

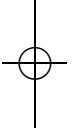


# 11

## Philosophical Messages in the Medium of Spoken Language<sup>1</sup>

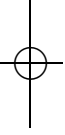
ROBERT E. REMEZ AND J. D. TROUT

### I. Introduction



The psychology of spoken language offers philosophical lessons about the potential for seduction by phenomenology, and for our readiness to adopt a reductionist metaphysics in building theory. To expose these themes, we will examine the most vexing issues in the science of speech perception, drawing evidence from an assortment of ordinary cases, from speech perception by the deaf and the cochlear implant user, and from extraordinary speech perception evoked by synthetic acoustic patterns created specifically to be impossible to vocalize. We then review the way in which dominant methods in the philosophy of mind use introspection, intuition, and reflection on phenomenal experience when drawing conclusions about the nature of perceptual experience. This contemporary research represents a substantial theory about how introspection and intuition work, a theory that can be wrong and shown to be so by appeal to empirical evidence.

Speech perception research is dominated by a theoretical perspective that, witting or not, relies on a reductionist perspective that offers comfort to the intuition that speech is, fundamentally, an articulatory act of sound production, despite the linguistic governance of the creation and understanding of every utterance. According to this dominant view, fundamental linguistic units are thought to have acoustic or articulatory characteristics essential to their identity. This version of psychological essentialism, irresistible to many, reflects reductionist commitments about the real causes of language performance: that the causes are sensory or motor properties concurrent with speech. Appealing



<sup>1</sup> The authors would like to thank Michael Bishop, an anonymous referee, and the editors for useful comments on an earlier draft.

to anti-reductionist accounts in different fields of naturalistic metaphysics, we will argue that the evidence heavily favors an alternative view in which significant speech units represent homeostatic properties.

## 2. The Character of Utterances

The technical study of spoken language begins with a puzzle. Talkers who are mutually intelligible—if not mutually comprehensible—share a common stock of linguistic forms. They share words, speaking Anglocentrically, and it would not incite controversy to propose that they also share the syntactic aggregation of words in superordinate phrases and clauses, as well as the syllables and phoneme segments subordinate to words. Speech is intelligible when a listener resolves linguistic form, a series of segments, syllables, words, phrases, and clauses that compose an utterance. The puzzle that launches the inquiry is the absence of close correspondence between the linguistic constituents that talkers share and the properties of physical expression. The articulation that produces an utterance, its acoustic effects borne on the air that envelops talker and listener, and the auditory qualities that a perceiver experiences are unique in each instance. Because each realization is different from any other, the relation between a linguistic form and its physical expression is one-to-many. There is neither an articulatory maneuver, nor an acoustic pattern, nor an auditory quality at the core of any linguistic form (Lieberman *et al.* 1967).

The research that defined the puzzle has a long history, and the search to identify criterial physical attributes of the symbolic properties of speech has often been spurred by practical as well as scientific purpose. The hierarchical nature of linguistic structure set the aims of this research. With the exception of homophones, words are distinguished by elementary units of contrast, the inventory of which composes the phoneme set of a language. English uses about three dozen of these phoneme segments expressing a dozen and a half contrast features. Typically, research on speech production and perception has taken this least constituent as its object, and in the lore that has developed among researchers, the perception of speech has come to denote the perception of the finite and small set of meaningless phoneme segments that compose the larger meaningful constituents. This has seemed like a sensible choice, inasmuch as the set of words known to a college sophomore can exceed 120,000, every one of which can be described as a series of phoneme segments drawn from the small set of three dozen (Miller 1951). Psychologically, the hierarchical nesting of structure fostered an account in which recognition of

a series of individual elementary contrasts expressed in an utterance permits a listener to distinguish words from each other. Of course, a listener might just guess that a talker said SHEEP, not JEEP or CHEAP,<sup>2</sup> but by attending to the fine grain of the incident sound, the expressed contrast is resolved perceptually, without speculating. But, if a perceiver is able to apprehend the linguistic form of an utterance perceptually, relying neither on guesswork nor telepathy, how is this accomplished?

### 2.1 *Reduction to Motor Types*

Accounts within psychology have favored reduction of phonemes to articulatory motor categories or to auditory sensory categories. Antique descriptions stood squarely on articulatory linguistic intuition—that is, on the impressions of a thoughtful talker considering the placement and motion of articulators when producing ordinary utterances. The apparent commutability of discrete segments—that RIGHT, TRY, and TIRE are different orders of the same phonemes, spelling notwithstanding—led to a conceptualization of articulation in similar form (Kühnert and Nolan 1999). A small segmental phoneme inventory was paired with a small inventory of hypothetical vocal acts, with different concatenation orders of the articulatory ingredients responsible for articulation of different words with the same segments. If this seems like the infinite use of finite means, it is not accidental, although this characterization of articulation proved to be false. Once evidence was pursued instrumentally, the puzzle was defined.

The advent of spectrum analyzers made the fine physical acoustic properties of speech observable, and nothing like commutable segments has ever been seen. In a classic metaphor, it is as if:

Easter eggs [are] carried along a moving belt; the eggs are of various sizes, and variously colored, but not boiled. At a certain point, the belt carries the row of eggs between two rollers of a wringer, which quite effectively smash them and rub them more or less into each other. The flow of eggs before the wringer represents the series of impulses from the phoneme source; the mess that emerges from the wringer represents the output of the speech transmitter. At a subsequent point, we have an inspector whose task is to examine the passing mess and decide, on the basis of the broken and unbroken yolks, the variously spread-out albumen, and the variously colored bits of shell, the nature of the flow of eggs which previously arrived at the wringer. Notice that he does not have to try to put the eggs together again—a manifest physical impossibility—but only to identify. (Hockett 1955: 210)

<sup>2</sup> The notation convention used here puts words mentioned as words in small caps. Phoneme segments are idealizations, indicated by slashes; phonetic transcriptions or plausible notations tied to specific instances are phonetic objects, and are indicated by brackets.

Indeed, the discrete, commutable phonetic intentions of a talker are restructured for expression as a pattern of continuous articulation. An anticipated segment or two is produced with the present one (Bell-Berti and Harris 1979), while the residue of the articulation of a prior segment or two still influences the postures and motions of the vocal organs (Lieberman *et al.* 1967). In this circumstance, it is not surprising that the acoustic correlates of any individual segment vary with its proximate phonemic environment. The coincident production of conceptually successive segments is *coarticulation*, and it is one source driving the variation in correspondence between a linguistic segment and its physical expression. Although intuition conforms to the linguistic properties, namely, of discrete and serially ordered commutable segments, the articulation contradicts this simple portrait of expression. The articulation of a phoneme series is an intricate rather than a straightforward expression of the linguistic sequence, and techniques to observe articulators directly, via x-ray movies and cinefluoroscopy, brought anatomical precision to the description inferred from acoustics (Honda 1996).

Coarticulation of conceptually discrete and commutable consonants and vowels amalgamates the canonical articulatory forms of speech, conferring a graded and seamlessly progressing form upon the motor acts of expression. This merges the production of linguistically distinct consonants and vowels, obliterating the conceptual linearity of the series, and thereby blocking any model of speech production similar to typing, in which each phoneme takes an invariant articulatory form in each instance. Instead, the articulation of a given segment incorporates aspects of the preceding and succeeding units in the series, amalgamating features of its phonemic context with its own distinctive attributes. In fact, expressive variation also incorporates changes in the rate of speech, in the precision of speech, in the casualness or formality of the diction, in the intrinsic anatomical and functional differences between talkers, and in the paralinguistic expression of affect that converges in the vocal tract with linguistic expression. If the expression of any linguistically governed phoneme, for instance, /d/, varies in physical form without end, there can be no simple perceptual standard, neither articulatory nor auditory, that is adequate.

The perceptual trick required of a competent listener, according to the Motor Theory of Speech Perception (Lieberman and Mattingly 1985; see also Chapter 10), is to apply intimate and expert knowledge of speech production during the perceptual analysis of the incident acoustic stream to achieve the inverse of the functions projecting phonemes into articulation and articulation into acoustics. Each listener is also a talker in this conceptualization, able to ascribe aspects of concurrent articulatory states to different moments in a speech signal and to ignore the articulatory blending of phonemes in an utterance.

The technical ingenuity of the account is admirable despite the presupposition that the phoneme, a linguistic object, is a motoric object, discoverable by unmixing the physical ingredients that the talker put into the articulatory soup.

