While doing that, they see the background as stationary on the screen, because its retinal signal and the efference copy have the same magnitude: the velocity of retinal image of the background (which defines retinal signal magnitude) is equal to the velocity of the eyes (which defines the magnitude of the efference copy). Now assume that during ocular pursuit of the fixation mark we slowly move the background in the same direction as the fixation mark, i.e. in the same direction as the eye movement. The magnitude of the retinal signal (of the background) is then slightly reduced. If with each consecutive sweep of the fixation mark we gradually increase background velocity, we can do so until the subject for the first time reports that the background is seen to move concurrently with the eye movement. At that point the threshold for background motion during the eye movement is reached, which means that the retinal signal has now become exactly one JND smaller than the efference copy signal. We then replicate the procedure for background motion in the direction opposite to the fixation mark, causing the magnitude of the retinal signal (of the background) to slightly increase. When that threshold is reached the retinal signal is one JND larger than the efference copy. Now we can make a prediction: if the fixation mark moves faster, the eyes will move faster as well. That causes both the efference copy signal and the retinal signal to increase. And thus, so should the JND between them, because according to Weber's law, a JND is a fixed fraction of signal magnitude. Consequently, the two thresholds for background motion should increase linearly with eye velocity. In a critical experiment the hypothesis was indeed confirmed (Wertheim 1981). The results provided strong support for Von Holst and Mittelstaedt's model, because this finding can only be explained by the assumption that two signals are involved in motion perception, their difference being the actual stimulus. In addition, the model is extended with a noise component, the magnitude of which corresponds to the threshold for perceiving motion.

However, another important aspect is still missing from the model. Imagine we repeat the experiment, but now the subject does not track the fixation mark with the eyes, but with the head, rotating the head such that the eyes remain stationary in their orbits. That would make the efference copy zero (the eyes are not moved in the head), but the retinal signal of the background would remain large. That would mean we should see the background move. But that does not happen. In such a situation a stationary background normally is still seen as stationary.

The solution to this problem is of course that the model should not be based on a signal that encodes how the eyes move in their orbits, but on how they move relative to external space. Von Holst and Mittelstaedt's model only applies to the special case where we do not move our head. Only then do eye movements in the head equal eye movements in space (see e.g. Swanston & Wade, 1988 for a similar argument).

So we need to assume that percepts of how objects move relative to external space do not arise from the difference between a retinal signal and an efference copy signal, but from the difference between a retinal signal and a signal encoding how the eyes move relative to external space. The latter signal has been termed the 'reference signal' (Wertheim, 1994). A reference signal should be generated whenever an observer's eyes are, for whatever reason, made to move in space, that is during any kind of self motion, whether this be eye movements, head movements or whole body (ego) movements.

A reference signal thus can be conceptualised as consisting of various components, each one being generated from afferent signals that inform the brain about these different kinds of self movements. Taken together these components form the necessary information about how the eyes move in space. One component has already been discussed: the efference copy. It contributes to the generation of a reference signal by providing information about how the eyes move relative to the head. Information about head and ego movements in space, stemming from vestibular afferents, forms a second component in reference signals. And there is a third component. The point is that we can also have ego-motion sensations in response to optic flow (this is usually termed 'vection'). Hence optic flow should also contribute to the reference signal. Thus, the reference signal can be envisioned as a compound signal composed of at least three components: an efference copy component, a vestibular component and an optic flow component.

This reasoning enables us to predict new phenomena. For example, thresholds for seeing motion should increase whenever reference signals increase in magnitude, because larger reference signals are noisier, i.e. they generate a larger JND. Hence, during any kind of self or ego motion that makes the eyes move in space, including driving in cars, airplanes, swivel chairs etc., thresholds for seeing object motion in space should increase and perceived object velocity in space should correspondingly be reduced (velocity underestimation).

Experiments in which subjects are moved on a moving chair or a sled in front of a screen showed these predictions to be correct (see Wertheim 1994 for a review). More recently, the same phenomenon has been demonstrated in subjects who were moved in a moving base car simulator while viewing visual stimuli moving on the screen of the simulator. Not only did the thresholds increase, but above threshold stimulus motion was underestimated correspondingly (Filliard et al., 2008).

