CPC, 1994, 13 (1), 60-68

COMMENTARY ON "AUDITORY-VISUAL SPATIAL INTERACTION" (M. RADEAU)

A pair of paradoxes and the perceptual pairing process

FELICE L. BEDFORD

University of Arizona

Understanding crossmodal interactions is essential for understanding perception, action, learning, and development. Radeau (1994) advances two novel, thought provoking claims about the mechanisms which handle conflicts between two modalities, vision and audition. One claim is that the Gestalt grouping principles of proximity and common fate are constraints on the auditory-visual pairing process, and the second is that the pairing process is a module, in Fodor's (1983) sense of the term. I divide this article into four parts: comments on Gestalt constraints, on modularity, on the relation between these two claims, and finally, I offer two paradoxes about perceptual learning in general.

Gestalt constraints

My own work on conflicts between vision and proprioception (Bedford, 1989, 1993a, 1993b) has uncovered rules and constraints that appear very different than Gestalt principles of organization. It is worth exploring whether there are genuine differences between crossmodal interactions that involve proprioception rather than audition, or whether the differing results can be reconciled in some other way.

Preparation of this article was supported by a grant from the National Science Foundation (BNS-8909825).
Correspondence should be sent to Felice L. Bedford, Department of Psychology, University of Arizona, Tucson, AZ 85721, U.S.A. (e-mail: bedford @ccit.arizona.edu).

Conflicts between where one localizes an object visually and where one localizes it with the arm, or proprioceptively, can be produced by looking through a prism. For instance, a 20 diopter base left wedge prism causes a nearly uniform shift of each visual location 11.3 degrees to the right of its true (and felt) location. A critical theoretical question is whether such mappings between visual space and proprioceptive space can be thought of as a list of independent visual-proprioceptive pairs of locations. That is, one possible mechanism underlying the recalibration pairing process involves a collection of individual associations between pairs of locations.

However, experiments designed to address this issue found that mappings could not be viewed in this way. A single pair is not the fundamental unit of the perceptual learning process. Instead, additional constraints are needed, one of which appears to be *linearity*. Briefly, some of the evidence for this conclusion comes from patterns of generalization to untrained locations. If trained on only a single novel visual-proprioceptive pair, e.g. 0,10, there is an equal-sized shift for all locations, rather than a decrease in the size of the shift with increasing distance from the trained location. Training at two novel pairs produces interpolation that conforms to a linear function. In addition, training with mappings that are non-linear often produces linear changes in pointing. Taken together, the data suggest that the single pair of locations is not the fundamental unit, individual pairs are not independent of each other, a mapping is not a collection of individual associations or pairs, and a linear rule plays an important role.

How does this relate to the visual-auditory constraints suggested by Radeau? Both Gestalt principles put forth as constraints, proximity and common fate, apply to a *single* visual-auditory pair. If a single visual stimulus appears in location v and an auditory stimulus in location a, then when $v \neq a$ there is a conflict. Proximity governs how close the values of v and a are, and common fate governs how close a match the onsets and offsets of v and a are when their presentations are extended in time. These constraints, which need only be applied to each pair of locations separately, would mesh well with a pairing process that consisted of a collection of associations between individual locations. Yet the data from visual-proprioceptive recalibrations suggest that the pairing process should not be viewed in this way. One possibility is that the rules on these two different crossmodal interactions are different. To my knowledge, auditory-visual analogs to the proprioceptive-visual experiments on generalization and the decomposability of a mapping have not been conducted. (Radeau and Bertelson's paradigm, e.g.,

1974, with pared down visual and auditory stimuli would be ideal to investigate the issue).

However, there is no need to postulate such a fundamental difference between different crossmodal interactions. Even though a single v-p pair is not the fundamental unit of the learning process, individual pairs are the unit of input (Bedford, 1989). The mechanism can operate on individual pairs of locations, even a single pair. "Consequently, the principles governing when the two stimuli of any one pair of stimuli will be registered as "going together" are crucial for understanding the adaptation learning process...For instance, contiguity may be a general principle whereby entities that are close to one another are viewed as going together." (Bedford, 1989, p. 245). Hence, we can view the Gestalt constraints as complementary - they restrict what values can go in a single pair, whereas the linear constraint restricts what kinds of interactions between pairs can occur.

