PAPER SUBMISSION FOR PSA 1998

The Dogma of Isomorphism: A Case Study from Speech Perception

Irene Appelbaum
Department of Philosophy
University of Montana

The Dogma of Isomorphism: A Case Study from Speech Perception

Abstract

It is a fundamental tenet of philosophy of the "special sciences" that an entity may be analyzed at multiple levels of organization. As a corollary, it is often assumed that the levels into which a system may be theoretically analyzed map straightforwardly onto real stages of processing. I criticize this assumption in a case-study from the domain of speech science. I argue (i) that the dominant research framework in speech perception embodies the assumption that units of processing mirror units of conceptual structure, and (ii) that this assumption functions not as a falsifiable hypothesis, but as an entrenched dogma.

WORD COUNT: 4,697

The Dogma of Isomorphism: A Case Study from Speech Perception

1. Introduction

It is a fundamental tenet of philosophy of the "special sciences" (Fodor 1974) that an entity may be analyzed at multiple, hierarchically organized, levels of structure. This claim has been interpreted in a number of ways. In some cases, a level of structure is taken to mean physical structure. In other cases, it is understood as abstract or conceptual structure. In still other cases, lower levels of structure are assumed to be physical, with higher levels becoming increasingly abstract. Thus, an organism may be analyzed in terms of successive levels of physical structure: atomic, cellular, physiological. A spoken utterance may be described at a number of levels of conceptual structure: phonological, syntactic, and semantic. And a cash register, as in Marr's well-known example (Marr 1982, 22-4) may be given a low-level description in terms of physical implementation or a high-level, abstract description in terms of the theory of addition .

In this essay I shall be primarily concerned with the claim that an entity may be described at successive levels of abstract or conceptual structure. I shall not be concerned to criticize or undermine this claim, but, rather, to distinguish it from another claim with which it is very often run together, and which I do wish to criticize. This is the assumption that successive levels of conceptual structure correspond to successive

stages of processing. That is, it is often assumed that the levels into which a system may be theoretically decomposed map straightforwardly onto the real stages of processing that produce the entity -- with the units in the first stage of processing mirroring the units at the lowest level of conceptual structure. More specifically, I will criticize this assumption insofar as it functions as a methodological constraint. That is, I will be concerned to reject the assumption that stages of processing *must* mirror levels of conceptual structure.

Another central tenet of philosophy of the special sciences is that philosophical questions are often not susceptible to global *a priori* answers, but must instead be addressed on a local, case-by-case basis. As such, my criticism will take the form of a case-study -- from the domain of speech science. My aim will be to show (i) that the dominant research framework in speech perception embodies the assumption that units of processing mirror units of conceptual structure, and (ii) that this assumption functions not as a falsifiable hypothesis, but as an entrenched dogma. Specifically, I will argue that the primary processing units in theories of spoken language processing are assumed to be isomorphic to units of phonetic structure -- the individual consonant and vowel sounds of language. Moreover, I will argue that in the face of contrary empirical evidence, rather than question this assumption of isomorphism, researchers simply stipulated that such isomorphism must exist at a different level of processing. Half a century of research, however, has failed to empirically confirm this stipulation. Yet the dogma that stages of processing recapitulate levels of conceptual structure

persists. The goal of this essay is to help loosen its grip, in speech science and, by example, in other special sciences as well.

2. Levels of Explanation

Although the idea of multiple levels of explanation is not new to the philosophy of science, it is only in contemporary discussions that it becomes a powerful tool. Hempel and Oppenheim, in their classic paper, "The Logic of Explanation" (1948) discuss the concept in connection with the idea of emergence. In this article, as in more contemporary discussions, the acceptance of higher levels of explanation is defended as being compatible with the generality of physics. Earlier as well as more recent discussions of higher-level explanation reject dualism and occult causation. However, the means by which each secures this defense of materialism differs.