Relativity

We now return to the relativity of perceived motion as illustrated in figure 1. The velocity of the eyes in space (magnitude of the reference signal) is plotted horizontally and the velocity of the image of a particular object on the retinas (magnitude of the retinal signal) is plotted vertically. The dotted line indicates where retinal and reference signals are equal. Thus, along that line the object is seen as stationary in space. Along the sloped line above the dotted one, the retinal signal is just one JND larger than the reference signal. Therefore, this line indicates the threshold for seeing the object move in space in the direction opposite to the direction in which the eyes move in space. Above that line we do see the object move in that direction and with a velocity that corresponds to the vertical distance to that threshold. Along the other sloped line, the one below the dotted line, retinal signals are one JND smaller than the reference signal. Thus, that line indicates the threshold for seeing object motion in space in the same direction as the eyes move in space. Below that line we are above the threshold for object motion in that direction and see the object move with a velocity corresponding to the vertical distance from that threshold.

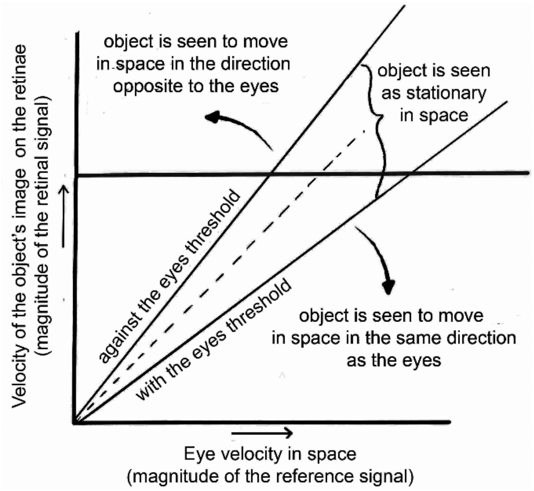

Figure 1
The relativity of perceived motion. See text for explanation.

The vertical distance between the two threshold lines thus indicates twice the magnitude of one JND between retinal and reference signals. Hence, in the area between the two drawn lines we cannot see the object move in space.

Now let us consider a retinal image which moves at a fixed velocity across the retinas. This is illustrated with the thick horizontal line in figure 1. If we move from left to right along this line, retinal image velocity remains unchanged, but the reference signal increases (i.e. the eyes move faster in space). On this line, close to the Y-axis, we see object motion in the direction against the eyes, and the more we move to the right the slower that motion will appear to be, until the threshold is reached. We then pass a range of eye velocities during which that same retinal image velocity is interpreted as object stationarity until the threshold is reached for seeing the object move in the other direction, i.e. in the same direction as the eyes. And the further we move to the right the higher its perceived velocity. Thus motion, stationarity, velocity and the direction of motion all depend on the relation between a retinal and a reference signal. In other words, we have complete relativity and symmetry: the physical relativity of motion, it being a symmetrical relation between two objects, is mirrored at the neurological level: perceived motion stems from a symmetrical relation between two neural signals: their difference.

Predicting illusions from the model

Thus we have found a solution to the problem with which we started: how to make the psychology of perception compatible with the language of physics. The perception of motion can now be described in terms of a symmetric relation between two signals, just like physical motion is defined as a symmetrical relation between two objects. It can be fun to play with this concept. For example, we may draw a perceptual parallel to Einstein's theory of relativity according to which light is moved away from a straight line in the presence of a gravitation field. Imagine that we are sitting in a rocket, drifting in empty space and looking in total darkness at a small stationary visual stimulus. We will then see the stimulus as stationary because its retinal signal is zero. Since we have no information about our own movement in space our reference signal will also be zero. Now imagine that a big star passes closely behind us while our computerised rocket begins to blast its engines automatically in order to keep the distance to that star constant. The star's gravity then acts on our vestibular system, which basically behaves as an accelerometer, gravity being an acceleration. (If we were to fall freely towards the star the vestibular system would not react and we would not sense any gravity). Since vestibular afferents generate a sense of ego motion, they also generate a reference signal. But the retinal signal remains unchanged. Consequently we will see our stimulus move. In other words, percepts of motion or stationarity relative to external space change in a gravitation field.