To imagine the action of the Motor Theory, consider a listener registering the sound of an utterance and sorting the waxing and waning coarticulated influences into a tidy sequence of ordered phonemes. Because the equation of phonemes and motoric types assumed that graded subphonemic variation can be attributed solely to coarticulation, it warranted a corollary claim that subphonemic variation is imperceptible. The reason is that the motoric knowledge acting as the engine of perception as well as production completely resolves a coarticulated utterance into its initiating phoneme series (Liberman 1970). Under many test conditions, psychoacoustic research on the perceptual differentiation of synthetically produced speech revealed that subphonemic variation is barely perceptible, as this reduction of perception to articulatory types claims (Liberman *et al.* 1957). This research defined the study of categorical perception. In other conditions, though, subphonemic variants are well resolved despite phonemic equivalence, as in the prominent contrast between casual and careful speech, or the contrast between the speech of Flatbush and the San Fernando Valley. The existence of salient subphonemic aspects of expression undermines the asserted equivalence of phonemic and articulatory types, and reflects a convergence of influences, of which the phoneme series is one, on the properties of utterances and on the perceptually resolvable properties of speech, in consequence.

Ultimately, the claims of the original Motor Theory fell due to empirical challenges to the equation of phoneme segments and articulatory segments. Chiefly, these empirical proofs showed that speech perception can occur in conditions that neutralize the contribution of production, as in cases of degenerate, immature, or absent productive capability. In one empirical falsification, infants too young to articulate speech were reported to perceive phoneme types (Eimas *et al.* 1971). In another, individuals who lacked the personal experience of articulation due to neural or functional incapacity remained capable perceivers of speech (Lenneberg 1967; MacNeilage *et al.* 1967). By resorting to direct means of examining the intrinsic physiology of speech production, it proved to be unimaginably difficult to identify traces of hypothetically discrete, commutable motoric ingredients in a coarticulated mixture, whether the examination focused on articulatory motion, anatomical configuration, or the motor–neural signals (MacNeilage 1970).

More recently, an original approach to motoric reduction has renewed the claim of equivalence of linguistic and articulatory types. Here, a theoretical breakthrough aimed to resolve the difference in structure between the abstract

linguistic segments—consonants and vowels—and the continuous, graded, and nonlinear articulation, recasting the linguistic description in a nonsegmental form (Goldstein and Fowler 2003). In this conceptualization, the phonemic properties that are expressed and recognized in speech are no more sequentially segmental than the components of articulation, relieving from the outset a prominent aspect of the discrepancy between symbolic and physical aspects of speech. In *articulatory phonology*, phoneme contrasts are conceptualized as potential shapes, placements, and phased actions of individual articulators. A word is represented phonemically in a gestural score describing the canonical form of production. By conceiving the phonology in discrete gestures that coincide temporally, the representation of linguistic properties falls into potential alignment with articulation. Lexical contrasts that had customarily received a segmental description in the grain of phonemes have been readily described as contrasts in an inventory of gestures. Under this description, a perceiver hears through the auditory effects of articulation to the graded and continuously progressing gestures that produce the sound as they mark the lexical constituents without additional elaboration or analysis. Once the shapes, placements, and phasing of the gestures are resolved, the perceiver achieves the linguistic grain requisite to identify words, at least insofar as canonical form of expression is concerned. Admittedly, an increase in realism is accomplished by designating a robust linguistic representation that inherently accommodates variation in articulatory rate as compression or expansion in the imbrication of gestures. Moreover, an emphasis on a linguistic description far closer to articulation than the traditional abstract phoneme potentially offers a natural means of explaining alternations, neutralization, assimilation, and other classic phonological phenomena. Instead of changing the symbolic composition of an expression, canonical gestural form persists, while gradient implementation of the score produces the variants.

As in the case of its predecessor, the Motor Theory, the empirical challenges to articulatory phonology are likely to determine the durability and range of its appeal. Of course, the evidence that had counted against the Motor Theory gets no traction against articulatory phonology simply because the more recent form of articulatory reduction claims no role for a faculty of production in perception. Instead, it uses productive types linguistically, to designate the phonemic form of words, and aims to relieve the mismatch in kind between the phonemic form regulated grammatically and the articulatory form responsible for expression. Nevertheless, the account seems vulnerable to challenges to the parity it asserts between canonical form and expressed form. For example, the new account predicts that the elevation of the tip of the tongue to touch the alveolar ridge of the palate behind the upper teeth

at the close of the words LINT and BUTT persists covertly when LINT BRUSH is pronounced LIMP BRUSH, or BUTT KICKER as BUCK KICKER. In these cases of regressive place assimilation, typical of many English speakers, old-fashioned segmental phonology considers the departure from canonical phonemic form to reflect the underspecification of the articulatory place of the class of coronal segments. In articulatory phonology, the canonical form of BUTT KICKER retains its composition, though, with a change in phase relative to the components of the constellation, the gesture of alveolar touch occurs during the palatal closure for /k/, starkly reducing its acoustic effects to the degree that the place of the prior consonant appears to have been assimilated to the latter. In the case of LINT and BRUSH, a phase shift of gestural components permits a labial closure to conceal the alveolar touch. Empirical studies have not yet decided whether the gesture predicted in the account occurs covertly, as if the linguistic object were indeed an articulatory gesture, or whether the gesture that appears to be assimilated is absent in fact, as if the change occurs at the level of linguistic form.

During the wait for the evidence to appear, it is useful to consider that articulation is seldom a result of canonical lexical form alone. Talkers can regulate the properties of expression secondary to the choice of words and the phonemes they comprise, as in instances of vocal mimicry or disguise, or in a modulation of style (Pardo and Remez 2006). When this involves substitution of a gesture, when ROPING and RIDING become ROPIN' and RIDIN', there is more than a simple shift in the relative phase of articulators at play. In a recent example of vocal play, a child was observed expressing a portion of the range available for expressive variation without altering the lexical message. In a stable pragmatic condition, plausibly commonplace, she taunted her father: [dæri] ... [diæri] ... [dædi] ... [dæ:ri] ... [dæ? di] ... [dæri] ... etc. Extrapolating from the language of articulatory phonology, we might say that the phonetic form related to the canonical phonemic form was edited on the run for expressive aim, although neither the lexical item nor the canonical phonemic form changed at all. The articulation followed the editing, and this unexceptional use is permitted by the principle that linguistic form and articulated form are related as the message is to its expression. They differ, and the symbolic representation of the expression cannot be reduced to its physical properties.

### 2.2 *Reduction to Acoustic and Auditory Types*

In a dialectic at large among researchers, a reduction of linguistic properties to motor components is typically weighed against reduction to auditory types, on the truism that speech must be heard to be understood. The germ of inspiration

for sensory reduction of symbolic linguistic objects was sown a century ago in the work of introspective psychologists attempting to catalog the elementary sensations in which perceptual experience is constituted. Variation in physical frequency, amplitude, and spectrum were correlated with the elementary sensory experience of pitch, loudness, and timbre, but additional properties, among them *vocality*, were identified (Köhler 1910). Even a pure tone evoked a vocal impression, according to this theme, which hindsight might view as unduly fanciful. However, were vocality to be primary, no account would need to derive audible phonetic qualities from simpler auditory sensations.

These days, the scene has shifted, and an explanation of speech perception appealing to auditory qualities uses the warhorses of psychological rationale, similarity, and likelihood, to make the case (Diehl *et al.* 2004; Mirman *et al.* 2006; Saffran 2003). The argument, put simply, is that individual phoneme segments are differentially correlated with specific acoustic properties. For a perceiver to be sensitive to the occurrence of a phoneme, a sensory function, albeit a tacit one, must tally the distribution of the precise sensory forms affiliated now and then with each of the phoneme types. The sensory forms associated with any phoneme are asserted to vary in likelihood as well as in auditory characteristics, with a gradient of similarity extending from the likeliest sensory form to less likely forms in this view of the orderly reduction of linguistic types to auditory norms. To recognize a phoneme, its auditory form—whether simple or complex—must be analyzed and held for comparison with the different remembered sensory gradients that compose a perceiver's experience of speech. Through an act of comparison of an unidentified auditory form with the remembered gradient of auditory correlates, a perceiver can estimate the likelihood that the present sensory experience is /p/ or /t/ or /k/, or none of these. The use of an actuarial method of recognition is often disguised in the garb of neural networks, although neither neural phenotypes nor actual exposure to spoken language are modeled in such enterprises.

