BEHAVIORAL AND BRAIN SCIENCES (1999) **22,** 337–340
Printed in the United States of America

Continuing Commentary

Commentary on **Alexander H. Wertheim (1994). Motion perception during self-motion: The direct versus inferential controversy revisited. BBS 17:293–355.**

Abstract of the original article: According to the traditional inferential theory of perception, percepts of object motion or stationarity stem from an evaluation of afferent retinal signals (which encode image motion) with the help of extraretinal signals (which encode eye movements). According to direct perception theory, on the other hand, the percepts derive from retinally conveyed information only. Neither view is compatible with a perceptual phenomenon that occurs during visually induced sensations of ego motion (vection). A modified version of inferential theory yields a model in which the concept of extraretinal signals is replaced by that of reference signals, which do not encode how the eyes move in their orbits but how they move in space. Hence reference signals are produced not only during eye movements but also during ego motion (i.e., in response to vestibular stimulation and to retinal image flow, which may induce vection). The present theory describes the interface between self-motion and object-motion percepts. An experimental paradigm that allows quantitative measurement of the magnitude and gain of reference signals and the size of the just noticeable difference (JND) between retinal and reference signals reveals that the distinction between direct and inferential theories largely depends on (1) a mistaken belief that perceptual veridicality is evidence that extraretinal information is not involved, and (2) a failure to distinguish between (the perception of) absolute object motion in space and relative motion of objects with respect to each other. The model corrects these errors, and provides a new, unified framework for interpreting many phenomena in the field of motion perception.

Are sense-specific reference frames so mutually exclusive?

Xavier M. Sauvan
Laboratory for Integrated Advanced Robotics, University of Genova, 16145 Genova, Italy. **sauvan@lira.dist.unige.it** **argo.lira.dist.unige.it/mseloy**

Abstract: We recently found orientation constancy with respect to direction of gravity in the alert monkey. This seems to rely on a polysensory interaction that involves different sense-specific reference frames. Thus, we will challenge the assumption that the sense-specific reference frames are mutually exclusive. At the same time, we will highlight the dynamic tuning of the receptors that might rely on cross-modal mechanisms.

Stoffregen (1994) claimed that Sauvan and Bonnet (1993) and Wertheim (1994) share the belief that the perception of self-motion is sense specific, especially visually specific. Moreover, Stoffregen stated that this would imply that the sense-specific reference frames are mutually exclusive. In particular, visual and vestibular signals would not compromise each other.

We shall argue that there is a cross-modal integration of visual and vestibular information related to gravity perception, in particular from our recent results. This is related to Wertheim's (1994) approach, which takes into account the fact that Newtonian and Einsteinian dimensions cannot be distinguishable for any observer on earth. We shall also emphasize that polysensory integration might show that sense-specific reference frames are not mutually exclusive.

Sauvan and Bonnet (1993; 1995) have already indicated that integration of visual (optical flow) and vestibular signals is involved in the perception of self-motion, especially linear self-motion. If our results (Sauvan 1998a; Sauvan & Bonnet 1993; 1995) highlight the role of the spatio-temporal structure of the optical flow in evaluating and controlling self-motion, our theoretical approach takes explicitly into account the cross-modal processing that is involved in the perception of self-motion.

Recently, Sauvan and Peterhans (1995; 1997; 1999) found that the perception of orientation of contours during body tilt relies on a polysensory integration that occurs as early as the occipital cortex, if not earlier, in the thalamus. They found neurons in the visual cortex of the awake behaving monkey that showed orientation

