1

Perception, Ecology, and Music

The Ecological Approach to Perception

Rather than considering perception to be a constructive process, in which the perceiver builds structure into an internal model of the world, the ecological approach emphasizes the structure of the environment itself and regards perception as the pick-up of that already structured perceptual information. The simple, but far-reaching, assertion is that the world is *not* a "blooming buzzing confusion", but is a highly structured environment subject to both the forces of nature (gravity, illumination, organic growth, the action of wind and water) and the profound impact of human beings and their cultures; and that in a reciprocal fashion perceivers are highly structured organisms that are adapted to that environment.

> The environment described [here] is that defined by ecology. Ecology is a blend of physics, geology, biology, archeology, and anthropology, but with an attempt at unification. The unifying principle has been the question of what can stimulate a sentient organism. (Gibson 1966: 29)

What is important is to consider what is *directly specified* by environmental information—not what a perceiving organism can interpret in, or construct from, a stimulus. The shape, mass, reflectance, density, and texture of a physical object directly determine the stimulus information that it gives off in different sensory domains when it is illuminated, struck, heated, scraped, blown with a stream of air, etc. For example, a hollow piece of wood will differentially reflect light of certain wavelengths according to its composition and the manner in which it has been cut and treated, and will vibrate with a certain pattern of frequencies if struck by another object (and as a function of the hardness and mass of that object) according to the degree to which it has been hollowed, and the specific size and shape of the cavity. This information directly specifies properties of the object itself to an organism equipped with an appropriate perceptual system. The amplitude and frequency distribution of the sounds emitted when this piece of hollowed wood is struck are a direct consequence of the physical properties of the wood itself—are an "imprint" of its physical structure—and an organism does not have to do complex processing to "decode" the information within the source: it needs to have a perceptual system that will *resonate* to the information:

> Instead of supposing that the brain constructs or computes the objective information from a kaleidoscopic inflow of sensations, we may suppose that the orienting of the organs of perception is governed by the brain so that the whole system of input and output resonates to the external information. (Gibson 1966: 5)

This "resonance" or "tuning" of the perceptual system to environmental information is different from the resonance of a string or hollow tube, for example, since these are fixed, and will only resonate to a specific kind of event—a particular frequency. The C-string of a cello, for example, will resonate sympathetically to a C sounded by another nearby instrument (or to some of the subharmonics of that C) but will not resonate to other kinds of sounds. Nor is this tuning like an analogue radio receiver, which can be tuned to any one of a great variety of broadcast frequencies but which needs someone to turn the knob. Perception is a *self*-tuning process, in which the pick-up of environmental information is intrinsically reinforcing, so that the system self-adjusts so as to optimize its resonance with the environment: "A system 'hunts' until it achieves clarity," wrote Gibson (1966: 271), a little like the scanning of a modern digital tuner, a device Gibson never encountered.

If ecological theory was simply the claim that organisms resonate to environmental information, which in turn directly specifies the objects and events from which it emanates, it would have little explanatory value: perception would be no more than a magical affinity between a perfectly structured environment and a miraculously endowed and adapted perceiver. There are three factors, however, that make the theory both more realistic and more interesting: the relationship between perception and action; adaptation; and perceptual learning.

Perception and Action

When humans and other animals perceive the world, they do so actively. Perception is essentially exploratory, seeking out sources of stimulation in order to discover more about the environment. This operates in so many ways and so continuously that it is easy to overlook: we detect a sound and turn to it; we catch sight of an object, turn our eyes to it, lean forward and reach out to touch it; we get a whiff of something and deliberately breathe in through the nose to get a better sense of its smell. These and countless other examples illustrate the constant orienting of the organism to its environment, the constant search to optimize and explore the source of stimulation. Actions lead to, enhance, and direct perception, and are in turn the result of, and response to, perception. Resonance is not passive: it is a perceiving organism's active, exploratory engagement with its environment.