Could a phoneme, a linguistic marker used to distinguish one word from another, be a sensory form? Could the perception of phonemes rest on normative characterizations of sounds? These questions about speech have been posed continually since the instrumental analysis of acoustic waves was possible, and the perspective is chronic if utterly implausible, despite a lack of encouragement of auditory sensory reduction from (1) physical acoustic analyses, (2) auditory physiological studies, (3) engineering projects, and (4) psycholinguistic research. Its survival expresses a forlorn wish despite a preponderance of disconfirming evidence.

The hypothesis is false. If the phonemes of English were physical acoustic properties sensed auditorily, and were this aspect of experience informed

by nothing more elaborate than a thorough actuarial practice that plotted distributions of differing likelihood, then typing on keyboards would have already been rendered obsolete by clever audio engineers and deft statisticians, in favor of vocal dictation. Sadly, these words are inscribed by a typing hand. The evidence was clear long ago that the acoustic properties of intelligible speech are not ordered normatively. Instead, it seems as if a logical function applies to sensory types including those well beyond the norm or even the physiologically possible, permitting the recognition of phonemes from auditory states that violate norms.

The claimed auditory definition of phonemes runs afoul of the hard evidence in several ways, whether the auditory component of perception is cast as a conduit faithful to the impinging spectra, or as a creative function in which criterial auditory sensory qualities are induced predictably from incipient speech. The most fundamental problem with this view is its presumption that auditory sensations are long-lasting, that is, persistent enough to permit comparison to stored distributions; and, both accurate and persistent enough to compose stored distributions.

In fact, very little of the auditory sensory effect of speech is left 100 ms after a wavefront strikes an eardrum, and nothing is left in 400 ms (Howell and Darwin 1977; Pisoni 1973). Human auditory experience is neither durably nor accurately remembered. These conclusions stem from studies of sensory acuity in which the ability to distinguish subtly different *discriminanda* was found to be inversely proportional to the interval occurring between them. In a test of this type, a listener is asked to distinguish the presentation of two physically identical sounds from two sounds of identical type differing in auditory form. When the interval between the two sounds is brief, discrimination performance can be based on a comparison of the auditory quality of the items as well as their perceived classification. As the interval grows in duration, the rapidly fading sensory trace of each item becomes less well resolved, and, therefore, less available for comparison; discrimination performance grows correspondingly poorer because at long lags it is based largely on the classification of the items, which is all that survives the lag. It is commonplace to encounter items that are discriminable at a brief lag of 50 ms yet not at a lag of 200 ms. In less than a quarter of a second, the sensory qualities are gone, leaving only the remembered experience to inform the act of discrimination. This distinction between fleeting sensory qualities and far more durable classified memory applies to sounds other than speech. Indeed, auditory quality is readily encoded as a mechanical description approximate to the physical cause of a sound (Hirsh 1988; Lakatos *et al.* 1997). Although it is possible to conjure a sound by thinking of it while the world is silent, it is implausible to suppose that this

experience occurs by recalling an exact trace of a waveform once heard. More likely, the experience is generated by way of a cognitive approximation to the mechanical causes of sound.

With respect to the components of an utterance, auditory persistence is apparently worse for consonants than for vowels; however, overall, this aspect of the early sensory experience of speech is just too ephemeral to be useful for more than initiating perception. Urgency imposed by the rapidly fading auditory sensory trace of speech warrants a prompt projection into a relatively more stable phonetic or phonemic code.

Auditory accounts of phonemes have presumed that the acoustic constituents of speech spectra and the elementary auditory experiences concurrent with speech perception are finite in variety, and that the acoustic and auditory elements that promote the perception of speech form a class. If the acoustic ingredients of speech were unique to speech, this premise might be true. Yet, the whistles, clicks, hisses, buzzes, and hums that are taken to be speech attain that status by virtue of the configuration that they compose, and not because of their characteristics considered in isolation (Remez 2005). Produced alone or extracted from a speech stream, an acoustic element of speech evokes an auditory quality without an accompanying phonetic impression. In fact, the aspect of the acoustic stream that appears critical for eliciting phonetic perception is its time-varying pattern, and not the elements that compose the pattern. A perceiver even tolerates the presence of impossible acoustic elements if the spectral configuration is speechlike, and such findings ultimately falsify the claim that perception is achieved by isolated acoustic or auditory elements because of their relation to the norms of experience. Perceptual functions apparently track the causal conditions of the incident sensory forms, and can be indifferent to the sensory details that auditory reduction places at the center of the account.

Several lines of evidence converge here. In one, speech perception can be evoked by a pattern composed of a small number of pure tones. None of these can be produced by a human talker, but the pattern imposed on the frequency and amplitude variation of the tones is derived from a speech spectrum over time, and perceivers report the linguistic properties of the speech that served as the model for the tone complex; the auditory qualities are described as unnatural for a voice (Remez *et al.* 1994). In another technique, a speech spectrum is analyzed into bands roughly 1 kHz wide, and all of the detail within the band is merged. A single integrated amplitude value is derived each moment in each band, and this is used to set the power of a noise source matched in center frequency and bandwidth to the analyzing band. The result is a signal composed of a small number of wide-band noise sources changing

amplitude according to the speech signal on which it is based, but without any of the acoustic elements of the original. No human vocal tract is capable of producing such a pattern, and this noise-band derivative of speech differs spectrotemporally from whispered speech in coarse and fine grain alike. This signal is also intelligible, despite an impression of a raspy voice (Shannon *et al.* 1995). In a chimerical variant of the noise-band method, the analysis is similar, though the source can be chosen freely rather than being noise. If the source is a recording of a construction site, the elements composing the resulting pattern are produced by a saw, a cement mixer, bulldozers, backhoes, shovels, loaders, fork lifts, and cranes. If the source is a jazz band, then the elements composing the pattern are trumpets, trombones, saxophones, piano, bass, and drums. The impression reported by the listeners is of the sentence from which the analysis is made, spoken by the sources, despite the impossibility of this event (Smith *et al.* 2002). Neither construction sites nor jazz bands talk, yet this is a listener's impression.

The claim that speech perception depends on resolving the normal and familiar auditory forms of phonemes is also problematic considered most broadly. In this case, the proof is provided by the speech perception prowess of the deaf. For many deaf individuals, speech reading, the visual perception of linguistic form in the absence of hearing, is quite successful, despite an utter lack of auditory sensory properties to drive perception (Bernstein *et al.* 2000). Some deaf perceivers whose speech reading is far short of perfect succeed with residual hearing despite sensitivity limited to low frequencies. Alone, such limited auditory perception is inadequate for perceiving speech, though in conjunction with vision a rather thin supply of auditory samples of speech is useful. Perhaps most remarkable is the use of an electrocochlear prosthesis to evoke auditory experience. An adult whose competence in language antedates the onset of deafness is often able to rely on auditory experience evoked in this unusual way; young children deafened before they attain linguistic ability have been observed to develop competence in perception and in production of language with this atypically caused auditory sensation (Svirsky *et al.* 2000). Once in place, an electrode causes effects that differ from ordinary auditory function, of course, but not only in the means by which neural activity is excited. For one, the inevitable spread of current along the cochlea creates a kind of frequency blur; a frequency difference as great as the interval of a minor third (the difference between the first two ascending notes of 'Greensleeves', or descending, of the 'Star Spangled Banner') is not resolvable (McDermott 2004). For another, the electrode does not penetrate to the apex of the cochlea, and is never proximate to the basilar region of lowest frequency sensitivity. This causes an upward shift of the experienced frequencies, perceptually.

Implant users who were adventitiously deafened report an impression of munchkin speech, likening the initial experience of quality elicited with the electrode to the high-pitched reedy voices of the little people in the movie, *The Wizard of Oz* (Shannon, personal correspondence). Over time, listeners say that they adapt to the odd auditory qualities of speech conveyed via an implanted electrode, though we can be confident that very little of the melodious lilt of speech is available to perception; those auditory properties are stripped from experience by the coarseness of the electrical transduction of the signal.