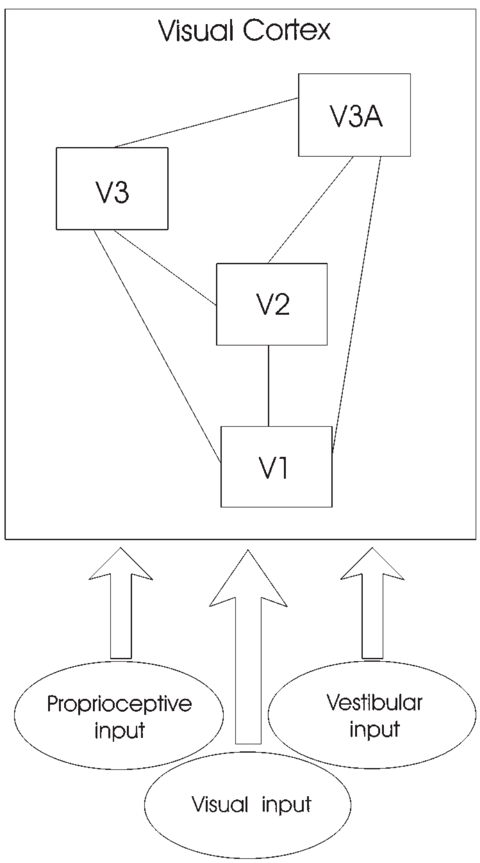

Figure 1 (Sauvan). The streams of retinal and extra-retinal information to the visual (occipital) cortex that might be involved in orientation constancy with respect to the direction of gravity during body tilt are shown. Also the visual areas in the occipital cortex and their connections (see Van Essen et al. 1992), which participate in this early cross-modal integration, are described.

https://doi.org/10.1017/S0140525X99221838 Published online by Cambridge University Press

constancy with respect to the direction of gravity. These neurons, called compensatory cells, showed no change in the preferred stimulus orientation irrespective of the body tilt. Similar results were found in a few studies made in paralyzed cats (e.g., Tomko et al. 1981). These compensatory cells were found mainly in the prestriate cortex (areas V2, V3, and V3A). This cross-modal integration might involve visual, vestibular, and/or proprioceptive signals (Sauvan 1998b; Sauvan & Peterhans 1995; 1997; 1999). Thus, our results suggest that mechanisms of orientation selectivity in the visual cortex integrate signals of nonvisual modalities which indicate direction of gravity, and that this convergence of modalities occurs at an early stage of cortical processing (Fig. 1).

This integration of visual and extra-retinal information about the direction of the gravity coming from the otolith organs, the neck muscles, the joints, and the gravity receptors of the trunk is a clear example of crosstalk between sense-specific reference frames. Moreover, our results suggest that the dynamics of receptors plays a role in perceiving orientation according to Wertheim's approach (1994) because extra-retinal cues can dynamically calibrate the tuning of visual receptors. Other findings showed an early polysensory integration of retinal and extra-retinal signals. For example, there is a modulation of disparity by vergence in area V1 that corresponds to the integration of visual and proprioceptive information (Trotter et al. 1996).

Our results do not imply that the dynamic properties of the body in interaction with the environment are not playing a role in perceiving orientation. We are making the assumption that the dynamics of the receptors corresponds to the dynamic properties of the body in the sense that changes of body orientation in space involve both a dynamic tuning of the receptors and adaptive kinematics. Again, the dynamic tuning of the receptors might rely on cross-modal mechanisms.

In conclusion, we think that a way to get a better understanding of the crosstalk between the sense-specific frames of reference is to study more systematically the central integrations between the information streams coming from the different kinds of receptors by using both psychophysical and physiological methods. This is especially true for the information streams involved in the perception of stimulus orientation with respect to the direction of gravity during a change of body orientation.

ACKNOWLEDGMENTS

This work was supported under the European Union VIRGO research network of the Training and Mobility of Researchers program (FMRX-CT96–0049).

Author's Response

Motion percepts: "Sense specific," "kinematic," or …?

