ECOLOGICAL PSYCHOLOGY
2016, VOL. 28, NO. 3, 160–165
http://dx.doi.org/10.1080/10407413.2016.1195188

Routledge
Taylor & Francis Group

Perceptual Organization and Lawful Specification

Robert E. Remez[a] and Philip E. Rubin[b]

[a]Program in Neuroscience and Behavior, Columbia University; [b]Haskins Laboratories

ABSTRACT

When a listener can also see a talker, audible and visible properties are ineluctably combined, perceptually. This perceptual disposition to audiovisual integration has received widely ranging explanations. At one extreme, accounts have likened perception to a blind listener and a deaf viewer combined within a single skin, resolving discrepancies in identification by each modality. At the other extreme, perception has been described as necessarily and automatically synesthetic. Useful descriptive and explanatory evidence was provided in a study of auditory-haptic presentation by Fowler and Dekle (1991) showing that neither familiarity nor congruence is required for perceptual integration to occur across modalities. Instead, the notion of conjoint lawful specification was proposed as a governing constraint. This principle treats sensory activity as proximal sampling of the properties of distal objects and events, and this essay notes that its corollaries offer a broadly applicable guide in contemporary investigations of perception.

It is not a perfect case of déjà vu, but accounts of ancient natural history sometimes have seemed just like direct realism. In describing the ancient origins of photosensitivity, a familiar story proposes that a free-ranging unicellular organism evolved an ability to detect the difference between day and night. An organelle responsive to light, its eyespot, sensed ambient illumination, by virtue of which this dinoflagellate could achieve longevity, such as it was, by hunting during the day for a bright region in the ooze where the sun could power its chloroplasts. It is hardly controversial, apparently, for this account to aver simply that a property of an environment is available for detection by a sensitive creature or that a creature can act to change its local conditions if these are noticed. Yet, for multicellular multisensory beasts such as ourselves, the premises of perceptual accounts have differed. A classical tradition in perceptual explanation describes the properties of objects and events as incommensurate with the properties of the energy arrays that excite the senses, and the subjective states deriving from sensory activity are described as bound to the eliciting energy, for which reason they must remain incommensurate with the properties of objects and events that perception somehow provides. In classical accounts, sensory states can be knowledge states for a protist but not for a thinker!

With an empirical project as springboard, Fowler and Dekle (1991) proposed the concept of "conjoint lawful specification" (p. 817) and in so doing asserted that linguistic properties of utterances can be sensed multimodally in patterns of available energy. Under such conditions,

CONTACT Robert E. Remez remez@columbia.edu Department of Psychology, Barnard College, Columbia University, 3009 Broadway, New York, NY 10027-6598.

perception can be veridical and direct, at least hypothetically. The critical premises of their exposition of specification include the causal and sufficient relations obtaining between: (a) objects and events and (b) the arrays of energy available to a sensitive agent, however many cells compose it, which invariantly exhibit a pattern specific to the objects and events. Prior generations of perceptual theories had asserted a causal relation (e.g., Mach, 1886/1897; Tolman & Brunswik, 1935) between object properties and patterned stimulation, which permitted accounts to warrant realism. Yet, the projection of the rich and varied dimensions of material objects into the poor and few immaterial dimensions of light or sound seemed to impose the condition of insufficiency. Specifically, the dimensions of objects (composed of wood, or metal, or stone, or glass, or muscle, or fluid, etc.; durable, perishable, rigid, pliable, graspable, insubstantial, etc.) differed from the dimensions of patterned energy (changing contours of intensity and hue differing solely in brightness, spatial frequency, and shape or changing contours of frequency and spectrum differing in intensity, duration, location, etc.). The upshot of this mismatch in dimensions obliged a psychological engine to convert the properties of stimulation into an apprehension of objects, constructing knowledge of objects out of sensory activity. This necessity of conversion, whether by logical function or by actuary, makes the classic account indirect.