The difference turns on whether or not the acceptance of materialism implies reduction. On the earlier account it does; on the more contemporary view it doesn't. What is important is that it is only on this latter view that higher-level domains become legitimate scientific domains. Hempel and Oppenheim keep accusations of occult causation at bay by treating the idea of higher-level explanation as an essentially stop-gap measure for dealing with the current limitations of our knowledge. That is, the idea of emergence is admitted, but only as an epistemological claim -- it flags our ignorance; it is not a claim about the world:

> [E]mergence of a characteristic is not an ontological trait inherent in some

> phenomena; rather it is indicative of the scope of our knowledge at a given time; thus it has no absolute, but a relative character; and what is emergent with respect to the theories available today may lose its emergent status tomorrow. (Hempel and Oppenheim 1948, 335-6)

What is objectionable about this way of understanding emergence is not that we venture explanations relative to the current state of our knowledge (as opposed to doing what?), but rather that the resting point for explanation is always at the level of micro-structure or physics:

> [T]he assertion that life and mind have an emergent status...can be summarized approximately by the statement that no explanation, in terms of micro-structure theories, is available at present for large classes of phenomena studied in biology and psychology. (Hempel and Oppenheim 1948, 336-7)

On the Hempel and Oppenheim understanding of emergence, then, higher-level explanations are saved from the occult only by in effect being deprived of their reality. The defense of higher-level explanations in current discussions of the philosophy of the special sciences makes no such sacrifice. Higher-level explanations, on this view, are irreducible not simply unreduced. In very nearly direct opposition to the above quotes, Fodor, for example, states:

> [T]here are special sciences not because of the nature of our epistemic relation to the world, but because of the way the world is put together: not all the kinds (not all the classes of things and events about which there are important counterfactual supporting generalizations to make) are, or correspond to, physical kinds. (Fodor 1974, 24)

The assumption that the commitment to materialism is to be severed from a commitment to reductionism is a fundamental tenet of in the philosophy of higher-level sciences. Defending philosophy of psychology, Putnam asserts, "while materialism is right...acceptance of these doctrines need not lead to reductionism" (Putnam 1981, 205) Fodor elaborates, "the assumption that every psychological event is a physical event does not guarantee that physics...can provide an appropriate vocabulary for psychological theories" (Fodor 1974, 17). And Wimsatt observes, "whatever the *promises* of reductionism, we do not in any interesting cases actually *have* the complete lower-level descriptions necessary to make upper-level descriptions and causal talk redundant" (Wimsatt 1976, 210).

An immediate consequence of recognizing the ontological claim of higher-level explanations is that it shifts the goal of explanation away from the lowest level in favor of the relevant one. That is, for some phenomena, a higher-level explanation may be "rock bottom" in terms of explanatory power. An explanation at too low a level of description, as Kitcher notes, will not only not increase explanatory power, it may

actually decrease it, by "adding a welter of irrelevant detail" (Kitcher 1984, 347). Putnam echoes this emphasis on relevance over reduction: "The relevant features of a situation should be brought out by an explanation and not buried in a mass of irrelevant information" (Putnam 1981, 206).

3. Levels of Structure vs. Stages of Processing

The distinction between levels of abstract structure and stages of concrete processing must be understood against the background of the assumption that a complex system is susceptible to a number of levels of explanation. But the structure/process distinction cannot be understood simply as a distinction between two different levels of explanation. To a first approximation, rather, two different levels at which a system's behavior can be explained correspond to two different structural descriptions of the system. Thus a particular spoken utterance might be analyzed in terms of its acoustic structure or in terms of its phonetic structure. More generally, an utterance may be given structural descriptions at successive levels of linguistic organization: phonological, syntactic, semantic. Such descriptions taxonomize the domain in terms of its content (i.e., in terms of its content described at some particular level). Successive levels of organization are compositional and hierarchical, but since there may be mismatches among the units at successive levels, the levels exhibit relative autonomy from one another.