A. H. Wertheim

TNO Human Factors Research Institute, 3769 DE, Soesterberg, The Netherlands. **wertheim@tm.tno.nl**

Abstract: In line with my model of object motion perception (Wertheim 1994) and in contradistinction to what Stoffregen (1994) states, Sauvan's data suggest that percepts of motion are not sense specific. It is here argued that percepts of object- or self-motion are neither sense specific nor do they necessarily stem from what Stoffregen calls "kinematic events." Stoffregen's error is in believing that we can only perceive object- or self-motion relative to other objects, which implies a failure to realise that percepts of absolute object- or self-motion in space (relative to the earth's surface) do exist as well, even when the earth's surface itself is not perceived.

Sauvan presents data which argue against Stoffregen's commentary (Stoffregen 1994) on my target article (Wertheim 1994t), in which I presented a model describing how visual percepts of object motion relative to external space (and relative to other objects) are attained by an observer who also moves in external space. In his commentary, Stoffregen criticized the model because it presumably describes percepts as "sense-specific." I agree with Sauvan's interpretation of his own data as challenging that criticism. However, I would like to argue here that in addition to his arguments there are also several theoretical reasons to reject Stoffregen's criticism.

To make this clear, it should first be pointed out that Stoffregen's term "sense-specific" can mean different things. At first glance we may assume that a sense specific percept is one that is derived exclusively from afferent sensory information, specific to a single sensory system, which in the present case is the visual system. But if this is what Stoffregen meant when he claimed that the model presented in my target article assumes that percepts of object motion relative to external space are sense specific, he has not understood its essential features. The model describes the two basic neural input signals used by the mechanism that generates percepts of object motion relative to external space. The two signals are: the retinal signal, which encodes retinal image velocity, and the reference signal, which encodes the velocity of the retinal surface in space. Neither of these signals should be understood as the neural representation of a percept. Instead, the model states that the percept of object motion relative to external space is born from the interaction between these two neural signals. This interaction was modelled by describing the two signals as vectors, each with its own magnitude, which are subtracted from each other vectorially, yielding a difference vector. The model then states that the magnitude of this difference vector minus a noise factor determines the percept of object velocity relative to external space. In terms of frames of reference, the reference signal thus functions as the algorithm with which the visual system recalibrates retinal image velocity from a retinal coordinate system into the coordinates of external "Newtonian" space.

Now let us examine Stoffregen's criticism more closely:

Are the percepts thus formed sense specific as defined above? I assumed that the retinal signal stems from the retinal receptor surface, and thus that it is a sensory afferent, specific indeed to the visual-retinal system. It is less clear, however, whence the reference signal stems. A large part of my target article was devoted to that issue. Basically, I proposed that the reference signal is really a compound one, composed of two components.

One component is the efference copy, which encodes how the eyes move in their orbits. The other component is a signal that encodes how the head moves in space. The efference copy and the head-in-space component add together to form the reference signal, which thus encodes the velocity of the retinal surface in space.

How sense specific is this reference signal then? First consider its efference copy component. This is presumably a neural corollary, a "copy," of an efferent oculomotor command signal, which means that it certainly cannot be viewed as derived from sensory afferents. As to the head-in-space component, I have proposed that it stems from sensory reactivity of the equilibrium system, from (mainly neck) muscle feedback (together termed vestibular afferents), and from retinal afferents generated by optic flow (termed optokinetic afferents). Hence, the head-in-space component stems from sensory afferents indeed (if feedback signals from muscles may also be called "sensory"), but not from just a single one. It is multisensory rather than sense specific.

In short, the compound reference signal not only derives from sensory afferents but also includes neural corollaries to (oculo)motor commands. Hence it can be viewed neither as a sensory afferent nor as specific to any sensory system.

It thus follows that Stoffregen's criticism – that percepts of object motion relative to external space, as described by my model, are "sense specific," in the sense that they derive from sensory afferents of a single sensory system (the visual one) – is certainly incorrect.