In the circumstances of entertainment and aesthetic engagement, however, overt manifestations of the perception-action cycle are often blocked or transformed, as Windsor (1995; 2000) has also discussed. Watching films and television, looking at paintings or sculpture in a gallery, and listening to music in a concert hall deliberately place perceivers in a relationship with the objects of perception that prevents them from acting upon or exploring those objects in an unhindered fashion. Many of the reactions that people have to these special circumstances (reaching out to touch a sculpture; approaching or standing back from a painting; laughing, crying, or flinching at a film; foot- and finger-tapping in response to music) are a residue of the more usual relationship between perception and action, as are the specific conventions that regulate these reactions (ritualized audience participation at pantomimes, "Please do not touch" signs at exhibitions, darkened auditoria, socially enforced silence and immobility at concerts, applause at regulated moments). The interruption or suspension of the perception-action cycle that characterizes some forms of aesthetic engagement is, of course, culturally specific; it is at its most extreme in some of the "high" art forms of the West and in circumstances in which formal ceremony and aesthetics interact. In many other contexts (folk traditions, popular cultures, some experimental art and music), active participation is the norm. The specific consequences of what might be called the contemplative or "disinterested" (Meidner 1985) perceptual attitude required or encouraged by the autonomous art forms of the West is an issue I take up again in chapter 5.

Adaptation

Organisms and their environments are constantly changing. The "goodness of fit" between an organism and its environment is not a matter of chance: it is the product of mutual adaptation brought about by an evolutionary process. The giraffe's long neck in an environment of savannah dotted with thorny trees is not a lucky break: it is the result of an adaptation that has left the giraffe as a successful competitor in an environment where the ability to reach the high branches of trees with sharp thorns lower down is an advantage. Similarly, the fact that the human basilar membrane demonstrates a logarithmic frequency distribution over much of its length (i.e., a fixed distance on the basilar membrane corresponds to a roughly constant ratio of frequencies) is no miracle of divine design or happy accident: it confers an advantage in a world where struck and blown objects tend to radiate sounds with harmonic series properties; and it is a particular advantage in a species for which speech and other forms of vocal/auditory communication are so important. It allows, for instance, for the equivalence of the same pitch profile in different registers—an important attribute when trying to respond appropriately to the vocalizations of individuals with different vocal ranges (men and women, adults and juveniles).

The resonance of a perceptual system with its environment is a product of evolution and adaptation in the same way that an organism's feeding behavior is adapted to the available food supply. It is no miracle that rabbits "resonate" with grassland: not only are they adapted to compete extremely well in that physical environment, but their presence in such an environment directly contributes to the continuation and even expansion of that environment itself. By eating the shoots of tree and bush seedlings that might otherwise compete with the grass, the rabbits help to create and sustain the environment in which they thrive.

Without suggesting too simplistic a leap from rabbits and grassland to humans and music, the same interdependency and mutual adaptation nonetheless apply. Human beings have exploited natural opportunities for music making (the acoustical characteristics of materials and the action-possibilities of the human body) and have also adapted themselves to those opportunities, and enhanced those opportunities, through tool-making of one sort or another—from drilled bones, through catgut and wooden boxes to notational systems, voltage-controlled oscillators and iPods. Once made, all these artifacts help both to sustain existing musical behaviors (i.e., they help to perpetuate the musical ecosystem) and to make new behaviors possible. This mutual adaptation between human beings and their (musical) environment is neither reducible to conventional evolutionary principles, nor is it independent of them: culture and biology are tangled together in complex ways, but nonetheless constitute a single connected system (see Cross 2003).