A phonetic transcription may be thought of as a structural theory of the sounds of a

language. Like any theory, it specifies primitives and rules for combining them. The primitives are the individual phonetic segments (or 'phones') and the primary well-formedness constraint is that these segments be concatenated left to right. Also like other theories, a phonetic transcription is a kind of idealization. It does not attempt to describe the speech signal in full-blown detail, but only to capture those aspects of the signal that are thought to be relevant to identifying and distinguishing the sounds of human languages.

For example, a 20 msec difference in the onset of voicing (i.e., vocal fold vibration) would not be represented in a phonetic transcription of the utterance if it were, for example, the 20 msec difference between +50 msec voice-onset time (VOT) and +70 msec VOT, because this 20 msec difference isn't used to mark a phonetic contrast in any language. By contrast, the 20 msec difference in VOT between +20 msec and +40 msec would be represented in the structural theory of the sounds of the language because this difference is sufficient to distinguish, for example, the class of voiced stops from the voiceless stops in a language such as English. In a particular case, this information might signal the difference between e.g., [b] and [p], and this difference, in turn, is sufficient to signal a difference in meaning (e.g. 'bat' vs. 'pat').

As with other theories, the formalism in which a transcription is expressed embodies certain commitments about the nature of speech. Most important among these commitments are the idea that the sounds of speech can in fact be isolated as units the

size of individual consonants and vowels; that these individual segments can be serially combined to produce words, and that individual segments of a single type are uniform and interchangeable. The English word 'bat', for example, is thought to be composed of the three individual phonetic segments, [b], [a], and [t], combined linearly to produce 'bat', and it is also thought that one could substitute the [b] from 'bit' to produce 'bat'.

Theories of *how* speech is perceived -- not of the content or structure of *what* one perceives, but of the mechanisms by which this goal is accomplished -- are theories of speech processing. A speech perception theory may have as its goal an explanation of how a perceiver processes the continuous acoustic signal to yield segmented phonetic mental structure, but simply specifying the structural decomposition of the speech signal is not sufficient for a theory of the process of speech perception. A theory of processing must in addition specify *how* this lower-level structure causes the perception of phonetic structure.

4. Speech Perception: A Case Study

Before World War II it was an unquestioned assumption that the serially-ordered individual consonant and vowel sounds into which we analyze speech, correspond one-to-one to discrete serially-ordered segments of acoustic structure. According to this view, speech was thought to be a kind of sound alphabet and perceiving speech was assumed to consist in a simple process of matching acoustic segment to perceived

phoneme. Phonemes, on this view, were said to be strung together like "beads on a string". Despite its widespread acceptance shortly after World War II, this was a view which speech scientists almost uniformly regarded as false.

What happened in the interval to cause this rapid and decisive reversal is not a matter of much controversy. During this period experiments using new techniques for analyzing and synthesizing speech first revealed that there was no simple mapping between segments of phonetic structure (e.g. individual consonant and vowel sounds) and segments of acoustic structure. The acoustic cues for individual phonetic segments proved to be both variable and intertwined. Phonetic segments could not be defined in acoustic terms and connected speech could not be segmented into serially ordered units. This newly amassed evidence made it difficult to avoid the conclusion -- no matter how deeply entrenched the assumption to the contrary -- that speech was not a sound alphabet.

But if not a linear arrangement of acoustic segments, what, the question became, is the nature of speech. A consensus soon emerged that the relation between the speech signal and the phonetic structure of an utterance was that of a code. The term 'code' was meant to indicate that the phonetic structure was represented in the acoustic signal in a complex manner. Instead of a simple one-to-one mapping, the relation between units at the acoustic level and units at the phonetic level was found to be many-to-many. Instead of being discrete and serially-ordered, acoustic cues for

individual consonant and vowel sounds were found to be overlapped.

Along with this new conception of the nature of speech came a new conception of the process of perceiving speech. For the view that the speech signal was a coded version of phonetic structure led naturally to the view that speech perception was essentially a process of de-coding this structure. While on the discarded view, the process of speech perception had been understood to be a straightforward one of associating unit sounds with unit symbols, according to this new consensus, speech perception was a process which extracted or recovered the phonetic structure from the acoustic signal in which it was said to be enmeshed.