However, this may well be an erroneous interpretation of Stoffregen's criticism, because in his commentary he appears to say that in terms of my model it is the perception of self motion which is sense specific. But if so, that would also be incorrect. The model concerns percepts of object motion during self motion but makes no claims about the nature of percepts of self motion. It only assumes that the head-in-space component in the reference signal stems from the same vestibular and optokinetic afferents that also serve as input to the mechanism that generates percepts of self motion. How the latter mechanism uses these inputs to generate those percepts of self motion is a matter not discussed in my target article.

This may still not correspond exactly to what Stoffregen really wanted to say. What I believe he meant was that the perception of self motion does not stem from just neural afferents, but from what he calls "kinematic events," that is, from interactions between the observer and the environment, or stated differently: from sensory afferents caused by and in interaction with concurrent (intended) motor activity.

However, this now appears quite a close description of how percepts of object motion stem from the interaction between visual (retinal) afferents and reference signals, which include efference copies coinciding with motor acts (eye movements). Thus, if anything, such percepts of object motion are not sense specific, but result from exactly the kind of kinematic interaction Stoffregen proposes. As mentioned earlier, however, if Stoffregen's criticism refers only to self motion perception, it is irrelevant to my model as such.

Nevertheless I do have two objections to the idea that percepts of self motion stem exclusively from such "kinematic events." First, I disagree with Stoffregen's argument that percepts of self motion must always stem from sensory afferents caused as a consequence of (intended) motor acts. My problem is with the implied exclusion of sensory afferents that are not caused by activity of the observer but are imposed upon the observer, irrespective of how he acts. It is well known that percepts of passive self motion may derive from sensory afferents (e.g., visual afferents generated from optic flow) that have no causal relation to the behavior of the observer at all but stem from independent external events (as when a percept of self motion develops when we move the striped wall of an optokinetic drum around a subject). In such cases there simply is no "kinematic event."

My second objection is more serious: Stoffregen's "kinematic event" hypothesis implies that relative motion between the environment and the observer is the only cue (or information) for perceiving self motion. In other words, we would only be able to perceive relative motion between our own body and our immediate environment. This means that when we do not see external space (the earth's surface), we cannot perceive whether or not we are moving relative to it. Apparently this is the crux of what Stoffregen believes: in such circumstances we can have no such percepts of "absolute" self motion. He tries to bolster this view by arguing that external "Newtonian" space cannot be perceived anyway because it does not exist. Observers do not need to know how they move relative to something that does not exist.

However, anyone who has ever experienced ego motion in an optokinetic drum (vection) knows that this is a percept of "absolute" self motion, relative to the earth's surface, even though we do not perceive that surface. The same can be said about the percept of "absolute" ego motion which is generated when we look through a train window at a neighbouring train moving out of the station. Here we definitely believe that we move relative to the earth's surface, although we do not see that surface. The issue of whether or not such a visually induced percept of self motion relative to external space is illusory is irrelevant here. The point is that these percepts do really exist. This is not to say that other percepts of self motion cannot exist as well. For example, when looking at the other train, or during vection in the optokinetic drum, observers are also capable of perceiving that there is relative motion between themselves and their environment (the train or the drum wall). But that is a symmetrical percept: the environment and the observer move relative to each other. This is the only kind of percept Stoffregen allows. On the other hand, "absolute" percepts of motion are often assymetrical: we move in space, the other train does not. And those percepts also exist.

What Stoffregen does not seem to grasp is that percepts of "relative" motion between oneself and one's visible environment and percepts of "absolute" self motion relative to invisible "Newtonian" space (relative to the earth's surface) are really two different percepts. To understand this we must realize that a percept is a cognition resembling an answer to a question. For example, when talking about objects, we can say that we perceive the color of an object. We can also say that we perceive the form of an object. These are different percepts, because they answer different questions (What is the colour of the object? and What is the form of the object?). Similarly, we can ask questions about relations between objects. To perceive the velocity between an object and one's own body is to answer the question: How fast do the object and I move towards or away from each other? To perceive the velocity of an object relative to external space, we answer quite another question: How fast does the object move relative to the earth's surface? These are different percepts although they both refer to a relation, just as percepts of colour and form are different percepts although they both refer to an object. The problem is with the word "motion," which is too general. It denotes a relation, but does not specify what relation. In talking about perceiving motion we should always stipulate what motion we mean, that is, what question the observer is answering. In section R6 of my Response (Wertheim 1994t) to the commentaries on my target article, I tried to make this clear.