A tolerance of acoustic elements that transcend normal experience is actually reasonable, and must not be taken to indicate an inherent perversity in the means by which perceptual knowledge of language occurs. In the nursery and on the boulevard, the variation in physical manifestations of spoken language is vast. Anatomically, talkers range from large adult males to small children, and the variation in scale is accompanied by concomitant variation in the acoustic correlates of production. Variation in dentition and soft-tissues that affect sound production is also commonly encountered in ordinary conditions, as are the occasional individuals suffering laryngitis or rhinitis. Speech can be shouted or growled, spoken with a *Partagas Robusto* clenched between the teeth or spoken with edentulous gums. The distorting effects of telephones and walkie-talkies are also familiar, and speech perception survives these conditions that prohibit the veridical transmission of the acoustic products of natural vocalization. The origin of this forgiving standard of articulatory, acoustic, or auditory constituents of speech is not understood, although evidence of its action is well documented. The qualitative experience of speech depends critically on the exact nature of the acoustic elements in the signal, but the linguistic form is apprehended by following the changing pattern composed by the elements, and not the elements themselves.

The freedom from fixity of the phonetic and acoustic details that express the phonemes is inherent in the nature of the contrast on which linguistic markers depend (Pardo and Remez 2006). The key is the use of distinctive oppositions. Talker and listener alike share the small set of phonemes, and though these constituents index words, neither production nor perception must be faithful in producing the segments in their canonical form. A talker must merely indicate at any juncture which of the possible phoneme contrasts is intended, with the understanding that the listener tracking the production has the same sense of possibility in mind, a consequence of experience with spoken language. This freedom to indicate rather than to replicate the articulatory choreography specific to a word licenses a talker's use of a wide expressive range; it also permits a listener to find the auditory and phonetic attributes that compose

the linguistic form of speech without a commitment to a single set of physical markers, regardless of their dimension.

### 2.3 *Prospects of Reduction*

Could reduction of linguistic phonemes to motoric or auditory elements succeed if researchers identified physical correlates of these linguistic markers with greater success? One way to hazard an answer is to acknowledge that the linguistic functions of phonemes are real, and that the prospect of reduction accordingly poor. It might be possible to correlate phoneme incidence with motor, auditory, neural, and even genetically controlled cellular phenotypes without reducing the linguistic function to its correlates. Practically, the only serious prospect available for eliminating the typist's keyboard in favor of a computer that takes dictation is to understand the conditions that allow a talker to depart from canonical articulatory form and its acoustic effects, for instance, and the dimensions of permissible departure. Likewise, there is empirical need to characterize the conditions in which a spectrotemporal aberration in a sound stream can be attributed to the talker's attempt to retain a piece of food in the mouth while articulating a poignant message. But in no case can the linguistic function of the phoneme be subsumed in the functions of articulation and auditory resolution, functions so often targeted for reductionist description appealing to the physiology of the vocal tract and of the cortex, respectively. Rather, the opposite is true—the linguistic functions of phoneme contrasts subsume their expression.

## 3. The Seduction of Psychology by Phenomenology and the Impression of Transparency

Phenomenological analysis begins innocently enough. We take the enormity of visual experience and examine its discernible components. The visual experience of the Taj Mahal presents it as having a color, a lightness, a shape, a texture, a location, and a duration. The auditory experience of Fido's bark has loudness, pitch (or at least a dominant one), timbre (a distinctive quality), heading, range, and duration.

No method is without presumption, and methodologies used in the philosophy of perception are not innocent. After what seems a brief flirtation with naturalism in theories of mental content, contemporary philosophy of mind

is once again firmly fixed on folk explorations in the phenomenology of perception.

Although contemporary philosophy of perception makes room for many different conclusions about the nature and limits of perception, its method of inquiry, in broad outline, continues the dominant legacy in the history of philosophy: We uncover the nature of perception first by describing perceptual experience, and we gain access to that perceptual experience by introspection. We peer into or focus attention on inner experience. So the procedure is utterly intuitive. We generate a first-person report of sensory experience and then introspect its phenomenal character. Contemporary treatises in the philosophy of mind offer a digest of how this process works:

When we introspect our experiences and feelings, we become aware of what it is like for us to undergo them. But we are not directly aware of those experiences and feelings; nor are we directly aware of any of their qualities. The qualities to which we have direct access are the external ones, the qualities that, if they are qualities of anything, are qualities of external things. By being aware of these qualities, we are aware of phenomenal character. (Tye 2000: 51)

But the philosophical work in this vein is not merely descriptive, it is also didactic; without acknowledging the possibility of intersubjective disagreement in what we find when we look inward, it supplies the description of the experience of human perceivers in the process of sensing and introspecting:

If you are attending to how things *look* to you, as opposed to how they are independent of how they look, you are bringing to bear *your* faculty of introspection. But in doing so, you are not aware of any inner object or thing. The only objects of which you are aware are the external ones making up the scene before your eyes. Nor, to repeat, are you directly aware of any qualities of your experience. Your experience is thus transparent to you. But when you introspect, you are certainly aware of the phenomenal character of your visual experience. On the basis of introspection, you know what it is like for you visually on the given occasion. Via introspection, you are directly aware of a range of qualities that you experience as being qualities of surfaces at varying distances away and orientations *and thereby* you are aware of the phenomenal character of your experience. (Tye 2000: 47; emphasis in original)

Patently, awareness of phenomenal character is not a quasi-scanning process. Our attention goes *outside* in the visual case, for example, not to the experience *inside* our heads. We attend to one thing—the external surfaces and qualities—and yet *thereby* we are aware of something else, the ‘feel’ of our experience. Awareness of that ‘feel’ is not direct awareness *of* a quality of the experience. It is awareness that is based upon direct awareness of external qualities without any inference or reasoning being involved. (Tye 2000: 51–2; emphasis in original)

Or, to take another example, this time of an inference to a metaphysical conclusion from phenomenological considerations:

Dispositions lie strictly outside of what is immediately perceptually presented, but colors figure in the very pith of perceptual presentation. To use a much-maligned term, colors are *given*, while dispositions are *posited*. It follows that, if colors were really dispositions, they would not be visible in the way they are. To be sure, there are dispositions associated with colors, and perhaps even inferable from them, but these dispositions are not given *in* the color. (McGinn 1996: 540)

Ordinarily, a perceptual function is nondemonstratively inferential if the ultimate structural description of the object is underdetermined by the sensory evidence (Fodor 1983). In the passage above, it is presumed that what is *given* and what is *inferred* would find consensus in any group whose members honestly reflected on the matter. Interestingly, Berkeley (1709) expressed the same confidence in *A New Theory of Vision*. And he used this premise to argue, quite falsely, that the outputs of the visual process were not arrived at by nondemonstrative inference.

There is the hope, and expectation, that an elaborate, and perhaps even technical, body of consequences can be extracted from the modest observation that phenomenal character conveys what it is like to have an experience. That expectation is implicit in the approach so widely taken in the study of phenomenal character:

Visual experiences have phenomenal character, or more simply a phenomenology. The phenomenal character of a visual experience is what it is like to have that visual experience. In general, I will say that events of sensing, such as seeing, have a sensory phenomenology. (Siegel 2006a: 484)

The scientific study of sensation and perception may have some relevance to philosophical inquiry, but this inquiry quickly and easily erects barriers to scientific improvement and criticism of philosophical analysis: ‘If the argument here is sound, our perceptual systems may include modular “input systems” of the sort described by Fodor (1983), but these systems will not be ones with which visual *phenomenology* is exclusively associated’ (Siegel 2006a: 501). Our perceptual systems may be modular, but the phenomenal features of their output require no allegiance to the scientific doctrine of modularity, or, for that matter, any scientific doctrine whatever. Indeed, it may be difficult to tell, from such passages, the place of the philosophy of perception within intellectual inquiry about the mind. More important, it is difficult to tell whether the methods and findings of science have any role at all to play in addressing evidently empirical claims that philosophers make about the phenomenal character of their experience. For example, just how far-reaching

is the scope of this inquiry? Can it only inform us about that part of the world that is phenomenally accessible? How does this method handle the voluminous findings that first-person reports, as well as the process of introspection, can be systematically unreliable? Are the tools of introspection so blunt that this research program, such as it is, can only tell us about the most basic distinctions in the field? Are our concepts so crude, so folk-bound, that they could not serve a taxonomic function in any project but the conceptual, first-philosophical one? And finally, what happens when intuition and scientific finding collide, as they do so often?