Having determined that there is no fundamental contradiction, however, there is still another consideration. Are the Gestalt constraints really in the recalibration process - or do they occur before that process ever begins? Radeau notes several times that the visual and auditory signals must be judged to refer to the same object in order to obtain recalibration (see also Bedford, 1993a, 1993b, 1994; Welch & Warren, 1980)[1]. Proximity and common fate are both relevant to that decision - if the two signals are near each other in space and turn on and off at the same time then they are more likely to come from the same object than if they are far apart or appear randomly. But the decision about object identity may be independent of the recalibration process itself. Independence is suggested by the fact that the object identity decision can take place without producing adaptation. If the system "decides" that what it is seeing in one place and what it is hearing in a different place

1. Bedford suggests that the reason for this is that we possess a constraint that *one object cannot be in two places at the same time*, a constraint essential for getting any perceptual learning to occur. Adaptation occurs if and only if there is a seeming violation of the constraint. If the different signals from the two modalities refer to the same object, this violates the internalized constraint that an object can't be in two places at the same time, suggests there must be an internal error, and perceptual adaptation results to correct the error. If the different signals refer to two different objects, then since there is no constraint that prohibits different objects from being in different places, there is no internal error, and no need for adaptation to occur. See also Paradox 1, this article.

are really two different objects, then perceptual learning isn't necessary and does not take place. Alternatively if what you are seeing and hearing are judged to be the same object but come from the *same* place, perceptual learning isn't necessary either. That is, neither a different-object judgement nor a same-object judgement forces recalibration. The same object/different object decision can function independently of any subsequent recalibration. If proximity and common fate operate at the level of the identity decision, and the identity decision is independent of the recalibration process, then those Gestalt constraints do not tell us anything about the mechanism of recalibration. The constraints are still interesting, but they tell us about factors involved in the identity decision and not about the recalibration process itself.

The separation of crossmodal processing into two distinct mechanisms, an object unity or identity decision with principles such as common fate and proximity, and a recalibration process with principles such as linearity, may also be helpful when considering another aspect of Radeau's claims.

Relation between Gestalt constraints and modularity

The two major claims advanced about the pairing process, Gestalt principles and modularity, appear to contradict each other. Proximity and common fate were argued to be very general principles that also apply to auditory-visual pairing. Radeau presents evidence of their generality: They apply to visual organization, to auditory organization, and to auditory-visual interactions for both spatial and non-spatial judgements. Note that proximity and common fate extend even further. For instance, there is an analog to common fate in Pavlovian conditioning. In conditioning, the greater the *contingency* between the conditioned stimulus (CS) and the unconditioned stimulus (US), the more effective the training (see, e.g., Rescorla, 1972). Contingency between the CS and the US is determined in much the same way as synchrony between visual and auditory signals - whether or not the two signals appear and disappear together over time. If the US (e.g., shock) appears without the CS (e.g., tone) then conditioning is reduced, which is analogous to the situation where the presentation of the auditory signal without the concurrent visual signal reduces capture and recalibration. Thus, synchrony, contingency, and common fate may be different manifestations of the same principle. The other principle, proximity, is known in the learning literature as *contiguity*, which refers to closeness in either time or space, depending on the problem.

The vast generality of proximity and common fate principles could, in different hands, be used as an *anti*-modularity argument. Arguments of modularity are often followed by a description of the specific and unique rules of that system, for instance, syntax processing in language. The existence of the same rules that cut across different domains and problems could be used to criticize the view that there are specific self-contained highly specialized systems.