The specific project of Fowler and Dekle (1991) led to a report of sensory sufficiency, stopping short of a proof that a specific energy pattern mediating between spoken syllable and perceiver was sufficient and invariant. Their project was based on the method developed by McGurk and MacDonald (1976) in which independent visible and audible displays of speech were presented for a perceiver to fuse or not, as the case might be (see Rosenblum, 2008). A simple, theoretically neutral question supplied the motive: What drives intersensory combination in the perception of consonant-vowel syllables? One set of conditions using sound and text derived from the classical indirect realism of perceptual explanation, namely, a psychologically powered projection from sensory states to the phonemic properties of syllables, achieved by reference to the probable association of each sensory sample and the linguistically governed segments in the phoneme inventory. Inasmuch as such projections are said to depend on distributions of sensory and linguistic properties committed to memory by rote, this explanation presumes that good function depends on familiarity with the sensory samples as well as the phonemic types. The relevant test included an assay of perceptual resolution under familiar conditions, as researchers mainly do. But, by modus tollens, a second set of conditions used sound and touch to test if perception failed when familiar conditions were denied to the perceiver. Because familiarity was not required for sensory combination to occur, and because sensory combination failed despite familiarity of the sensory samples, Fowler and Dekle advanced an alternative explanation, invoking direct realism and appealing to specification jointly across two sensory modalities.

To be more specific, they argued that a speaker of English can listen to a syllable and determine whether it begins with B or G, and because this listener can also read, a printed character presented concurrently on a display can also indicate orthographically whether a syllable begins with B or G. The experimenter's trick was to vary the sound and the text independently and to see if the perceiver's reports showed evidence of blending the heard and read property. Now, each of us ordinarily has a history of listening to speech beginning in the nursery and only a slightly shorter span of familiarity with print. Yet, despite extensive familiarity with the audible characteristics of utterances and the visible appearance of print, only linguistic properties and their auditory sensory effects are causally related. An orthographic contour is arbitrarily and conventionally related to the phoneme that it represents,

whereas the acoustic properties of speech are a consequence of physical sound production in linguistically governed movements of tongue, jaw, and lips. If perceptual impressions follow the sensory effects of specification—a pattern of energy sampled by sensory modalities structured by objects and events and sufficient to convey their attributes—then despite the familiarity of print, the arbitrariness of orthography must preclude integration with audible speech.

In their second set of multimodal conditions, a native English-speaking listener was presented with an utterance of a syllable beginning with B or G and was asked concurrently to place a hand on the face of the talker with eyes closed and to feel her produce speech. Once more, these heard and felt events were controlled independently, and the experimenters looked for evidence of fusion in the identification reports. Because of the relative rarity of perceiving speech by relying on touch (called Tadoma, eponymously, for the first two deaf-blind individuals who were taught the method, Tad and Oma; see Alcorn, 1932), this condition of sound-plus-touch was an effective control for the condition of sound-plus-print. Whereas print was familiar but arbitrary in relation to the articulation causing the acoustics of speech, the feel of the articulating face was causally related to the acts of sound production and hypothetically sufficient to convey their sequence, but thoroughly unfamiliar. The results of the study, nuisances of method aside, showed that perceptual reports reflected the integration of familiar acoustics and unfamiliar haptics, whereas familiar text had meager influence on the identification of the spoken syllables. It seemed to the authors as though touch and listening combined because "information in different media that are joint consequences of the same event … serve jointly to specify the event to the perceiver … without prior familiarity with it because of the causal chain that supports perception. Stimulation caused by an environmental event has causal effects on sensory receptors so that its structure is, in part, transmitted to a perceptual system. By hypothesis, the perceiver comes to know an event in the environment by way of its impact on the perceptual systems as transmitted by proximal stimulation" (Fowler & Dekle, 1991, p. 817).

The notion expressed here is an echo: "patterns in the flux of sound, touch, and light from the environment may be equivalent to one another by invariant laws of nature" (Gibson, 1966, p. 115). The challenge for the perceiver, well met in the study by Fowler and Dekle (1991), is to resolve common patterning equivalently or harmoniously in different senses despite the unique dimensions of each sense. But Gibson's oracular diction must not discourage us from noting obvious and intrinsic departures from equivalence. The optic array is fundamentally superficial, that is, its patterning is imposed by the properties of surfaces and shallow translucent subsurfaces (Gibson, 1966). In contrast, acoustic waves are a consequence of the restoring forces of objects that return a substrate deformed by a driving force to its initial state, and this is an effect of its material composition. Acoustic waves are produced when this motion is imposed on enveloping air. For this reason acoustics provides a pattern tied more closely to substances than to surfaces (Gaver, 1993). Indeed, some objects—the chest wall, for instance—are opaque to light and transparent to sound, a difference in the patternability of energy media with useful consequences as well as some annoyances commonly noted by engineers in the soundproofing business. Touch and especially haptic sensitivity allow a squeeze to resolve the compressibility and distensibility of an object when neither property readily patterns an optic or an acoustic array. Whatever versatility in perception is permitted by common patterning of flux, the properties available in each modality must differ irreducibly as a consequence of the physical causes that impart a