The phenomenological project appears to be descriptive; it attempts to detail, through introspection and report, the nature of perceptual experience, along with the mechanisms, processes, states, and properties underlying it. But it is also explanatory. It attempts to account for features of perceptual experience, once again through introspection and first-person report. Finally, the processes described in contemporary phenomenological analysis are not restricted to those with the technical vocabulary to discuss it; those processes are thought to be universal. Its characteristic treatment is that important aspects of our phenomenal experience are species-general. There is a way that phenomenology is for all perceptual experience of objects. As one philosopher puts it:

The conclusion of this paper is that there is a phenomenological constraint on object-seeing. There is a specific sort of visual phenomenology that perceivers must have to see objects, and it is that specific sort of phenomenology that plays a role in making the situation one in which the perceiver sees a particular ordinary object *o*, as opposed to seeing no object at all. (Siegel 2006b: 432–3)

If commitments to transparency or intuitiveness underlie such philosophical analyses, it is difficult to get too enthusiastic about the real scope or reach of philosophy into psychology. After all, where is the evidence that putatively transparent assertions are actually accurate? Where is the evidence that the taxonomic philosophical categories of perceptual experience designated in this contemporary work are found in the assertions of psychologists of perceptual experience, or in the vernacular perceptual assertions of the citizens of the Republic? It is one thing to get swept up in a project of conceptual analysis, quite another to forget the humble roots of the inquiry, the limited tools for excavation, and the culturally local body of evidence addressed.

Psychologists have not been similarly drawn in. Indeed, psychologists put scant stock in the transparency of the *contents* of perception, and even less in the power of introspection to isolate and characterize perceptual functions. By contrast, the philosophy of perception, as currently practiced, often assumes

that useful, robust, generalizable—even nomic—information can be harvested from complex sources and the percepts they prompt. This assumption proved false in physics and biology, and much of 20th century psychology has been devoted to unmasking the conceits at the foundations of judgment, problem solving, memory, and self-assessment. It is precisely these conceits on which philosophical method has relied.

In the end, it is not clear what these phenomenological analyses are supposed to show. People who use them seem to believe that they reveal more than just the trenchant intuitions cultivated by training in analytic philosophy at a particular moment in history. Yet, the standard philosophical practice of giving epistemic primacy to intuitions and their ‘refinement’ in reflective equilibrium has been roundly criticized in the areas of moral theory (Doris 2002) and standard analytic epistemology (Weinberg et al. 2001; Bishop and Trout 2000a, 2000b). This critical work explains how the justification of epistemic and moral claims has a structure embarrassingly similar to the basis of any ideology. As it is practiced in the English-speaking world, epistemology and moral theory reflect the favored intuitions of the West. But these intuitions and practises are not necessarily compatible with our best sciences, applicable to a wide range of human populations, or associated with documented success. If cultures of the East are less disposed to the intuition that knowledge is justified true belief, which experiments have revealed, why should we expect our analytic, Western, first-person reports and intuitions about the phenomenal character of experience to be superior? They might prove to be, but we need to see the argument and evidence. Without it, the claim to its superior reliability would be cultural prejudice of the crudest sort. Notice, for example, that the enterprise of standard analytic phenomenology includes no attempts to address similar appeals to transparency, or to phenomenality, that occur in the philosophical texts of other cultures, for example, Nyāya Sūtras.<sup>3</sup>

<sup>3</sup> From the Nyāya Sūtras of Gotama:

I.1.4 Perception is that knowledge which arises from the contact of a sense with its object, and which is determinate [well-defined], unnameable [not expressible in words], and non-erratic. (Radhakrishnan and Moore 1957: 359, brackets in original)

I.1.16 The mark of the mind is that there do not arise (in the self) more acts of knowledge than one at a time. (360)

Vātsyāyāna’s commentary (also in Radhakrishnan and Moore) on I.1.16 explains the above:

... even though at one and the same time several perceptible objects... are in close proximity to the respective perceptive sense-organs, ... yet there is no simultaneous cognition of them; for from this we infer that there is some other cause [namely, the mind], by whose proximity cognition appears ... If the proximity of sense-organs to their objects, by themselves, independently of the contact of the mind,

This is not to say that analytic phenomenology of perception is *necessarily* compromised in the same way that much moral theory and standard analytic epistemology are. Our cognitive intuitions about our perceptual outputs may be different in just the way they need to be in order to escape this charge. But that case needs to be made.

In order to rely legitimately on first-person reports and on intuitions, we need to have evidence of their reliability beyond the fact that we can generate the reports and that we have the intuitions. They must actually be reliable, not just feel reliable. In particular, we need to have evidence that first-person reports are correct, and that they are general. With no mention of whether the same individual might respect the same constraint over time, or that the constraint is honored in the conduct of different people, we have little warrant for trusting the generalizations of phenomenal analyses. Securing this evidence would require an empirical enterprise, one that tracked the accuracy of reports and intuitions, and tested them against the reports and intuitions of individuals with very different philosophical attachments. Practitioners could respond to this rather obvious challenge by examining the strength of their methods and the scope of their conclusions. Another, complementary path would lead to experimentation on different populations. What do the people of Toraja report when asked to examine the separability of dimensions in perceptual outputs? What do Mongols say? What does Oprah say?

Two decades ago, philosophers commonly contrasted perceptual appearances with realities. Through careful experimentation, researchers uncovered underlying functions that promote the resolution of the attributes of objects and events. Some are anisotropies of sensitivity that ordinarily escape notice, though scrutiny under rarefied conditions permits more or less faithful subjective impressions. Others are less available—for instance, an experience of perceptual priming. In this experimental method, the occurrence of one event affects sensitivity to a second, independent event, as if the first event altered the threshold of detection of the second. Admittedly, this is an intuitively plausible phenomenon in olfactory sensitivity, for example, in which exposure to lemon scent immediately increases the concentration of orange scent

were the sole cause of cognitions, then it would be quite possible for several cognitions to appear simultaneously. (359, brackets, ellipses in original)

And Vātsyāyāna's commentary on section II.1.31 sheds further light:

When the observer cognises the tree, what he actually perceives is only its part nearest to himself; and certainly that one part is not the 'tree'. So that (when the man cognises the 'tree' as a whole) what happens is that there is an inference of it (from the perception of its one part), just like the inference of fire from the apprehension of smoke. (366)

required for detection. Gradually, the baseline sensitivity returns. In auditory word recognition, priming is not accompanied by veridical subjective states at all. A spoken word will briefly lower the threshold for detecting other words to which it is related. Naturally, a semantic relationship is evident, in which the presentation of DOCTOR facilitates the recognition of NURSE and BREAD facilitates BUTTER; yet, neither does DOCTOR facilitate the recognition of BUTTER nor does BREAD facilitate the recognition of NURSE, indicating one requirement for priming—that is, semantic relatedness.

If this dimension of relation has the ring of plausibility, it is hardly the intuitive experience of a perceiver hearing a sentence to be flooded with impressions of the semantic relatives of the resolved words. Imagine the experience of a listener hearing mother dictate, ‘Eat a little something!’ decomposed perhaps as ‘Absorb attack bite bolt chew devour digest dine ingest gorge nibble snack sup wolf a little tiny infinitesimal miniscule paltry insignificant something object commodity substance thing’, and this example is limited, falsely, to synonymy. Additionally, a spoken word also facilitates the identification of words to which it is related only in phonemic form. The occurrence of the word SHEEP facilitates the similar words SHEET, CHIC, CHEAP, JEEP, SEEP, SEAT, SEEK, SHIP, CHIP, SHEAF, SHE, SEE, and more. None of these functions is phenomenally accessible. We can only introspect the outputs that are phenomenally accessible. The intermediate values derived in these functions do not form the ingredients of the cognitive capabilities of storage and report, producing a kind of opacity of ordinary perceptual functions and contents (Fodor 1983).

These findings have exerted a powerful influence on psychological and physiological research on the senses, but almost none on philosophical analyses of perception. Moreover, the unobservable causes of our phenomenal experience would seem explanatorily relevant, because experiences, like any complex taxonomic item, can be individuated not just by what they cause but by what causes them. Despite the obvious relevance of scientific psychology to an intellectually responsible phenomenological analysis, philosophical practitioners unburden themselves of science by insisting that they are interested specifically in the study of *experience*—the introspectible output of our perceptual systems—and so presumably not concerned with its theoretical causes or effects.