The two claims of general Gestalt laws and Modularity are not necessarily mutually exclusive, but their co-existence would require some type of reconciliation. One possibility is that Gestalt-like principles evolved separately and independently in all the different systems (homologous evolution) because they evolved in the same world to solve similar problems. Another possibility is that the principles evolved in one module, and then were duplicated or "blueprinted" for use in other modules (see Rozin, 1976). An alternative resolution involves the idea mentioned earlier, that principles of common fate and proximity apply to the object identity decision, which is separate from the recalibration process. Whereas the recalibration process may be modular, the object identity decision may not be. Even if the object decision cannot be influenced by conscious knowledge, it could still borrow freely from a set of other criteria shared by many different systems.

Modularity

Does the proposed visual-auditory module help us to understand the pairing process, or understand the architecture of the mind? It may be too early to tell, but there are a number of questions that would have to be answered.

What precisely is the scope of the module? Radeau suggests that the module may be distinguished from auditory-visual interactions involving speech, but also suggests that speech and non-speech may be part of the same underlying system. And if the module does exclude speech, is there one module for all non-speech judgements, or are there different ones for space, for form, and other judgements? What happens when a third modality is involved? In the real world, objects will not only be heard and seen, but felt at the same time. Suppose we introduce a conflict such that vision says the object is at 20 degrees, and both touch and audition say it is at 10 degrees - the very situation we can induce with a prism. Do two separate modules, one for visual-auditory interactions and one for visual-proprioceptive interactions (which can't talk to each other until the recalibration is complete because of information

encapsulation) deal with the conflicts as if they were two separate independent conflicts? Or is there one module that can handle the input from all three modalities to reach to the best decision? - in this case that vision is at fault and should be recalibrated. Knowing the exact content of a module is a problem for auditory-visual interactions, as it is for potential modules in other domains.

A second issue concerns how useful the concept of modularity is for dealing with plastic systems in particular. Consider the criterion of innateness. Radeau reviews evidence which suggests that infants are born knowing that different modalities go together in some way, but it is not clear they are born knowing precisely how. They may know that they should look when they hear something, but may not know precisely where to look. Is this evidence for innateness or not? How precisely does one apply the innateness criteria to a system whose very function it is to produce change with experience? A more profitable approach may be to identify what parts constitute the preexisting structure, and what parts require input from the environment.

Similarly for the issue of conscious penetrability (informational encapsulation), the modularity claim forces one into an all-or-none position. Radeau presents impressive evidence that conscious knowledge involving the realism of the stimuli does not affect recalibration. But should we conclude that all kinds of conscious knowledge of the stimuli can never affect recalibration? How then does one make sense of the handful of studies Radeau mentions that do seem to suggest a complicated influence, such as the finding that when observers get to look at the visual stimulus, conscious instruction about whether there is a single object now influences the outcome. Also, the short-term resolution of auditory-visual conflict, e.g., ventriloquism, does seem more susceptible to conscious knowledge; Radeau notes that conceptual factors may just be influencing subject responses, but how can we be sure that they are not also affecting perception? Here too, investigation into how consciousness can and cannot be influential may lead to a better understanding of crossmodal interactions.

The article also explores some weaker views of modularity -that modules can be learned, and/or that they can be influenced by higher-order conceptual knowledge, but need not be to work (e.g., Marr, 1982). While these are both reasonable considerations, in what sense would these processes be *modules*? The notion of a "system" was around long before Fodor or Marr. What makes Fodor's claims novel (even if wrong) is the strong package of requirements that go above and beyond what we mean by a system. If the package is wrong, system-

atically taking things away from the requirements simply puts us back to where we were before.

But it is far easier to be destructive that constructive. The application of modularity may take the empirical study of crossmodal conflicts into a new profitable direction.

Perceptual learning paradoxes

Finally, I offer two paradoxes relevant to the issues of crossmodal conflict and object identity. The first paradox is about how sensory information in general can ever produce sensory changes. Once that problem is solved, the second paradox concerns how the specific sensory information we need for recalibration doesn't seem to be present. I state them using examples of vision and touch, but they apply to audition as well.