In short, if Stoffregen's claim that percepts of self motion are not sense specific implies the assumption that they can only stem from "kinematic events," then his criticism must be rejected. It leads to the absurd consequence that (visually induced) percepts of self motion cannot be relative to the earth's surface when that surface is not visible. In reality, visually induced percepts of "absolute" self motion and of "relative" self motion are both possible and may well arise in similar circumstances. But they are nonetheless different percepts. The interesting question is not whether or not such percepts should exist, but how they come about.

Note that this rejection of Stoffregen's claim that percepts stem only from "kinematic events" does not imply the acceptance of what Stoffregen apparently sees as the only (wrong) alternative: that percepts of (self) motion are sense specific. I hope to have made it clear that, in terms of my model, percepts of object motion are neither sense specific nor caused purely by "kinematic events," and that it is my belief that this also holds for percepts of self motion.

The cells reported by Sauvan are quite intriguing, but they may or may not relate to my model, as their relation to motion (i.e., "absolute" motion) has not been studied. So far, Sauvan has only shown them to relate to "absolute" orientation, that is, to orientation relative to the direction of gravity. But even if they turned out not to be related to percepts of motion, but only to percepts of orientation, it still remains to be seen whether they form part of a mechanism responsible for the perception of "absolute" self orientation, or for percepts of "absolute" object orientation.

References

Sauvan, X. M. (1998a) Linear vection as a tool to study the visual control of self-motion. In: *Vision in Vehicles,* Vol. 6, ed. A. G. Gale, I. D. Brown, C. M. Haslegrave & S. P. Taylor. Elsevier. [XMS]

Sauvan, X. M. (1998b) Early integration of retinal and extra-retinal information: Recent results and hypotheses. *Reviews in the Neurosciences* 9:291–99.

Sauvan, X. M. & Bonnet, C. (1993) Properties of curvilinear vection. *Perception and Psychophysics* 53:429–35. [XMS]

Sauvan, X. M. & Bonnet, C. (1995) Spatiotemporal boundaries of linear vection. *Perception and Psychophysics* 57:898–904. [XMS]

Sauvan, X. M. & Peterhans, E. (1995) Neural integration of visual information and direction of gravity in prestriate cortex of the alert monkey. In: *Multisensory Control of Posture,* ed. T. Mergner & F. Hlavacka. Plenum Press. [XMS]

Sauvan, X. M. & Peterhans, E. (1997) Orientation constancy in visual cortex of the awake monkey. *Investigate Ophthalmology and Visual Science* 38:S625. [XMS]

Sauvan, X. M. & Peterhans, E. (1999) Orientation constancy in neurons of monkey visual cortex. *Visual Cognition* 6:43–54. [XMS]

Stoffregen, T. A. (1994) "Sensory" reference frames and the information for self-motion versus object-motion. *Behavioral and Brain Sciences* 17(2):332–33. [XMS, rAHW]

Tomko, D., Barbaro, N. & Ali, F. N. (1981) Effect of body tilt on receptive field orientation of simple visual cortical neurons in unanesthetized cats. *Experimental Brain Research* 43:309–14. [XMS]

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S. & Imbert, M. (1996) Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1. *Journal of Neurophysiology* 76:2872–85. [XMS]

Van Essen, D. C., Anderson, C. H. & Felleman, D. J. (1992) Information processing in the primate visual system: An integrated system perspective. *Science* 255:419–23. [XMS]

Wertheim, A. (1994) Motion perception during self-motion: The direct versus inferential controversy revisited. *Behavioral and Brain Sciences* 17(2):293–355. [XMS, rAHW]