This is not to say that phenomenology cannot be part of empirical psychology,<sup>4</sup> but in empirical psychology the phenomenal character of experience has a far more modest and settled role. It is useful to begin with questions about

<sup>4</sup> On the contrary, good examples can be found in Noë *et al.* (2000), and Noë (2002).

how we attribute properties to objects, and how we represent those properties. But this starting place is tentative and utterly defeasible. And if we take these analyses any farther, we begin to work the philosophical theses beyond the evidence we have. We might say, for example, that our experience of color has both representational and nonrepresentational features. Our evidence might consist in examples of each, or what we would say about each case. How are we to know that a given visual experience has both representational and nonrepresentational features? Because we seem to detect this difference when we apply our first-person attention to the object.

#### 4. Transparency

One of the most seductive features of phenomenological analyses of perceptual experience is the conviction that important aspects of perceptual output are transparent—that, no matter how short the lifetime of our sensory impressions, or how faded their traces, there is something decisive of the content that can be harvested. The perceptual material uncovered by phenomenological analysis—shape, color, timbre, bitterness, heat, alliacousness—is patently transparent in a certain respect. When their intensity reaches human thresholds, their presence is open to casual inspection. The fact that they are present is introspectible.

Different accounts of transparency have been offered, but for our purposes the precise account does not matter. Whatever the account, the transparency of mental contents is an honored thesis in the history of philosophy. And, the integrity of phenomenological analysis presupposes *some* kind of transparency. Otherwise, there would be nothing distinctively phenomenal about the examination. Instead, the process would be a purely conceptual theoretical exercise.

Whether out of arrogance or ignorance of the empirical evidence to the contrary, a substantial class of philosophers subscribe to something like an official theory, to the effect that we have privileged access to the content of our own minds. This tradition is nicely summarized by Gilbert Ryle:

[A]ccording to the official theory, a person has direct knowledge of the best imaginable kind of the workings of his own mind. Mental states and processes are (or are normally) conscious states and processes, and the consciousness that irradiates them can engender no illusions and leaves the door open for no doubts. A person's present thinkings, feelings, and willings, his perceivings, remembering, and imaginings are intrinsically 'phosphorescent'; their existence and their nature are inevitably betrayed to their owner. (Ryle 1949: 154)

Descartes famously thought that he knew what he perceived and believed simply by looking inward, as it were, by immediately observing the contents of his mind. As Descartes (1637) put the point: ‘nothing can be in me, that is to say, in my mind, of which I am not aware’. Some commentators on Descartes hold that he restricted his transparency thesis to *occurrent* mental states, but we shall not digress to the vicissitudes of Descartes’s exegesis; our focus is on positions, not figures. In Descartes’s corpus, the transparency thesis takes the form that ‘there can be nothing in me, that is in my mind, of which I am not conscious’ (Wilson 1981: 98).

In modern philosophy, the attachment to the transparency thesis flowed deeply beneath the surface contours that otherwise divided the landscape of early modern philosophy. Along with the rationalist Descartes, the empiricists Locke, Berkeley, and Hume all embraced versions of the transparency thesis.

Asking the reader to use introspection as a test, Locke announced that ‘a man cannot conceive himself capable of a greater certainty than to know that any idea in his mind is such as he perceives it to be; and that two ideas, wherein he perceives a difference, are different and are not precisely the same’ (Locke 1689). Hume, advertising a different putative benefit of transparency, consults introspection to justify the incorrigibility of transparent mental contents:

For since all actions and sensations of the mind are known to us by consciousness, they must necessarily appear in every particular what they are, and be what they appear. Everything that enters the mind, being in *reality* as the perception, tis impossible anything should to *feeling* appear different. This were to suppose that even where we are most intimately conscious, we might be mistaken. (Hume 1739)

Bishop Berkeley’s arguments for transparency may be the richest philosophically—they are certainly the timeliest—because they make the very connection between transparency and the non-inferential nature of perception that modern psychologists and some philosophers resist:

But those *lines* and *angles*, by means whereof *mathematicians* pretend to explain the perception of distance, are themselves not all perceived, nor are they, in truth, ever thought of by those unskillful in optics. I appeal to any one’s experience, whether, upon sight of an *object*, he compute its distance by the bigness of the *angle* made by the meeting of the two *optic axes*? Or whether he ever think of the greater or lesser divergency of the rays, which arrive from any point to his *pupil*? Nay, whether it be not perfectly impossible for him to perceive by sense the various angles wherewith the rays, according to their greater or lesser divergence, do fall on his eye. Every one is himself the best judge of what he perceives, and what not. In vain shall all the *mathematicians*

in the world tell me, that I perceive certain *lines* and *angles* which introduce into my mind the various *ideas* of *distance*; so long as I myself am conscious of no such thing. (Berkeley 1709: sect. 12; emphasis in original)

He attempts to nail down that connection by declaring, 'Since, therefore, those *angles* and *lines* are not themselves perceived by sight, it follows from Sect. X., that the mind does not by them judge of the distance of *objects*' (Berkeley 1709: 16).

The father of geometry, Euclid, had an 'emission theory' of vision: Rays of light are emitted from the eye, and objects become visible when they 'catch the rays' of the eye. In this way, vision was conceived as a species of touch. Al-Kindi, the first great philosopher of the Islamic world, resuscitated Euclid's view. He urged others, 'not to be ashamed to acknowledge truth and to assimilate it from whatever source it comes to us, even if it is brought to us by former generations and foreign peoples' (Lindberg 1971: 469). To overcome the obstacles of time and culture, the emission theory had to have considerable charm. And it did. But notice that the mechanics of visual perception were no more transparent to Al-Kindi than they were to Berkeley. If they had been, introspection alone could have revealed which of these theories of vision would have been recognized as correct.

Consider an example in the philosophical analysis of visual experience, in the service of establishing the philosophical thesis of representationalism. This thesis states that the phenomenal character of experience should be explained by its status as a species of mental representation. Philosophers have defended representationalism by appealing to the transparency of visual experience to first-person acts of attention. According to this defense, our awareness of an object passes through our experience directly to the object we are attending to. As a result, we cannot focus on the intrinsic features of our visual experience. Our attempts to do so are like trying to handle a wet fish, slipping seamlessly from the experience of blue to the object's properties of blueness or squareness.

These arguments work the apparent transparency of phenomenal processes. Some distinctly philosophical projects infer the epistemic property of incorrigibility from transparency. Other such projects infer models of psychological organization conditioned on the accuracy of our first-person access. After all, if our visual experiences are transparent, then self-observation can go proxy for experimentation. Reliance on transparency raises basic scientific and philosophical questions about the *scope* of a foundational program that attempts to analyze the constituents of perceptual outputs, the more basic elements of perceptual *experience*. If philosophical analysis proceeds with introspective tools and perceptual impressions that are culturally specific, the philosopher can do

little more than armchair conceptual anthropology, cataloging and charting the perceptual experience of people similarly cultivated. This is true even when the philosopher reclining at the helm is ingenious, motivated, and equipped with specialized maps and terminology.

This view certainly brings to earth the lofty goals classically associated with certain kinds of philosophical inquiry about the mind. With its domain pared down in this way, can philosophers even achieve the goal of the psychological cataloging and charting? After all, we do seem to have access to some features of our global perceptual outputs—the appearance of a tree, or the sound of a cat’s meow. Is phenomenology a fit tool for the analysis of these complex contents? In the first place, once sensory flow evokes an ultimate perceptual effect, we are left with the constructive process of *interpreting* that effect. And that has been the job of phenomenology. But interpretation of this sort is notoriously cognitively penetrable, influenced by our culturally imbued beliefs and desires. Careful philosophical analysis of scientific findings can be invaluable. But when scientific material is ignored, or replaced by intuitive material, we are left to ask whether phenomenological analysis is really just a kind of ethno-phenomenology, the anthropology of academic introspection of the West.

These cautions are timely. Because philosophical investigations have returned to phenomenological analysis, we are left to wonder whether this marks a principled reaction to the inadequacies of a naturalistic approach, or the untethered drift of interest we find in research programs in fields like literary theory. If the latter, then recent interest in old-style conceptual analysis may be largely sociological in origin. By itself, this issue is not important. It becomes significant only when philosophy begins to map the terrain of other disciplines, disguising a normative view of legitimate areas of study in a descriptive philosophical vocabulary. This occurs when philosophers extend folk phenomenology to related questions in perception. Is perceptual psychology reducible to biology? What are the constituent functions in the causal chain linking our percepts to their objects? These connections cannot be introspected. Perhaps the contents of experience, our perceptual outputs, can. But these outputs are typically multiply interpretable, and often these reports vary with culture.