Paradox 1

Assumption 1) The purpose of perceptual adaptation is to keep the sensory systems functioning optimally, in the face of natural problems such as growth of the body and drifts out of alignment.

Assumption 2) Consequently, in order to get adaptation to occur there must be some kind of evidence that the sensory systems are not functioning correctly. Otherwise perception would change needlessly and hurt our ability to perceive the world accurately.

Assumption 3) The information that drives adaptation seems to come through the sensory systems - such as when vision "says" one location and touch says another.

The question: How could information that comes through the sense modalities possibly tell you that your senses are wrong? Why wouldn't any new information simply be interpreted as reflecting something new about the world?

The solution: Not all information does come through the senses. We have internalized fundamental, often inviolable constraints about the behavior of objects, for instance, that an object cannot be in two places at the same time. It is the constraints together with the sensory information that allows us to conclude something about our sensory systems.

When the vision and touch signals do not agree about location and they are judged to come from the same object, we say there is a "conflict" between the two modalities that must be eliminated. What makes the situation a conflict, that demands recalibration, is that we

have a constraint that an object cannot be in two places at the same time. Otherwise, there would be no need or incentive for the perceptual systems to correct themselves and we would instead learn something new about the world - e.g., that an object *can* be in two places at once.

Paradox 2 (Held's paradox?)[2]

Assumption 1) Normally, when you are seeing and feeling an object at the same time, vision and touch agree about location. This coincidence of spatial locations from the two modalities at the same time indicates that you are interacting with one object.

Assumption 2) Many situations can make vision and touch no longer agree about location. Adaptation is the name of the process by which they are made to agree once again.

Assumption 3) In order to get adaptation, the system must first conclude that the vision and touch signals refer to one object. Otherwise there is no reason to do anything about the non-coincidence.

The question: If adaptation requires the judgement that there is only one object, but the non-coincidence of vision and touch tells you there are two objects, then how can adaptation occur?

The solution: The coincidence of spatial locations for vision and touch is not a necessary condition of object identify. There is long list of criteria relevant for the object decision, no single one of which is necessary. The final decision about whether there is one or two objects present is reached by weighing the information from all the different cues.

When vision and touch are made to disagree about location, either because of natural circumstances or experimenter intervention, the discrepant location information leads to a "vote" that there are two objects. But the remainder of the cues "vote" that there is one object, outweighing the exception. The final decision reached is that there is only one object, which sets off the correction process for one or both modalities, and enables the location judgement from the two modalities to be restored to alignment.

---

2. I suggest the name Held's paradox, because of something Richard Held said in conversation (March 1991) that made me view the situation as a paradox. I was talking about how the conclusion of one object was necessary to get adaptation. He replied that he always viewed it the opposite way - that after adaptation the coinciding locations tells you that there is one object.

REFERENCES

Bedford, F. L. (1989). Constraints on learning new mappings between perceptual dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 232-248.

Bedford, F. L. (1993a). Perceptual and cognitive spatial learning. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 517-530.

Bedford, F. L. (1993b). Perceptual Learning. In D. Medin (Ed.), *The psychology of learning and motivation* (Vol. 30, pp. 1-60). New York: Academic Press.

Bedford, F. L. (1994). Constraints on perceptual learning: Objects and dimensions. Revision submitted for publication. Manuscript available from author.

Fodor, J. A. (1983). *Modularity of mind.* Cambridge, MA: MIT press.

Marr, D. (1982). *Vision.* New York, NY: W. H. Freeman and Co.

Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 13,* 3-51.

Radeau, M., & Bertelson, P. (1974). The aftereffects of ventriloquism. *The Quarterly Journal of Experimental Psychology, 26,* 63-71.

Rescorla, R. A. (1972). Informational variables in Pavlovian conditioning. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 6). New York: Academic Press.

Rozin, P. (1976). The evolution of intelligence and access to the cognitive unconscious. In J. A. Sprague & A. N. Epstein (Eds.), *Progress in psychobiology and physiological psychology* (Vol. 6). New York: Academic Press.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88,* 638-667.