Fragmentation, coherence, and the perception/action divide

Erik Myin
Department of Philosophy and Artificial Intelligence Laboratory, Free University of Brussels (VUB), B1050 Brussels, Belgium. emyin@vub.ac.be
http://homepages.vub.ac.be/~emyin/

Abstract: I discuss Stoffregen & Bardy’s theory from the perspective of the complementary aspect of input conflict, namely, input coherence – the unity of perception. In a classical approach this leads to the famous “binding problem.” The conceptual framework the authors construct leaves no space for a binding problem to arise. A remaining problem of perceptual conflict, arising in cases of inversion of the visual field can be handled by the theory the authors propose.

Sensory conflict plays an important role in the target article. In this commentary, I will look at the paper with a complementary theme in mind: sensory coherence, or the unity of perceptual consciousness.

According to the classical view, perceptual input information is not only ambiguous, therefore requiring inference, but it is also fragmented. Fragmentation applies at a variety of levels. First, the input information is shattered to pieces as it spread out over thousands or even millions of receptor cells. At a higher level, fragmentation occurs because of the separateness of the sense: separate modules in the various modalities deal with the shattered pieces of input to build modality-specific higher-level representations. These higher level representations are supposed at least partially to solve both the ambiguity problem and the fragmentation problem. Ambiguity has been solved because knowledge not present (such as Marr’s rigidity constraint, 1982, pp. 209–210) in the stimulus itself has been brought to bear upon it. Fragmentation is partially undone because from the receptor mosaic emerge sense-specific representations, perhaps in a “canonical” format. Yet fragmentation remains a problem. Now how do all these outputs from these separate modules get glued together? The situation is aggravated by the fact that it is widely assumed that there is intermediate level fragmentation within the modules; within the visual module there are submodules for color, form, movement, and so on (Zeki 1993). So the traditional view creates for itself the famous “binding problem”. How do all these separate representations get together so as to lead to the unified perceptual consciousness we normally enjoy? The favorite solution within the tradition is well-known: it is neural synchronisation that is doing the job (Crick & Koch 1990).

Whatever the merits of the synchronisation idea, it is still unable to deal with still higher levels of fragmentation. What, for example, binds lower level sensory representations (as outputs from perceptual modules) to their conceptual representations? Even after the color of the bucket, its shape and its cold feel have been bound, it still needs to be bound to how I conceive it: as the bucket I used to scare the attacking dog on Aunt Margaret’s farm, for example. And there is yet a further level of fragmentation: how does the passively received input representation, even when fully bound with each other and with conceptual representations, become something I can act upon? How is the perception-action diad crossed? Does it need to be bound with plans, or with goal-representations? Clearly, we’re on the verge of a combinatorial explosion, or worse, conceptual impossibility (cf. Shanon 1993).

One of the many nice aspects of this paper is that it shows conclusively that these problems of fragmentation and binding, as the side coin of problems of input conflict, arise not only within the traditional view, but also within all forms of Gibsonian theories that assume separate senses. With hindsight, one sees that it could not be otherwise: fragmentation at separate senses (accepted by many Gibsonians) is just a higher-level version of fragmentation at sensory receptors (criticised by all Gibsonians). It is only with S&B’s proposal of unifying the senses ab initio, that the conceptual space closes so as not even to leave room for the problem of perceptual unity to emerge.

From the various remarks in the target article concerning the dependence on the goals of action as the functional basis of the interaction with the global array (e.g., end of sect. 4.5, also sect. 6.2.5), one sees how the theory also contains the ingredients to cross the last gap created by fragmentation: the gap between perception and action. A consequence of this is that the distinction S&B make between perception and action can only be heuristic. Just as they characterise their view as one in which perception is not seen as the parallel action of a group of systems, but as the unitary action of a system with parts, they, seem to me, would have to apply this to action and perception and see these as aspects of one global thing; the organism in its interaction with the environment (as I think they would certainly be willing to do).

More interesting, bridging the perception/action gap in this way allows for a real form of fragmentation: erroneous or disunited perception when either the organism’s perceptual situation is problematic when viewing conditions are suboptimal or the perceptual apparatus is damaged (cf. sect. 6.2.3) or when its action possibilities are suboptimal. An interesting situation occurs when both are suboptimal, as in the well known case of people wearing goggles that invert the retinal image.