sensory pattern. Nonetheless, to fulfill the declaration that conjoint lawful specification was probably at work in the perception of sound-plus-touch requires identifying the pattern that is hypothetically present in sound and the adjustments of hand shape. If this is largely a matter of physical description (Stoffregen & Bardy, 2001), an assay to determine whether a hypothetical pattern lies within the range of human sensitivity brings it into the realm of psychophysics, and an appraisal of the granularity matching the English initial phoneme inventory sharpens the descriptive and theoretical focus to the perception of speech.

For Fowler and Dekle (1991), performance level was the surrogate for a technical description of a hypothetical amodal flux pattern available to touch and hearing alike. They noted the greater plausibility of conjoint specification than probabilistic discrimination in describing their finding. However, they left to other researchers who follow in their wake to determine whether fluency or accuracy in this empirical procedure indicated an effect of invariant patterns of energy eliciting auditory-haptic perception. Regarding the properties of phonemes, the existence of invariant patterns remains hypothetical, and unsecured. That is, adopting Gibson's clairvoyance, Fowler confidently promised that "phonological gestures are the public actions of the vocal tract that cause structure in acoustic speech signals. By hypothesis, *they will be found to cause specifiers or invariants*" (Fowler, 1996, p. 1731, emphasis added).

In the 25 years since it appeared, the report by Fowler and Dekle (1991) has been widely cited and is acknowledged as a model for research on perception in several concurrent senses. Theoretically, it draws fire now and then (e.g., Mitterer & Jesse, 2010), but this reflects the devotion of our dismal era to the methods of insurance companies and an attendant commitment to portray human knowledge in incidence reports of sensations and the events with which they happen to be associated. No convincing evidence has yet appeared to show that cases of perceptual success in the absence of familiarity should be attributed to statistically rationalized schematic representations of familiar sensory states projected into perceptual types. In this regard, history teaches that success without familiarity is an old story, constituting a principal and vexing concern of E. L. Thorndike (described in Hilgard & Bower, 1975) over decades. Taking a long view about a local dispute, it seems that subjects in experiments, whether rats or people, are not very sensitive to differential incidence. Had Fowler and Dekle pushed a bit harder at the explanation they disfavored, it might simply have failed and left the building. (See also Luce, 2003, for the argument that differential incidence is only ever salient in a two-alternative forced-choice task. Otherwise, the richness of perceptual experience precludes attention to mere differences in incidence or its perceptual counterpart, differential familiarity.)

Come what may, the appealing simplicity of their original explanation must be relinquished to accommodate new evidence. To take one of the more charming examples, consider a pair of studies on the topic of voicing in obstruent consonants (Derrick & Gick, 2013; Gick & Derrick, 2009). In American English, the dominant acoustic property conveying voicing is the presence or absence of aspiration (Lotz, Abramson, Gerstman, Ingemann, & Nemser, 1960). That is to say, despite the vulgar near synonymy of "voicing" and Voicing Onset Time (VOT), the perceptually criterial acoustic difference is absence of aspiration (voiced) or presence of aspiration (voiceless). In an auditory tactile presentation, participants in a study by Gick and Derrick (2009) integrated the audible properties of speech and the tactile experience of a puff of air presented to the neck or hand in perceiving the voicing feature. Their report notes that perceivers are often close enough to talkers

to experience the aerodynamic consequences of speech production along with the acoustic, even though the tactile presentation in their project included only an explosive release and its lagging airflow for 80 ms, considerably briefer than the roughly 400 ms duration syllables. Under such circumstances, the finding is readily resolved to the conditions of the bad old cognitive view in which probability distributions are consulted covertly to make otherwise inscrutable sensory experience portray the linguistic properties of syllables. Yet, according to the premises of conjoint lawful specification, an acoustic pattern at the ear and an air current at the neck or hand might well be joint consequences of the same event and readily combined in multimodal sensory sampling. A succeeding study (Derrick & Gick, 2013) relieves this uncertainty while introducing a new problem.

