quism phenomenon that occurs in spatial scene analysis. When light flashes and sound trains come from moderately separated locations, the spatial separation is generally unnoticed, and perceptual recalibration is also manifested in aftereffects (Radeau & Bertelson 1974). The criteria for pairing visual and auditory signals from different locations are sensory factors like the timing of the signals (although strict synchrony is not required) and the distance between them. As demonstrated in barn owls raised from birth to adulthood with prisms (Knudsen & Knudsen 1989) and in human adults (Colin et al. 2001), ventriloquism decreases with increasing spatial separation, being maximal until 20°. Cognitive factors do not play any role. A context simulating a real-life situation, such as seeing the face of a speaker or the hands of a man playing bongos while hearing the sounds displaced, does not enhance ventriloquism beyond the level reached in more artificial situations, as when diffuse light is modulated by the sounds (Radeau & Bertelson 1977) or when the speaker's face is presented inverted (Colin et al. 2001). The system underlying ventriloquism has been considered as being based on primal knowledge of the Gestalt principles of common fate and proximity (Radeau 1994a), used both in visual grouping and in "auditory scene analysis" (Bregman 1990).

Contrary to the ventriloquism effect, which concerns localization, the McGurk effect concerns speech identification and is subtended by different spatial and cognitive rules. It is unaffected by the degree of spatial separation between the signals (Colin et al. 2001), but it decreases in cases of face-voice gender discrepancy for familiar speakers (Walker et al. 1995) or of face inversion (Bertelson et al. 1994; Colin et al. 2001; Jordan & Bevan 1997; Massaro & Chen 1996).

The two effects are probably achieved by specific mechanisms in a way consistent with their different functions (Radeau 1994b). Neurophysiological studies of vision in nonhuman primates have provided evidence for the "what" and "where" problems involving distinct neural pathways (Ungerleider & Mishkin 1982). Recent neuropsychological data from human patients with left hemisphere lesions argued for a "what" versus "where" distinction in the auditory modality as well (Poremba et al. 2003).

The discovery of multimodal neurons helps in understanding crossmodal responses because sensory convergence on individual neurons may well be the underlying neural mechanism. Multisensory neurons have been found in many species and in many parts of the brain. Especially relevant here are the audiovisual neurons found in the deep layers of the superior colliculus and in the polysensory cortex of cat and monkey (Stein & Meredith 1993). Although these neurons often fail to respond to unimodal stimulation, they exhibit vigorous responses under bimodal stimulation provided the stimulations come from locations not too far apart. Enhancement is eliminated around 20° of spatial disparity, and it is inversely related to temporal disparity without being restricted to temporal coincidence. The rules that govern responses of audiovisual neurons are therefore very similar to those that underlie ventriloquism, so these neurons could well constitute the neural substrate of this phenomenon.

What about the development of multisensory functioning? Does it result from amodal representations that are functional early in life or is it learned from experience of co-occurrent unimodal informations, as assumed by empiricist philosophy?

Probably due to the immaturity of the superior colliculus of the newborn cat, there is no evidence for multisensory enhancement before several weeks after birth (Stein et al. 2000). However, there is much behavioral evidence to indicate that there is a primitive unity of the senses, the sensory systems becoming gradually differentiated during development (Bower 1974; Gibson 1966; Marks 1978). In the first months after birth, neonatal humans (Lewkowicz & Turkewitz 1980) and rats (Spear & McKinzie 1994) respond to stimulation in all modalities; further, these responses are dominated by quantitative aspects of the stimulation without distinction of modality. On the other hand, synesthesia (joined sensation) is very important in the first month of life and decreases during development, being two to three times more frequent in infants than in adults (Marks 1975; Maurer 1993).

Data from studies on perinatal sensory surstimulation or sensory deprivation also provide support for early auditory-visual connections. Unusually early experience in a late-developing system interferes with sensory functioning in earlier-developing systems. Exposure of bird embryos to visual stimulation several days prior to hatching results in an auditory deficit, with ducklings (Gottlieb et al. 1989) and quail chicks (Lickliter & Banker 1994) failing to learn the maternal call.

Moreover, perinatal deprivation in a sensory system can affect functioning in the remaining modalities. Deprivation of patterned visual stimulation by binocular eyelid suture in ferrets (King & Carlile 1993) and barn owls (Knudsen et al. 1991) results in anomalous responses of auditory neurons.