It has been well documented that once behavioral adaptation to the goggles occurs, and once people learn to act in the “normal” way again, their perception returns to normal. However, perceptual adaptation does not occur in an “all at once,” but in a fragmented way. Perceptual adaptation seems to depend on which behavioral capacities have adapted, and they do not all adapt at the same pace (for discussion, see Hurley 1998, p. 347–48, and O’Regan & Noe, in press). What this suggests is that just like perceptual unity, perceptual fragmentation is not an internal affair. Rather it is a matter of behavioral “disunity.” At last, we have a real case of fragmentation, but one that perfectly fits within the framework Stoffregen & Bardy sketch.

ACKNOWLEDGMENT
Thanks to the Flemish Community and the Free University of Brussels (VUB), project GOA 5 for financial support.

Act globally, think locally

Michael F. Neelon and Rick L. Jenison
Department of Psychology, University of Wisconsin, Madison, WI 53706. fnneelon@students.wisc.edu; jenison@wavelet.psych.wisc.edu
www.wavelet.psych.wisc.edu/jenison.html

Abstract: The authors attempt to prove that single energy arrays cannot specify reality. We offer contrary evidence that motion structures the acoustic array to specify fundamental attributes of the source. Against direct detection in general, we cite evidence that humans weight acoustic inputs differentially when making perceptual judgments of auditory motion.

Stoffregen & Bardy (S&B) attempt to prove the inadequacy of perceiving reality via single energy arrays through the following syllogism: perception in general is largely the perception of motion; motion is relative and often indicated divergently among energy arrays; hence, reliable perception of the environment is not possible via single arrays. We agree that motion is vital to successful perception. But in addition, dynamics can impose sufficient constraints that allow even single sensory arrays to specify other distal properties relevant to an animal’s behavior beyond motion per se.

For example, analytical demonstrations exist showing that auditory motion can structure the dynamic acoustic array to specify fundamental attributes of the sound source, such as its position, velocity, and time-to-contact (TTC) (Jenison 1997). These higher order variables are inversely determined from the forward equations describing the physical mapping of intensity, frequency, and interaural time delay (ITD) from source to observer. Such inverse solutions are evidence that single energy arrays can specify reality provided that sufficient input dimensions within the modality exist.

We believe there is a more general inquiry into S&B’s hypoth-
Commentary/Stoffregen & Bardy: On specification and the senses

esis concerning the adequacy of individual arrays for perception. Namely, how many sources of sensory input do they take to specify a distal property? In our view, this is a simple question of algebra. Specification of a higher order property implies the existence of an inverse solution from the proximal inputs to the distal property. If there are not enough "knowns" (input) to "unknovns" (distal variable), then the property may not be identifiable. Hence, increasing the number of inputs should generally increase specification. This is the root of the Stoffregen & Bardy's argument for the necessity of multimodal ("global") information. Again, we would generalize this position by saying what is needed to specify the world is simply enough proper input information, independent of whether that information is conveyed within ("multiple specification") or across modalities ("amodal specification"). By "proper" we mean that there is a lawful physical relationship between the distal property and the input.

While this approach explicates the physical journey from environment to receptor, we also believe in a corresponding journey from receptor to perceptor. The achievement of perception is "harder" than just saying a variable has or has not been detected, as most ecological theorists seem satisfied to claim (Marr 1982). Typical in such arguments, there is no description by the authors of how higher order information is in fact directly detected by the animal (e.g., are there transducers for global variables [Fodor & Pylyshyn 1981]? Rather than appealing to direct detection, we formally question: (1) how are input sources or physical dimensions combined when detecting properties specified by multiple inputs? and (2) how may perception be achieved through a form of statistical estimation, given the stochastic nature of neural transduction and transmission?

To investigate these questions, we have simulated the estimation of higher order auditory variables (e.g., auditory TTC) using two models of input fusion. The first employs a Kalman filter which can successfully estimate higher order terms using as input the noise-corrupted observed intensity, frequency and ITD of a simulated moving sound source (Jenison 1981). In theory, one could extend the Kalman filter to integrate information from different modalities, obviating any debate over a qualitative versus quantitative difference between uni- and multimodal input. What is required are representations of the forward equations mapping distal states to physical input dimensions across modalities. These equations allow the filter to dynamically adapt the fusion of input information in order to improve its estimate of source states.