In any approach that individuates properties not just by what they cause but what causes them, phenomenological analysis *just is* folk exploration into perceptual functions, not simply, if at all, perceptual experience. Disconnected from a science about the source, phenomenological analysis assumes that our experiences are classically well-defined; they assume a kind of narrow psychological essentialism.

## 5. The Official View: Psychological Essentialism, Old-Style

If philosophical methodology is not innocent, then different methodologies may drive us to distinct conclusions. This is a point pressed in much contemporary philosophical discussion of heuristics and biases, particularly in approaches to the problems in the philosophy of mind and biology (see especially Wimsatt 2006). For example, conventional wisdom and some enlightened theories have supposed that properties essential to species identity or, for that matter, the identity of a phoneme, have necessary and sufficient conditions, though cladistic approaches have offered an alternative account of species-membership. We may have had purely pragmatic or instrumental reasons for applying these standards. But applying them may lead more surely to reductionist viewpoints than other methods.

Phenomenological states, for example, are not individuated causally. They are individuated by their internal properties, by their sensory qualities, or by their seemings. Change the world as you like, our phenomenological states may remain unchanged. Or so we imagine (Trout 2001). Characterizing these properties as internal to the subject is hardly novel. But the observation can be worth repeating because it has been so regularly ignored. If psychological explanations were driven by seemings, those explanations would be internalist. They occur within the skin, and are presumed to supervene on their physiology.

These kinds of biases arise again in more scientific endeavors. True to our prediction, grouping or taxonomic schemes for speech sounds that are phenomenologically based lead to reductionist and old-style essentialist accounts of the objects of perception. On this view, phonetic classes necessarily or essentially possessed an acoustic invariant, and there was thought to be a one-to-one mapping of taxonomic speech units onto reduction bases like either vocal gestures, intended gestures, or acoustic classes (Appelbaum 1999, 2004). This essentialist fiction gets a foothold from the phenomenal character of a speech unit presented in isolation, but why has it flourished?

### 5.1 *The Doctrine*

If you consult Ladefoged's (2005) classic resource for phonetics, we find a catalog of phonetic segments. These segments can be arrived at by comparing two sequences in a minimal pair, such as MAT and MAD. This list treats each segment as though it belongs to a class that is distinguished by an acoustically invariant feature. Each phonetic segment is an island; no part of any segment is part of another. This impression is abetted by the fact that the minimal-pair

difference that defines each class is phenomenologically accessible: You can hear the difference between the [t] and the [d] in MAT/MAD. But hearing the difference is not the same as identifying distinctive features, because you can do the former without doing the latter. However, this does not change appearances. From the appearance that phonetic segments are assembled serially /m/ /æ/ /t/, people infer that they are spoken serially, with clearly identified boundaries.

But speech is not produced in that way. Instead, the vocal gestures that produce these features occur simultaneously; they are coarticulated. For example, the /m/ in MAD is produced by simultaneously closing and releasing the lips with the velum lowered while the larynx is buzzing; to produce /æ/, the blade of the tongue is lowered in the front of the mouth and the root of the tongue is drawn forward in the laryngopharynx, again, while laryngeal action produces the buzz of phonation. The lowering of the jaw required for the vowel cannot occur until the labial closure and release required for the /m/ are accomplished, but the tongue shape appropriate for the vowel can be set well in advance of the labial release and the lowering of the jaw—this description of the concurrent production of subjectively sequential segments is typical of articulation. This general character of articulation is unlike the subjective impression of linear concatenation of separate segments in the language's phonology (Fant 1962; Liberman *et al.* 1967).

### 5.2 *Underlying Functions and Psychological Opacity*

Perceptual inference is a complex matter. In the case of sound perception, we saw that there are levels of coding involved as the sound percept evolves from the sensory periphery to neural centers, and then connects to the cortical territory that guides our motor functions. These are codes that have a chemical and biochemical vocabulary, but not one that we think in, or ones that we speak. They do not have the properties necessary for our transparent experience of them. They are not accessible to the mechanisms of storage and report. We cannot make them the objects of our attention.

Berkeley's account of transparency, then, presupposed a now discredited view of this coding process. Given the psychological opacity of most of perception—even aspects of its output—we cannot infer that a psychological function does not exist just because we are unable to detect it upon introspection or reflection. While the absence of a psychological function could explain our failure to detect it, there might be other explanations. If we were able to peer into any of our perceptual functions at any stage, at any time we liked, we might delay their appointed rounds, and so slow their operation tremendously.

But a good part of the value of nontransparency resides in the fluency of perception it captures. Perceptual functions are shielded from the meddling of attention and memory, so that they can discharge their action-guiding duties, sometimes with life-saving quickness.

The level of function, and the unavailability to awareness of its details, makes the intermediate qualities opaque to introspection.

## 6. Conclusion

If explanations for spoken language understanding do not proceed by appeal to well-defined phenomenal attributes or to biological features of talkers and listeners, what items are implicated? Most scientific explanations account for effects by invoking natural kinds—classes of objects that play a taxonomic, counterfactual-supporting role in explanation. Open sets of properties, processes, states, and objects can endure over time even though they depend on a changing environment. Homeostasis offers a way of explaining the stable covariation of these properties. Animals, for example, are designed to maintain certain internal states, such as a body temperature range. With these features in place, the stability may invoke an even wider circle of properties. Consider a group of properties associated with birds: a high metabolic rate, feathers, hollow bones, nest building, and flight. There are birds that lack one or another of these characteristics. But homeostatic property clusters have the effect of favoring the presence of some properties, and by doing so erect barriers to the reception or development of others (Boyd 1999).

Put generally, in a homeostatic kind, no single property is a necessary condition for membership in the kind. If the homeostatic account accurately depicts the relationship among the properties in a natural kind at least in biology, psychology, and the social sciences, then we have an explanation for the field's resistance to reduction. At a minimum, we must abandon old-style essentialism. Our foregoing discussion of phenomenal attributes in speech perception stands as a disconfirming instance of the claim that explanatory kinds are classically well defined, at least in one area of psychology, and as a confirming instance of a homeostatic view of natural kinds in that field.

If our analysis is fair, then phonemic classes are not defined by necessary and sufficient conditions—not by gestural components, auditory qualities, or phenomenal characteristics. They are defined only by their role as contrast markers distinguishing words within a language. As we argued in Section 2, while their phenomenal properties may suggest as much, an utterance is not

a set of articulator shapes and motions, nor sound segments strung together like commutable beads. The conditions for understanding an utterance tolerate huge variation in the structure of a vocal tract, in the nature of articulatory action, and in concurrent nonlinguistic aim. When it comes to the expression of taxonomic kinds like phonemes, class membership is open. The ultimate linguistic meaning of a kind of spoken sequence depends on properties that are diagnostic, but not extensionally definitive, of particular speech sounds. These homeostatic properties may be deeply theoretical. And theoretical states and processes are neither transparent, nor routinely accessible, to intuition.

And if our analysis is useful, we will learn much more about the nature of the mind when we use a scientific methodology in studying durable issues about the mind. If introspection and phenomenological analyses have a place in intellectual inquiry, their accuracy must first be justified using methods that do not presuppose their reliability. The range and reliability of introspection—at least about short-lived sensory states—are amply documented in the history of psychophysics, and there the evidence supports the careful use of introspection in philosophical analyses of sensory experience. But this is a modest subset of the domains in which introspective insights have been pressed for philosophical gain. Our own philosophical tradition has held at one point or another that nearly every mental state was transparent to introspection and that our intuitions were high-fidelity guides to accurate theories about the mind. That tradition in philosophy remains a formidable obstacle to knowledge in the study of the mind.