This, of course is just a funny thought experiment. But the mechanism explained above can also predict some curious illusions that could actually occur in real-life situations. For example, if we watch a looming optic flow pattern on a screen while riding a bicycle ergometer (see figure 2), or while running on a treadmill, we may 'fool' the brain into believing that we, and thus our eyes, are moving in space. Consequently a reference signal should be generated and with it a higher noise level, causing an elevation of the threshold for object motion. As a result we should underestimate the velocity of the looming pattern and, if it moves slowly enough, perceive the pattern as completely stationary. In several pilot studies at the Max Planck Institute in Tübingen, Germany, we have indeed observed this phenomenon (see also Wertheim & Reymond, 2007). In fact, such elevations of object motion thresholds during real or illusory self motion can actually be found in the literature (Buechele, Degner, & Brandt, 1980; Probst, Brandt, & Degner, 1986; Wallach 1987; Pelah & Thurell 2001; Thurell & Pelah 2002; Durgin, Gigone, Scott, 2005), although in those papers the present explanation of these effects was not proposed.

Figure 2
The looming image appears to stop moving when the subject 'fools' his brain that he is moving, i.e. that his eyes move in space.

Recently, an even stranger prediction was supported in a first pilot study in the Renault car simulator near Paris, France. Here we induced slow vection, simulating linear forward ego motion with a complex slowly looming visual pattern projected on a large screen inside the simulator. When subjects sensed linear forward vection, i.e. when they perceived forward ego motion, we accelerated the simulator forward. A vestibular signal should then be generated. According to our model this should add a component to the reference signal, thereby enhancing its magnitude. Given the consequent increase of the JND between retinal and reference signals, this should result in an increase of the threshold for perceiving object motion, causing the pattern to appear stationary on the screen. But since viewing a pattern that appears to be stationary does not result in vection, the sensation of ego motion should stop. This is indeed what happened. Thus, here we have an example of how the addition of a vestibular signal may suppress a sensation of ego motion rather than enhance it.

Another example of a rather counterintuitive effect which derives from the present model is illustrated in figure 3. It can easily be perceived inside a car when viewing a slowly moving grating on a laptop. If the car is stationary or moves at a constant speed while we look at the grating, we will see the grating move because there is a retinal signal but no reference signal (the eyes and head do not move). However, when the car accelerates, breaks or takes a sharp turn, the vestibular systems of the people inside the car react and a reference signal is generated. However, since the laptop remains stable in front of our eyes, the retinal signal does not change. When the magnitude of the vestibularly induced reference signal increases enough to approximate that of the retinal signal by less than one JND, the grating suddenly appears to be stationary on the screen. The illusion is so strong that it can even be perceived if one holds the laptop and walks around with it, or when being pushed while moving on a bureau chair with wheels. In fact this effect could explain roadway accidents due to velocity underestimation of traffic, especially when a driver accelerates or takes a sharp bend at high velocity to enter a highway. Similarly, visual stimuli moving at a constant velocity inside a head mounted display will also be seen to stop moving when the head moves (Pavard & Berthoz, 1977).

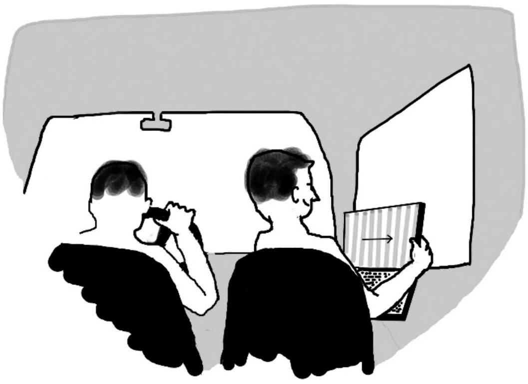

Figure 3
When the car accelerates and stimulates the vestibular system, a slowly moving grating is seen to stop moving across the screen of a laptop.

Some philosophical consequences

The relativity of perceived object motion is also quite interesting from a philosophical point of view. The point is that perceived object motion results not from either an externally induced retinal signal, or a self-generated reference signal, but from their vectorial difference. This illustrates that percepts come about as the consequence of an interaction between the physical world and the subjective world, and cannot be reduced to something sprouting from either world alone. This reminds one of the Kantian view according to which we cannot 'directly' perceive things purely as they physically are. Translated into the language of psychology this means that we cannot generate percepts that correspond purely to the physically induced excitation of our sensory systems. We always need an interpreting internally generated signal as well. Or do we?