It may be the case that the Kalman algorithm does not strictly represent neurophysiological mechanisms. However, it has been proven equivalent to a form of dynamic Bayesian estimation (Meinhold & Singpurwalla 1983), a process which we believe better describes how neural systems may extract information from sensory input. As a model embodying fewer assumptions, we have also trained a recurrent neural network to estimate TTC from the same acoustic dimensions (Neelon & Jenison 2000). Performance of both models degrades as a function of reducing either the number of inputs or their signal-to-noise ratio (e.g., increase sensor noise or virtual distance of the sources). This again illustrates that perceptual estimation is likely to be a continuous (though non-linear) function of the quantity and fidelity of lower order inputs.

Ecological theorists may remain unconvinced that real observers combine sensory inputs to perceive higher order information. There is evidence, however, that human subjects differentially weight acoustic inputs when making judgments about moving sources. First, we have tested listeners in tasks requiring them to discriminate between the intrinsic frequency and TTC of two moving sound sources (Neelon & Jenison 1997, 2000). Subjects show a pattern of results similar to that of model simulations as sensory input is degraded. Further evidence is provided by Lutl & Wang (1996), who indirectly measured the weights humans assign to different acoustic cues: level, frequency and ITD when discriminating between moving sources. They correlated listener response with experimentally controlled input variation to estimate how much influence each input had on performance.

They discovered that as task and source parameters changed, so did the weighting schemes.

To conclude, we agree in principle with Stoffregen & Bardy's arguments that multiple, dynamic inputs should provide better specification of the external world than single or static inputs. However, they offer no evidence that global energy arrays are necessary for identifying even a single invariant property; on the contrary, our analyses show the sufficiency of the dynamic acoustic array to specify several attributes of moving sound sources. The degree to which the world is inversely determined by sensory inputs (i.e., perception) is a product of the lawful relationship of energy propagation between them and the statistical reliability of the neural representations of that input.

Input-driven behavior: One extreme of the multisensory perceptual continuum
Kelvin S. Oie and John J. Jeka
Program in Neuroscience and Cognitive Science and the Department of Kinesiology, University of Maryland, College Park, MD 20742.
jjke@umail.umd.edu keso3713@glue.umd.edu
www.inform.umd.edu/EdRes/Collies/HLHP/KNES/faculty/jeka
www.glue.umd.edu/~kseo3713/Welcome.html

Abstract: The propositions that the senses are separate and that the global array may be sufficient for adequate perception are questioned. There is evidence that certain tasks may be primarily "input-driven," but these are a special case along the behavioral continuum. Many tasks involve sensory information that is ambiguous, and other sources of information may be required for adequate perception.

Despite the long scientific tradition of separating perception into separate sensory systems, it has long been recognized that different sensory systems interact in ways that remain poorly understood. Our present understanding of multisensory interactions sits at the level of phenomena such as the McGurk effect, rather than with organizational principles of how senses interact.

Stoffregen & Bardy (StB) address the problem of multisensory interaction from the perspective that separate senses do not exist. Unfortunately, their arguments against separate senses are not compelling. StB reject an anatomical basis for separate senses due to interactions of structures only within a sense (two eyes/2 ears that work together), not between. There is little doubt that anatomical substructures have developed to be sensitive to particular forms of energy. With eyes closed, how well could one determine the intensity of a light source directed at the pinnae? Moreover, the fact that animals with different receptors interpret the same form of energy in different ways does not argue against separate senses, but against the unique meaning of a stimulus to a perceiver.

More problematic is the question of whether structured energy fields provide "sufficient/insufficient" information for accurate perception. The problem may be one of nomenclature. The ecological view suggests that information for behavior is specified uniquely in the ambient array, individual or global, and any non-1:1 mapping negates specification. But, why is specification an either/or concept? Why not view the specification of behaviorally relevant information along a continuum? Under certain task conditions, the stimulus array dominates the response, which one might call "input-driven" behavior or perception. Time-to-contact (τ) (Lee 1981) is a classical example of such input-driven perception. Change the parameters of the task conditions, however, and the same sensory information may now be ambiguous, requiring other processes and sources of information to be recruited for adequate perception (e.g., memory). This view is more in line with current thinking about cognition, which stresses the dynamic nature of processing inputs from multiple sources (cf. Beer 2000).

In the search for specification, StB appeal to the concept of the global array; essentially a higher-dimensional version of the ambi-