Perceptual Learning

Adaptation between an organism and its environment occurs over evolutionary time, not in the life span of a single individual. But this does not mean that individuals come into the world with perceptual characteristics that remain fixed and determined throughout their lives: from the moment of their first encounters with the world, organisms are immersed in a continual process of perceptual learning—a matter to which the ecological approach to perception has paid considerable attention, in particular through the work of Eleanor Gibson (e.g. Gibson 1969). Cognitive psychology has also recognized the importance of changes in the perceptual capacities of humans and other animals, but has tended to treat the question in terms of the enrichment, or increased "coding power" of perception through experience and learning (as discussed in Gibson and Gibson 1955). According to a cognitive view, perceptual skills develop through the accumulation of knowledge that guides and informs them, and which fills in the information that is missing in a chaotic and imperfect environment. By contrast, the ecological approach views perceptual learning as progressive *differentiation*, perceivers becoming increasingly sensitive to distinctions within the stimulus information that were always there but previously undetected. A newborn human infant is equipped with a relatively small number of very powerful, but as yet rather undifferentiated, perceptual capacities. Exposure to the environment shapes these perceptual capacities, and distinctions that previously went unnoticed become detectable. As the infant explores these new discoveries, further distinctions that were previously unperceivable are revealed, and a cascade of successive differentiations ensues.

The overwhelming majority of this perceptual learning occurs "passively"—though this is a misleading term. What is meant is that there is no explicit *training* involved, no human supervisor pointing out distinctive features and appropriate responses. It is "passive" in the sense that it is not under the direct guidance of any external human agency, but it is, of course, profoundly active from the perspective of the organism itself. As already observed, perception and action are inextricably bound together, and the differentiation of attention that is described here takes place because the actions of the organism on the environment reveal previously unnoticed distinctions which in turn result in modified actions.

As a musical example of passive perceptual learning, consider a young child's discovery of loudness and pitch on a xylophone. On first encountering a xylophone, the child's more-or-less unregulated experiments with hands or sticks will result in all kinds of accidental sounds. With unsupervised investigation, the child may discover that different kinds of actions (with more force/with less force, to the left-hand side of the object/to the right-hand side of the object, with the fingers/with a stick) give rise to differentiated results (louder/softer, low pitched/high pitched, sharp attack/dull attack), and even that these distinctions can themselves be used to achieve other goals—funny sounds, scary sounds, surprises, etc. Perceptual learning about pitch height, dynamics, and timbre resulting from manual/aural exploration leads to further perceptual learning about the possibilities of tune building, or expressive function.

As well as the continual passive perceptual learning that goes on in a rich environment, there is also directed perceptual learning—the differentiation of attention that goes on when one person points out a distinction to another, or deliberately puts an individual in a situation designed to elicit perceptual learning. In the xylophone example, an adult might encourage or direct the child to try out certain actions and to pay attention to specific aspects of the resulting sounds. Aural training provides numerous examples of precisely this kind of process: an early skill in traditional aural training, for example, involves learning to recognize that a triad consists of three notes. Untrained listeners—and certainly children prior to training—tend to regard a chord as a single entity. This is a perfectly reasonable and "correct" perception: chords, and especially chords played on the piano, typically consist of notes with closely synchronized onsets, homogeneous timbres, and very similar dynamic levels—all of which help to produce fusion between the chord components, as Bregman (1990: 490–493) points out. So it is perfectly appropriate to hear a triadic chord as a single "thing": it *is* a single thing. But when an instructor points out that this single thing can also be heard to consist of a number of components, he or she is directing the learner's attention to a feature that was always available in the stimulus information but was previously undetected. Awareness of this information is nearly always achieved by a perception/action cycle: the learner is encouraged to "sing the middle note" or produce some other kind of overt action which has the effect of directing attention and consolidating the new perceptual awareness—a "reinforcement" of the perceptual information through the perception/action cycle. Thus the three factors discussed here (perception/action; adaptation; perceptual learning) explain how the resonance of a perceiver with its environment is not preordained or mysterious: a newborn infant (who, research increasingly reveals, has already had many weeks of prenatal perceptual learning; see Lecanuet 1996) has a limited range of powerful perceptual capacities and predispositions that give it a foothold in the world; but the overwhelming majority of an adult's more differentiated perception develops from these simple but powerful beginnings by virtue of environmental exposure/exploration and enculturation.