Visual event-related potentials (ERPs) have been recorded in congenitally deaf cats (Rebillard et al. 1980) and humans (Neville 1990) over temporal brain areas, which in the hearing subject contain the auditory cortex. However, there was no change in humans who became deaf after the age of four years. Moreover, in congenitally blind humans, auditory and somatosensory ERPs have been found to have a more posterior distribution than in control subjects (Kujala et al. 1992; 1995). The observed compensatory changes can thus reflect stabilization of transitory connections in one modality (Changeux & Dehaene 1989; Edelman 1987) in the absence of competing input from another modality.

There is some neuroanatomical evidence for transient auditory to visual cortex connections around birth that disappear in the fourth week of age in the kitten (Innocenti & Clarke 1984) and in the ferret (Kennedy & Dehay 1993). Connections have also been found between the retinas and the somatosensory and auditory nuclei of the thalamus in the hamster less than 1-week old (Frost 1990). In primate newborns, auditory ERPs have been recorded over the occipital visual cortex of human 6-month-old babies but not in older children (Neville 1995).

All of these data argue for initial sensitivity to structures in the global array, experience probably leading to sensitivity to structures in single-energy arrays.

ACKNOWLEDGMENTS
This work has been supported by grants A.R.C. (96/01–203) and F.R.F.C. (8.4519.96) to Monique Radeau.

Retinae don't see

John T. Sanders
Department of Philosophy, Rochester Institute of Technology, Rochester, New York 14623. **jtsgsh@rit.edu**

Abstract: Sensation should indeed be understood globally: some infant behaviors do not make sense on the model of separate senses; neonates of all species lack time to learn about the world by triangulating among different senses. Considerations of natural selection favor a global understanding; and the global interpretation is not as opposed to traditional work on sensation as might seem.

As Stoffregen & Bardy (S&B) (2001) contend, the theory of "direct" perception does indeed seem to be incompatible with the idea that sensory data should be understood as being gathered independently via several senses. In suggesting that the conflict be resolved by rejecting the latter idea, rather than simply by abandoning the theory of direct perception, they choose the more controversial route. In this brief response I offer a few further reasons to take their suggestion seriously.

In the first place, some infant behaviors simply do not make sense on the assumption that at birth the several senses begin providing independent information which can be brought together only after further experience and comparison. Excellent examples of such behavioral phenomena are provided in the studies of infant imitation done by Meltzoff and Moore (1977; 1983a; 1983b) in which they found that infants can contort their faces in imitation of another person's facial expressions within minutes of birth. Since these children have no idea how their faces look, visual cues thus seem to be leading to direct and immediate matches with tactile/proprioceptive sensations, with no time and no mechanism available for learning about correlations among distinctly apprehended sensory modalities.

Second, consideration of experiments like these suggests, more generally, that the early lives of neonates of all species must be rich in similar cross-modal sensory integration, given the sophistication of what they are able to do almost immediately. Although there is certainly no reason to argue that the world *never* presents itself to animals via singular sensory modalities, such presentations must be rare. Yet animals need to work with information in their environments almost immediately, and it is often the nuanced information that crosses and combines modalities that is most crucially needed, whether for balance, for reaching, for flight, or whatever. Animals simply do not have time to learn about the world by triangulating among independent contributions of several different senses.

Third, the idea that sensation is primitively multimodular makes considerable evolutionary sense. It seems likely that the several sense organs of each species have evolved as specializations of earlier less-specialized organs, based on the proven value of each specialization in enabling species members to survive and propagate. More general sensation must arrive on the evolutionary scene earliest, and then becomes more specific as a result of the contingencies of the niche. This seems consistent with a treatment of sensation that understands it most primitively as a global sensitivity to the environment, focused by opportunities and dangers available there, organized and differentiated by natural selection over time in terms of various sensory surfaces.

Finally, students of perception have frequently disagreed about how to understand the role of the several sensory modalities. It has seemed plain to anyone who has ever thought about sensation that colors are remarkably different from sounds, and although this has seemed to require a sharp distinction among the contributions of the several senses, there have also been suggestions that this distinction must be moderated in any suitable analysis.

George Berkeley (1709; 1713; 1733; Jessop 1937), for example, tried to distinguish between "immediate and proper" seeing and a more liberal sense of "seeing." The first – the bare immediate and proper "seeing" – is to be understood as uninterpreted, stripped of all learned associations. But Berkeley's own texts show that he was himself very uncomfortable with the traditionally conceived implications of this distinction.