Let us consider the perceived velocity of two cars. Car A moves at 100 km/h, and car B moves at 80 km/h. According to our model, the perceived velocity of car A corresponds to:

$$\mathrm{Vperc}_A = \mathrm{Vret}_A - \mathrm{Vref} - \mathrm{JND} \quad (1)$$

where Vperc_A is the perceived velocity of car A on the road, Vret_A is the velocity of the image of car A across the retinas, Vref is the reference signal, and the JND stands for the Just Noticeable Difference between the two signals which can be interpreted as intrinsic signal noise.

Correspondingly, the perceived velocity of car B is:

$Vperc_B = Vret_B - Vref - JND$ (2)

where $Vperc_B$ is the perceived velocity of car B on the road, $Vret_B$ is the velocity of the image of car B across the retinas, Vref is again the same reference signal (we can have only one reference signal) and the JND again is the noise between them.

What then is the perceived velocity of one car relative to the other, which is physically 100 - 80 = 20 km/h? The answer of course is the difference between their perceived velocities on the road. But if we subtract formula (1) and (2) from each other, the reference signal drops out of the equation. Hence the relative motion between objects corresponds to the difference between $Vret_A$, and $Vret_B$ minus a noise factor.

This implies that the movement of objects not relative to space but relative to other objects is still perceived on the basis of the interaction between two signals. We get the same graph as in figure 1, but now the two axes reflect the velocities of the two retinal images. Hence the symmetrical relativity of such motion percepts remains as it is. However, now the two relevant signals stem from only retinal image motion, i.e. from externally induced excitation of our sensory system. The subjectively induced interpretative reference signal is absent. There only remains a certain noise level that must be taken into account.

The present model thus has implications for the nature of perception: The absolute dimension of object motion, i.e. motion relative to space, i.e. relative to the three-dimensional frame of reference defined by the earth's surface and the direction of gravity, can only be perceived 'indirectly'. To generate such percepts, the physical excitation of our sensory systems must always interact with a subjective self-generated interpretative signal. But motion of objects relative to each other can be perceived 'directly', i.e. purely in terms of external physical excitation of our sensory systems without the need for subjectively generated interpretative signals. Hence, *relations* can be perceived 'directly'. In other words, the famous Kantian argument is correct that a tree cannot be perceived veridically on the basis of only physically incoming information. Such a percept is the outcome of an interaction between externally induced sensations and internally generated interpretational information. Thus, it should not come as a surprise that in figure 1, which illustrates how we perceive absolute motion, physical parameters of the real movements of the object are completely absent. On the other hand, percepts concerning *relations* can be perceived from external information only. If we want to perceive whether one tree is higher than another, the subjective element, which is present in the percept of each tree separately, drops out when they are subtracted from each other. Nevertheless since such percepts stem from the interaction between two signals (their difference), these percepts still reflect the relativity of motion.

What does this mean with respect to the veridicality of our percepts? Nothing! Although relational percepts and absolute percepts differ in terms of whether or not they derive from physical information only or from the interaction between physical and subjective information, they do not necessarily differ in terms of their veridicality with respect to the real world. If the magnitude of the signals is correct, i.e. when movements of retinal images and movements of the eyes in space are correctly encoded in the respective signals, no illusions will happen and the percepts will be veridical, apart from the fact that neural noise still creates a bias. Illusions occur when retinal or reference signals become less correct (see Wertheim, 1994 for many examples not discussed here). But even if the signals do indeed encode information correctly, and even if a percept concerns relational motion between objects, strong illusions may still occur because of the presence of neural noise. One example is what is known as the freezing illusion (Mesland & Wertheim, 1996). If a monitor with a slowly moving grating is itself moved, the motion of the grating on the monitor is seen to stop. This 'freezing' effect can be explained as caused by the noise between the retinal signal of the grating and that of the monitor. As long as the difference between these two retinal signals is smaller than the noise (i.e., the JND) between them, the brain assumes that they are equally sized, and we see no relational motion between the grating and monitor (see Wertheim & Reymond, 2007, for details).