Reinforcement for the view that speech was a code came from an understanding of the processing limitations of both the auditory and articulatory systems. The rate at which phonetic segments are transmitted in human speech is far too high for them to be serially communicated. That is, listeners can perceive speech at rates of about 30 phonetic segments per second. If segments were communicated one-by-one at this speed the auditory system could not resolve them fast enough to perceive them distinctly. Indeed, at this rate not only wouldn't speech sounds be distinguishable from one another, they would not be identifiable as speech at all. By contrast, the understanding of speech as a code implies that the acoustic cues for individual phonetic segments are transmitted in parallel; the number of individual sounds that has to be processed at any one time is thereby greatly reduced.

Facts about the production process offered an explanation of how the encoding of phonetic structure occurred. Associated with each phonetic segment is a static vocal tract configuration. But in the production of successive phonetic segments -- that is, in the production of normal speech -- the articulators are in continuous transition from one configuration to the next. As a consequence, the articulatory configurations for individual segments become intertwined and the configuration for any particular phonetic segment is influenced by that of the preceding and succeeding segment. The anatomical and physiological design of the articulators seemed to guarantee that the attempt to produce a series of discrete phonetic segments would yield a signal in which the acoustic cues for individual segments are both overlapped and variable.

The speech signal, then, was shown to be constrained both by features of the production process and by features of the articulatory process. In both cases, these constraints proved to be incompatible with the assumption that speech was a linear arrangement of discrete sound segments. In other words, the claim that speech *was* not an acoustic alphabet gained support from the recognition that it *could* not be.

Still, if the encoded view of speech resolved a problem that the alphabetic view left unresolved, it also created a problem that the alphabetic view did not create. For although the view that speech was a code made sense of the experimental evidence, and was compatible with physiological constraints, it left unanswered the central question of how speech was perceived. On the earlier understanding of speech - the

view of speech as a sound alphabet, the question hardly arose: since phonetic structure was thought to be transparently represented in the acoustic signal, perceiving the signal and perceiving the phonetic structure were assumed to be one and the same.

But if, as the new consensus maintained, phonetic structure was represented in the acoustic signal in a manner which was both complex and variable, the question of how phonetic structure was perceived became a puzzle. Since individual speech sounds were perceived as being serially ordered but were not present in the acoustic stream as such, a question arose as to how -- by what mechanism -- the encoded phonetic structure was decoded and the linear ordering of segments (that one perceived) restored.

Since the acoustic cues for particular phonetic segments were not only encoded but also variable, there was an additional question to be answered. For if the acoustic cues for a particular phonetic category were neither stable nor isolable -- if different tokens of a phonetic type might have little in common acoustically -- it was far from obvious how instances of a particular speech sound could be recognized as instances of the same speech sound. That is, without invariant acoustic properties, it was far from obvious how tokens of the same speech sound could sound the same.

Since it seemed clear that the acoustic signal (as displayed in spectrographic analyses)

lacked invariant properties, a crucial assumption was made that invariant properties must exist at some *other* level of the production-perception process. It was these "surrogate" invariant properties which were thought to be recovered from the speech signal in the decoding process. And it was these invariants that were thought to be identified with the individual serially-ordered phonetic segments that one perceived. What remained unanswered, though, was what these invariant properties were.

Solving the "lack of invariance problem", as this problem came to be known, became, and in large measure remains, the central goal of research in speech perception. The numerous speech perception theories that have developed during the past 50 years may be

distinguished largely in terms of their proposed solutions to this problem. Thus, the *motor theory of speech perception* (Liberman, Cooper, Shankweiler, Studdert-Kennedy 1967; Liberman and Mattingly 1985) claims that phonetic segments are to be identified with invariant neural structures; the *ecological theory to speech perception* (Fowler 1986, 1989) claims that they are to be identified with the articulatory level of production; and the *theory of acoustic invariance* (Blumstein and Stevens 1981; Stevens and Blumstein 1981). remains committed to the view that the invariants in speech perception are acoustic properties -- though of a different sort than those made salient in early speech experiments.