Ecology and Connectionism

One of the complaints that cognitivists make about the ecological approach is that it appears "magical." By rejecting the dominating role of internal representations, and with it the idea of explicit processing stages that are intended to explain perception and cognition, the ecological approach seems to retreat into a quasi-mystical belief that perception "just happens" as a result of a miraculous tuning of perceptual systems to the regularities of the environment. That charge is based on a fundamental misrepresentation of the ecological approach—one that completely ignores the central role of perceptual learning. The tuning of a perceiver's perceptual systems to the invariant properties of the environment is no happy accident, nor the result purely of some kind of Darwinian biological adaptation: it is a consequence of the flexibility of perception, and the plasticity of the nervous system, in the context of a shaping environment. Perceptual systems become attuned to the environment through continual exposure, both as the result of species adaptation on an evolutionary time scale, and as the consequence of perceptual learning within the lifetime of an individual.

But this still seems to beg the question: perceptual systems may be plastic, and the environment may be highly structured, but how does the shaping that is supposed to arise out of the interaction of the two actually take place? If internal representations and all of the mechanisms of a more standard cognitivist account are rejected, what is there instead? Having adjusted and adapted in some manner, in what sense does a perceiving organism actually perceive or know anything? And how does it ever know anything more than, or different from, the cumulative impact of the specific encounters with the environment that constitute its history? How can it ever generalize to novel situations, or be sensitive to certain aspects of the environment and not others?

One way to understand how this is possible is to consider connectionist models of perception—not as literal models of the actual structures and processes that may be involved, but rather as a relatively concrete "metaphor" for the ecological approach.[1] Despite a critique by Costall (1991), who argues that connectionism ignores the mutualism and evolving nature of the relationship between organism and environment, there are features of the approach that shed interesting light on ecological principles—however partial a representation the metaphor might in the end turn out to be.

Connectionist modeling, which was widely discussed in psychology and computer science following the publication of Rumelhart and McLelland's influential book *Parallel Distributed Processing* (Rumelhart and McLelland 1986), differentiates itself from traditional Artificial Intelligence (AI) by claiming that perceptual and cognitive processes can be modeled as the distributed property of a whole system, no particular part of which possesses any "knowledge" at all, rather than as the functioning of explicit rules operating on fixed storage addresses which contain representations or knowledge stores (a crude characterization of AI). A connectionist model typically consists of a network of nodes, interlinked with connections that can take variable values representing their strength (or weight). A layer of input units is connected to a layer of output units, with a variable number of "hidden layers" (usually no more than about two or three) in between. When input units are stimulated, a pattern of activation spreads through the network, the pattern depending on the structure of the connections and the weights assigned to them, and converging on a number of output units. Typically, the network is initially set up with random values assigned to the connection weights, so that the first "activation" results in random behavior of the system as a whole. Thereafter, the behavior of the system becomes more or less structured either on the basis of supervised learning, or according to a principle of self-organization. In supervised learning, the network is guided towards an intended final behavior by means of an explicit set of target values, provided by the experimenter/programmer. By contrast, in a self-organizing network, the final state of the system is not known in advance (although the experimenter/programmer will have an idea of the pattern of behavior that the network is supposed to model), and the system changes over time simply through repeated exposure to "stimulus" information (i.e. input).