## References

- Appelbaum, I. (1999). 'The Dogma of Isomorphism: A Case Study from Speech Perception'. *Philosophy of Science*, 66: S250–S259.
- (2004). 'Physical Segments and Functional Gestures', In A. Agwuele, W. Warren, and S. Park (eds.), *Proceedings of the 2003 Texas Linguistics Society Conference*. Sommerville, Mass.: Cascadilla Proceedings Project.
- Bell-Berti, F. and Harris, K. S. (1979). 'Anticipatory Coarticulation: Some Implications from a Study of Lip Rounding'. *Journal of the Acoustical Society of America*, 65: 1268–70.
- Berkeley, G. (1709). *A New Theory of Vision*. New York: E. P. Dutton & Co.
- Bernstein, L. E., Demorest, M. E., and Tucker, P. E. (2000). 'Speech Perception without Hearing'. *Perception & Psychophysics*, 62: 233–52.
- Bishop, M. and Trout, J. D. (2005a). *Epistemology and the Psychology of Human Judgment*. New York: Oxford University Press.

- (2005b). 'Pathologies of Standard Analytic Epistemology'. *Notûs*, 39: 696–714.
- Boyd, R. (1999). 'Homeostasis, Species, and Higher Taxa', in R. A. Wilson (ed.), *Species: New Interdisciplinary Essays*. Cambridge, Mass.: MIT Press, 141–86.
- Descartes, R. (1637). *Discourse on Method, Optics, Geometry, and Meteorology*, P. J. Olscamp (trans.). Indianapolis, Ind.: Bobbs-Merrill.
- Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). 'Speech Perception'. *Annual Review of Psychology*, 55: 149–79.
- Doris, J. (2002). *Lack of Character*. New York: Cambridge University Press.
- Eimas, P. D., Siqueland, E. P., Jusczyk, P., and Vigorito, J. (1971). 'Speech Perception in Infants'. *Science*, 171: 303–6.
- Fant, C. G. (1962). 'Descriptive Analysis of the Acoustic Aspects of Speech'. *Logos*, 5: 3–17.
- Fodor, J. A. (1983). *Modularity of Mind*. Cambridge, Mass.: MIT Press.
- Goldstein, L. and Fowler, C. A. (2003). 'Articulatory Phonology: A Phonology for Public Language Use', in A. S. Meyer and N. O. Schiller (eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Berlin: Mouton de Gruyter, 159–207.
- Hirsh, I. J. (1988). 'Auditory Perception and Speech', in R. C. Atkinson, R. J. Herrnstein, G. Lindzey, and R. D. Luce (eds.), *Stevens' Handbook of Experimental Psychology, Volume I: Perception and Motivation*. New York: Wiley-Interscience, 377–408.
- Hockett, C. F. (1955). *Manual of Phonology*. Bloomington, Ind.: Indiana University Publications in Anthropology and Linguistics, 11.
- Honda, K. (1996). 'Organization of Tongue Articulation for Vowels'. *Journal of Phonetics*, 24: 39–52.
- Howell, P. and Darwin, C. J. (1977). 'Some Properties of Auditory Memory for Rapid Formant Transitions'. *Memory & Cognition*, 5: 700–8.
- Hume, D. (1739). *Treatise of Human Nature*, D. F. Norton and M. J. Norton (eds.). Oxford: Oxford University Press, 2001.
- Köhler, W. (1910). 'Akustische Untersuchungen, II' [Acoustic investigations]. *Zeitschrift für Psychologie mit Zeitschrift für Angewandte Psychologie und Charakterkunde*, 58: 59–140.
- Kühnert, B. and Nolan, F. (1999). 'The Origin of Coarticulation', in W. J. Hardcastle and N. Hewitt (eds.), *Coarticulation: Theory, Data and Techniques*. Cambridge: Cambridge University Press, 7–30.
- Ladefoged, P. (2005). *A Course in Phonetics*, 5th edn. Belmont, Calif.: Heinle/Thomson.
- Lakatos, S., McAdams, S., and Causse, R. (1997). 'The Representation of Auditory Source Characteristics: Simple Geometric Form'. *Perception & Psychophysics*, 59: 1180–90.
- Lenneberg, E. H. (1967). *The Biology of Language*. New York: Wiley & Sons.
- Liberman, A. M. (1970). 'Some Characteristics of Perception in the Speech Mode', in D. A. Hamburg (ed.), *Perception and its Disorders*. Baltimore: Williams & Wilkins, 238–54.

- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). 'Perception of the Speech Code'. *Psychological Review*, 74: 421–61.
- Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). 'The Discrimination of Speech Sounds within and across Phoneme Boundaries'. *Journal of Experimental Psychology*, 54: 358–68.
- and Mattingly, I. G. (1985). 'The Motor Theory of Speech Perception Revised'. *Cognition*, 21: 1–36.
- Lindberg, D. C. (1971). 'Alkindi's Critique of Euclid's Theory of Vision'. *Isis*, 63: 469–89.
- Locke, J. (1689). *Essay Concerning Human Understanding*. Oxford: Oxford University Press.
- McDermott, H. J. (2004). 'Music Perception with Cochlear Implants: A Review'. *Trends in Amplification*, 8: 49–82.
- McGinn, C. (1996). 'Another Look at Color'. *Journal of Philosophy*, 43: 537–53.
- MacNeilage, P. F. (1970). 'Motor Control of Serial Ordering of Speech'. *Psychological Review*, 77: 182–96.
- Rootes, T. P., and Chase, R. A. (1967). 'Speech Production and Perception in a Patient with Severe Impairment of Somesthetic Perception and Motor Control'. *Journal of Speech and Hearing Research*, 10: 449–67.
- Miller, G. A. (1951). *Language and Communication*. New York: McGraw-Hill.
- Mirman, D., McClelland, J. L., and Holt, L. L. (2006). 'An Interactive Hebbian Account of Lexically Guided Tuning of Speech Perception'. *Psychonomic Bulletin and Review*, 13: 958–65.
- Noë, A. (2002). 'Experience and the Active Mind'. *Synthese*, 29: 41–60.
- Pessoa, L., and Thompson, E. (2000). 'Beyond the Grand Illusion: What Change Blindness Really Teaches Us about Vision'. *Visual Cognition*, 7: 93–106.
- Pardo, J. S. and Remez, R. E. (2006). 'The Perception of Speech', in M. Traxler and M. A. Gernsbacher (eds.), *The Handbook of Psycholinguistics*, 2nd edn. New York: Elsevier.
- Pisoni, D. B. (1973). 'Auditory and Phonetic Memory Codes in the Discrimination of Consonants and Vowels'. *Perception & Psychophysics*, 13: 253–60.
- Radhakrishnan, S. and Moore, C. A. (1957). *A Sourcebook in Indian Philosophy*. Princeton: Princeton University Press.
- Remez, R. E. (2005). 'Perceptual Organization of Speech', in D. B. Pisoni and R. E. Remez (eds.), *The Handbook of Speech Perception*. Oxford: Blackwell, 28–50.
- Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (1994). 'On the Perceptual Organization of Speech'. *Psychological Review*, 101: 129–56.
- Ryle, G. (1949). *The Concept of Mind*. Chicago, Ill.: University of Chicago Press.
- Saffran, J. R. (2003). 'Statistical Language Learning: Mechanisms and Constraints'. *Current Directions in Psychological Science*, 12: 110–14.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). 'Speech Recognition with Primarily Temporal Cues'. *Science*, 270: 303–4.

- Siegel, S. (2006a). 'Which Properties are Represented in Perception?' in T. Szabó Gendler and J. Hawthorne (eds.), *Perceptual Experience*. Oxford: Oxford University Press, 481–503.
- (2006b). 'How Does Visual Phenomenology Constrain Object-Seeing?' *Australasian Journal of Philosophy*, 84: 429–41.
- Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). 'Chimaeric Sounds Reveal Dichotomies in Auditory Perception'. *Nature*, 416: 87–90.
- Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., and Miyamoto, R. T. (2000). 'Language Development in Profoundly Deaf Children with Cochlear Implants'. *Psychological Science*, 11: 153–8.
- Trout, J. D. (2001). 'Metaphysics, Method and the Mouth: Philosophical Lessons of Speech Perception'. *Philosophical Psychology*, 14: 261–91.
- Tye, M. (2000). *Consciousness, Color, and Content*. Cambridge, Mass.: MIT Press.
- Weinberg, J., Nichols, S., and Stich, S. (2001). 'Normativity and Epistemic Intuitions'. *Philosophical Topics*, 29: 429–60.
- Wilson, M. (1981). *Descartes*. London: Routledge and Kegan Paul.
- Wimsatt, W. (2006). 'Reductionism and its Heuristics: Making Methodological Reductionism Honest'. *Synthese*, 151: 1–31.