The current period of research in speech perception is thus guided by the assumption

that speech is a complex and variable code; that perceiving speech involves a decoding process which recovers serially ordered invariant units which need not themselves be acoustic; and that discovering the precise nature of these invariant units is the primary goal of a theory of speech perception.

Thus, the view of speech which developed in response to the surprising results of early experiments in speech synthesis and analysis, and which continues to dominate speech perception research today, is widely thought to mark a radical departure from the long-standing view of speech as a simple sound alphabet. Still, while it is difficult to overlook the differences between these two views, it is easy to overstate them. In particular, the current view of how speech is processed continues to be guided by the assumption that the primary processing units must be isomorphic to units of phonetic structure.

It is not that the acoustic properties of the speech signal which spectrographic analyses showed to be encoded are alphabetic after all, but rather, that the claim that the speech signal is encoded, itself presupposes a view of speech as alphabetic. For the code which the speech signal represents is precisely a code of phonetic -- that is, alphabetic -- structure. It is phonetic segmental structure which, according to the current view, is encoded in the production process and decoded in the perception process. So, although the current view denies that there are alphabetic segments in the acoustic signal, it presupposes the existence of such segments at both ends of the speech

chain: the production process originates with them and the perception process ends with them.

The invariant properties which are thought to be recovered from the acoustic signal, and by which the phonetic percepts are thought to be identified, also represent an alphabetic structure: a linear arrangement of discrete symbols each one standing for a single consonant or vowel sound. On the previous view, the process of perception was thought to be a straightforward process of matching invariant acoustic units one-to-one to phonetic units. On the current view, the invariant units need not be acoustic and the process which recovers them from the acoustic signal is a complex rather than straightforward one; nevertheless, the basic process of mapping invariant units one- to-one to phonetic units remains the same.

What has thus been heralded as a radical change in the nature of speech was really only acoustic-surface deep. The change from thinking of speech as a sound alphabet to thinking of it as a sound code was not a change in the understanding of the nature of speech, but only in the locus of it. For the recognition that the acoustic signal was not alphabetic did not lead researchers to abandon the assumption that speech was alphabetic, but only to seek this alphabetic structure elsewhere in the speech chain. The current consensus that the goal of a speech perception theory is to identify a set of discrete, serially ordered invariant units -- that is, that the goal of a speech perception theory is to solve the lack of invariance problem -- is testimony not to the decisiveness

of the rejection of the earlier view, but to the tenacity of its grip. Thus, the central role which solving the lack of invariance problem occupies in current speech perception research is evidence, not that the alphabetic conception of speech has been replaced, but on the contrary, that it has never been seriously questioned.

5. Conclusions

The assumption that speech is a sound alphabet and that it is perceived by a process that matches serially ordered perceptual processing units to serially ordered phonemes was perhaps the "natural" starting point for theorizing about speech perception. However, when experimental results clearly showed that this view was false, instead of discarding the background methodological assumption embodied in this view, speech perception researchers simply discarded the most superficial implementation of it. That is, when forced to abandon the thesis that speech is an acoustic alphabet, they rejected only the view that it is acoustic, not the view that it is alphabetic. Speech perception scientists never questioned the assumption that the primary processing units -- the invariant objects of perception -- must be discrete, serially ordered units corresponding one-to-one to phonetic or alphabetic segments. Rather, taking for granted the view that speech *must* be perceived by a process which recovers invariant, alphabetic units, and faced with evidence that these were not simple acoustic units, speech perception researchers were almost forced to conclude that these invariants exist at some other level of processing.

Understanding the lack of invariance problem against the backdrop of the assumption that stages of perceptual processing mirror levels of linguistic structure helps to explain how the *fact* of lack of invariance became the *problem* of lack of invariance. For it is only if one is committed to such an assumption that one will expect to find isomorphic mappings between linguistic units (in the present case, phonetic units) and units of processing. Despite the fact that experimental evidence has shown the multiple-realizability and context-dependence of phonetic cues to be pervasive in speech, such facts must, if one subscribes to this underlying assumption, be explained away. And it is the identification of underlying alphabetic invariants that is thought to hold to the key to such an explanation.