The critical condition in the procedure used by Fowler and Dekle (1991) was the lack of familiarity of the auditory-haptic experience so readily transcended by their participants. For Derrick and Gick (2013), this lack of familiarity was accomplished by running the latex hose that delivered the puff of air down to the participant's ankle; it was guaranteed to create an unfamiliar experience for any participant who spends little time with a talkative podiatrist. Under those circumstances, an acoustic-plus-tactile combination satisfied the condition of unfamiliar presentation that Fowler and Dekle had described. Despite all, perception of consonant voicing exhibited an undeniably weird blend of audible and tangible properties, although it should be conceded that the participants did not sample an utterance concurrently in two modalities, as in the method of Fowler and Dekle. Rather, while the acoustic sample of a syllable was available, a brief burst of air was imposed. Evidently, perceptual integration allows an analogy in which a puff of air on the ankle stands for a puff of air of an aspirate consonant release.

The discovery here is complicated. Briefly, a "stands for" relation differs from an identity relation that obtains when a single event is tracked in two modalities. With these conditions of presentation, it is difficult to claim without special exclusion that the tangible and audible properties in feeling with the foot and listening with the ear are naturally merged as joint consequences of the same event, much as feeling a talking face and hearing it speak were. Worse, the finding introduces the possibility that the original study could also have expressed this allowance for analogy rather than expressing a direct sensitivity to an invariant that transcended modality. Impressed with the versatility of perceivers, Derrick and Gick (2013) conclude their report with an agnostic note about how it could have happened, saying again that familiarity is evidently unimportant to perceptual organization.

Of course, we knew this from decades of research specifically about the perceptual organization of speech (Remez, Rubin, Berns, Pardo, & Lang, 1994; Remez, Rubin, Pisoni, & Carrell, 1981; Remez & Thomas, 2013). That terrifying story is better left for another day. For now, we note that the value of a theoretical claim is its clarity and its pertinence, for it will surely be wrong, if not immediately, then eventually. May we all be wrong in useful ways, and may we continue to find inspiration in the idea of conjoint lawful specification.

Acknowledgment

We thank Emily Thomas for applying her gimlet eye to a draft of the essay and Michael Studdert-Kennedy for counsel on the topics of grammar and rhetoric.

Funding

For supporting this project, we are grateful to the National Institute on Deafness and Other Communication Disorders for Grant DC000308 and to the National Institute of Child Health and Human Development HD 01944.

References

Alcorn, S. (1932). The Tadoma method. *The Volta Review*, *34*, 195–198.
Derrick, D., & Gick, B. (2013). Aerotactile integration from distal skin stimuli. *Multisensory Research*, *26*, 405–416.
Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, *99*, 1739–1741.
Fowler, C. A., & Dekle, D. J. (1991). Listening with the hand and eye: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 816–828.
Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, *5*, 285–313.
Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin.
Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*, 502–504.
Hilgard, E. R., & Bower, G. H. (1975). *Theories of learning* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
Lotz, J., Abramson, A. S., Gerstman, L. J., Ingemann, F., & Nemser, W. J. (1960). The perception of stops by speakers of English, Spanish, Hungarian, and Thai: A tape-cutting experiment. *Language and Speech*, *3*, 71–76.
Luce, R. D. (2003). Whatever happened to information theory in psychology? *Review of General Psychology*, *7*, 183–188.
Mach, E. (1897). *Contributions to the analysis of the sensations and the relation of the physical to the psychical* (C. M. Williams, Trans.). Chicago, IL: Open Court. (Original work published 1886)
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
Mitterer, H., & Jesse, A. (2010). Correlation versus causation in multisensory perception. *Psychonomic Bulletin & Review*, *17*, 329–334.
Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, *101*, 129–156.
Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, *212*, 947–950.
Remez, R. E., & Thomas, E. F. (2013). Early recognition of speech. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*, 213–223. doi:10.1002/wcs.1213
Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*, 405–409.
Stoffregen, T. A., & Bardy, B. G. (2001). On specification and the senses. *Behavioral and Brain Sciences*, *24*, 195–261.
Tolman, E. C., & Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological Review*, *42*, 43–77.