This discomfort emerges quite clearly in Berkeley's examination of whether distance can be seen. His view was that although it cannot be seen "of itself and immediately," it can nevertheless be seen in a less restrictive sense. Indeed, according to Berkeley, there are a great many factors involved in determining our visual perception of distance, one of which is plainly the apparent magnitude of the thing seen. But this did not lead him to conclude that distance is seen indirectly via such cues as size, since that latter perception is often based on how far away we think the object is. Neither cue is less "directly" perceived than the other. Each can help in discerning the other. Another factor involved in determining *both* perception of magnitude and perception of distance is the apparent faintness of what is seen. But that is simply another factor in a very complicated contextual situation.

Berkeley argued, finally, that what one sees with regard to distance and magnitude is determined in part also by the posture of the head and eyes, and perhaps with the help of contributions from other sensory modalities. What is important here, though, is his forceful argument that no "judgment" or inference is involved in such seeing – the distance is suggested immediately. In sum, Berkeley's claim is not really that we do not see depth, for he explicitly says that we do. But he insisted equally that it is *we* who see . . . not our retinae (see Sanders, forthcoming).

This line of thinking shows, in any case, that even Berkeley's thoughts on the subject examined by S&B were not as antithetical to their thesis as might be imagined. There is also support here for the idea that a theory of "direct" perception need not be as counterintuitive as has sometimes been maintained. The line of study urged by S&B would amount to a Gestalt switch of sorts, it is true. And it is important to acknowledge, even in their proposed research program, the importance of studying the separate sensory modalities in order to further understand their contribution to sensation. The upshot would be, though, that sensation would be understood not as taking place at sensory surfaces, in particular, but throughout a larger global sensory system which has those surfaces as parts.

Multi-sensory processing facilitates perception but direct perception of global invariants remains unproven

Lawrence Warwick-Evans

Department of Psychology, The University of Southampton, Southampton, SO17 1BJ, United Kingdom. **lawe@soton.ac.uk**
http://www.soton.ac.uk/~psyweb/staffpages/lawe/lawe.html

Abstract: The existence of sensory convergence does not establish that the senses function as a single unified perceptual system. Reality is fully specified only by a one:many mapping onto the totality of energy arrays, and these provide alternative frames of reference for movement. It is therefore possible that higher order crossmodal relationships are detected by skilled perceivers, but this has not been confirmed empirically.

After the initial agenda setting in the target article (sect. 1), Stoffregen & Bardy (S&B) (2001) consider in section 2 the three arguments (anatomy, energy, and neurophysiology) that are used to support the idea of the separate existence and functioning of each sense. They correctly reject the anatomy argument by pointing out that inputs are required from the two (anatomically separate) ears in order to localise sound; this establishes the point that anatomically distinct origins are not sufficient to preclude later functional interdependence. But the energy argument is not, as they claim, circular: the electromagnetic and vibrational energies to which the eye and ear respond can be defined independently of the eye and ear. The neurophysiology argument is a variation of the anatomy argument and similarly establishes that initial separation of sensory inputs is not evidence against later convergence and unified perception. In short, the anatomy and neurophysiology arguments are correct but not new, whereas the energy argument is incorrect. The authors' arguments from example do entail the limited (weak) conclusion that sometimes two or more senses act as a single perceptual system, but do not justify the universal (strong) claim that the senses never function as separate systems and that they all and always function together in a unitary and irreducible manner.

Section 3 of the target article distinguishes four possible relationships (no specification, modal, multiple independent, and multiple amodal specification) between reality and the ambient array. The authors conclude that "each (relationship) . . . is confronted with problems" (sect. 3.4) and that "all theories of perception derived from existing views of specification are compromised by fundamental errors" (sect. 1). Their pivotal reason for rejecting all four views is that (1) each view involves the possibility of sensory conflict, and (2) since the senses act as a unified single perceptual system, there cannot be conflict between the separate senses. This argument would not be sufficient to reject the four views even if the strong version of sensory unity were correct, because there is no a priori reason why there cannot be conflict within a system. Additionally, there is a major inconsistency between rejecting the independent amodal view of specificity defined as "a one:many mapping, with properties of reality being

https://doi.org/10.1017/S0140525X04250208 Published online by Cambridge University Press