Nevertheless, the sobering fact is that after nearly five decades of effort to identify such invariants, most theories of speech perception cannot produce a single empirical invariant property. The one theory that has produced at least a candidate invariant property is the theory of acoustic invariance (Blumstein and Stevens 1981; Stevens and Blumstein, 1981) -- a theory which claims that the invariants in speech perception are acoustic, after all, but of a different sort than those traditionally expected. But even in this case only a single invariant property has been suggested for a single class of consonants, and it is found in only about 85% of cases.

It is important to note that although I have been arguing that speech perception researchers should have interpreted the early evidence of complex mappings between

phonetic and acoustic structure as evidence against the isomorphic assumption, rather than simply against its acoustic implementation, one must understand the sense of 'should' here in an appropriately anachronistic sense. For I am not suggesting that this was a mistake early researchers could reasonably have been expected to have forseen. Indeed, it is a mistake that can only be properly diagnosed *a posteriori*. The point here is not simply that one could not reasonably choose between criticizing the assumption that speech is a sound alphabet and that it is alphabetic *simpliciter*, in advance of the results of future research. The important point, I think, is that it is only in light of such future research developments and failures that this assumption first becomes fully articulated and disambiguated. That is, it is reasonable to assume that early researchers did not identify the problem at the level of the underlying isomorphic assumption, in part because they had not yet isolated it *as* an assumption -- they had, thence far, no reason to do so. Thus the rejection of the isomorphic assumption as a guide to explaining speech perception that I have been urging is, to some important extent, made possible by the very persistence of this assumption.

But for current speech perception scientists the situation is different. After half a century of failed research, the persistence of the isomorphic assumption -- in the form of the continued search for invariants -- can only be seen as a dogma.

REFERENCES

Blumstein, S. and K. Stevens (1981), "Phonetic Features and Acoustic Invariance in Speech", *Cognition 10*: 25-32.

Fodor, J. (1974), "Special Sciences, or the Disunity of Science as a Working Hypothesis", *Synthese 28*: 97-115.

Fowler, C. (1986), "An Event Approach to the Study of Speech Perception from a Direct-Realist Perspective", *Journal of Phonetics 14*: 3-28.

Fowler, C. (1989), "Real Objects of Speech Perception: A Commentary of Diehl and Kluender", *Ecological Psychology 1*: 145-60.

Hempel, C. and P. Oppenheim (1948), "The Logic of Explanation", in H. Feigl and M. Brodbeck (eds.), *Readings in the Philosophy of Science* New York: Appleton-Century-Crofts, pp. 319-352.

Kitcher, P. (1984), "1953 and All That. A Tale of Two Sciences", *The Philosophical Review 93*: 335-73.

Liberman, A.; F. Cooper; D. Shankweiler; M. Studdert-Kennedy (1967), "Perception of

the Speech Code", *Psychological Review 74*: 431-461.

Liberman, A. and I. Mattingly (1985), "The Motor Theory of Speech Perception Revised", *Cognition 21*: 1-36.

Marr, D. (1982), *Vision*. New York: W.H. Freeman & Company.

Putnam, H. (1981), "Reductionism and the Nature of Psychology", in J. Haugeland (ed.), *Mind Design*. Cambridge, MA: MIT Press, pp. 20-219.

Stevens, K. and S. Blumstein (1981), "The Search for Invariant Acoustic Correlates of Phonetic Features," in P.D. Eimas and J.L. Miller (eds.), *Perspectives on the Study of Speech*. New Jersey: Erlbaum, pp. 1-38.

Wimsatt, W. (1976), "Reductionism, Levels of Organization, and the Mind-Body Problem", in G. Globus; G. Maxwell; I. Savodnik (eds.), *Consciousness and the Brain: A Scientific and Philosophical Inquiry.* New York: Plenum Press, pp. 199-267.