Conclusion

The model describing the psychological mechanism which causes us to see motion can be expressed in terms that are compatible with those used in physics to define the concept of motion, and it provides insight into the empirical, the psychological and the philosophical nature of perception.

References

Buechele, W., Degner, D., & Brandt, T. (1980). Thresholds for object motion perception raised by concurrent head movements. *Pfluegers Archives Supplement* 348: R33.

Durgin, F. H. Gigone, K., & Scott, R. (2005). The perception of visual speed while moving. *Journal of Experimental Psychology, Human Perception and Performance, 31(2)*, 339-353.

Filliard, N., Reymond, G., Wertheim, A. H., & Berthoz, A. (2008). *Vestibular stimulation causes underestimation of object velocity.* Poster presented at the European Conference of Visual Perception, Utrecht, the Netherlands.

Gibson, J. J. (1979). *The ecological approach to visual perception*, Hillsdale, NJ: Lawrence Erlbaum.

Helmholtz, H. (1910). *Handbuch der physiologichen Optik, Vol. 3*, Hamburg/Leipzig: Voss.

MacKay, D. M. (1973). Visual stability and voluntary eye movements. ln: Jung, R. (Ed.) *Handbook of sensory physiology, vol. 7(3A)*, (pp. 307-331). Berlin-Heidelberg-New York Springer.

Mesland, B. S., & Wertheim, A. H. (1996). A puzzling percept of stimulus stabilization. *Vision Research, 36(20)*, 3325-3328.

Mittelstaedt, H. (1990). Basic solutions to the problem of head-centred visual localization. In R. Warren and A. H. Wertheim (Eds.): *Perception and control of self-motion.*:, (pp. 267-287). Hillsdale, NJ: Lawrence Erlbaum.

Pavard, B., & Berthoz, A. (1977). Linear acceleration modifies the perceived velocity of a moving visual scene. *Perception, 6*, 529-540.

Pelah, A., & Thurell, A. E. I. (2001). Reduction of perceived visual speed during locomotion: evidence for quadrupedal perceptual pathways in human? *Journal of Vision, 1, 307a*; http://www.Journalofvision.org/1/3/207.

Probst, T., Brandt, T., & Degner, D. (1986). Object-motion detection affected by concurrent self-motion perception: Psychophysics of a new phenomenon. *Behavioral Brain Research, 22*, 1-11.

Swanston, M. T., & Wade, N. J. (1988). The perception of visual motion during movements of the eyes and of the head. *Perception and Psychophysics, 43*, 559-566.

Thurell, A. E. I., & Pelah, A. (2002). Reduction of perceived visual speed during walking: effect dependent upon stimulus similarity to the visual consequences of locomotion. *Journal of Vision, 2*, 268a; http://www.Journalofvision.org/2/7/628.

Von Holst, E. (1954). Relations between the central nervous system and peripheral organs. *British Journal of Animal Behaviour, 2*, 89-94.

Von Holst, E., & Mittelstaedt, H. (1950). Das Reafferentzprinzip (Wechselwirkungen zwischen Zentralnervensystem und Peripherie). *Naturwissenschaften, 37*: 464-476.

Wallach, H. (1987). Perceiving a stable environment when one moves. *Annual Review of Psychology 38*, 1-27.

Wallach, H., & Kravitz, J. H. (1965). The measurement of the constancy of visual direction and of its adaptation. *Psychonomic Science, 2*: 25-29.

Wertheim, A. H. (1981). On the relativity of perceived motion. *Acta Psychologica, 48*, 97-110.

Wertheim, A. H. (1994). Motion perception during self-motion. The direct versus inferential controversy revisited. *Behavioral and Brain Sciences, 17*, 293-355.

Wertheim, A. H., & Reymond, G. (2007). Neural noise distorts perceived motion: The special case of the Freezing illusion and the Pavard and Berthoz effect. *Experimental Brain Research 180, 3*, 569-576.

Wertheim, A. H., Van Gelder, P., Lautin, A., Peselow, E., & Cohen, N. (1985). High thresholds for movement perception in schizophrenia may indicate abnormal extraneous noise levels of central vestibular activity. *Biological Psychiatry*, *20*: 1197-1210.