Consider a connectionist approach to modeling two examples of musical behavior: listeners' preferences for simple melodies of various types and listeners' awareness of tonal functions.[2] Let us assume that for the first problem there are some data collected from a previous empirical study which show that for a particular collection of short unaccompanied melodies, listeners prefer those which start and finish on the same note, generally move in a stepwise manner, but contain at least two intervals of a major third or more. A network is then constructed that takes as its input the intervals between adjacent notes in each of the experimental melodies (since it is intervals rather than pitches that seem to determine preference), and has as its output a simple binary classification (like/dislike). The connections between the input units (sequences of intervals) and the output units (like/dislike), via some number of hidden layer units, are initially random—so that the first melody input will result in a response which is randomly "like" or "dislike," and therefore equally likely to be "correct" or "incorrect" in relation to the empirical data. In order to train the network by means of supervised learning, the network's correct responses (i.e. responses that conform to the empirical data) are reinforced by adjusting the connection weights between the input units and the output units. Incorrect responses are inhibited by changing the connection weights (increasing or decreasing them as appropriate) so as to steer the system towards the target relationship between input and output. In this way, over a period of supervised training that makes use of a subset of the melodies from the empirical study, the network becomes more and more differentiated in relation to its originally randomly organized state, as it is "shaped" by the supervisor.

At some point this training phase finishes, and the behavior of the system is then observed in relation to a number of melodies from the original set that have *not* so far been used. If the structure of the network and the principles on which it is based are appropriate, and the training period has been sufficient in terms of the number and variety of melodies presented, the network should now be able to classify these new melodies, employing the same general principles, in a way that mimics the preferences of the listeners on which the model is based. Melodies that listeners liked should activate the corresponding "like" output unit in the model, and melodies that the listeners disliked should activate the "dislike" unit.

In this imagined example, supervised learning seems appropriate because the experimenter/programmer knows in advance about listeners' preferences and is trying to train a network to exhibit those same preferences in order then to explore how well they generalize. It is essentially analogous to what happens when a rather prescriptive music teacher instructs a class about the differences between well-formed and ill-formed melodies, by playing them simple tunes and asking the students to judge whether each one is "good" or "bad," and providing feedback (i.e. the "right" answer) after each example. After a while, the expectation would be that the students might be able to generalize their classifying abilities (assuming that the training had been based on principled behavior of some kind!) to new examples of simple tunes not taken from the training set. The same expectation (and the same proviso about principled behavior) can apply to a network.

The parallel between the model and an ecological approach is the implicit manner in which both the human listeners, and the equivalent network, acquire their skills. Listeners asked to make preference judgments of the kind described here generally do not have explicit knowledge of how they make them, and in the same way the connectionist model described here[3] has neither been instructed with, nor does it contain, any explicit rules, and contains no explicit processing stages or knowledge representations. The adapted behavior arises out of a process of shaping, the effect of which is distributed throughout the network and is only seen when the network engages with its "environment"—in this case an environment consisting only of sequences of melodic intervals and "like" or "dislike" responses.

Using similar distributed principles, self-organizing networks arrive at some kind of "solution," or structured behavior, without any kind of explicit instruction. As empirical studies have shown, suitably enculturated listeners can make systematic judgments about tonal structure in music (expressed, for instance, in terms of the perceived completeness or stability of a sequence) without any experience of "supervised learning" or formal music instruction (see e.g. Krumhansl 1990). The belief is that people become attuned to this property of music through simple exposure, due to the interaction of the regularities of the tonal environment with certain fundamental perceptual capacities of the auditory system. A variety of self-organizing methods have been explored in the connectionist literature to model this kind of unsupervised or "passive" learning, particularly those proposed by Kohonen (1984) and Grossberg (1982), which depend in one way or another on what is known as "competitive learning." The basic idea of competitive learning is that an existing connection (which may have been made fortuitously) is strengthened every time the connection is reiterated, while adjacent ("competing") connections are weakened. The consequence of this is that regularities in the environment progressively shape the network simply by virtue of their recurrence and co-dependence. If certain combinations of environmental events occur more frequently than others, then the corresponding connections in the network will become increasingly heavily weighted, and adjacent connections will become attenuated.

In a series of publications, Bharucha (1987; 1991a; 1991b; 1999; Bharucha and Todd 1991) has presented and developed a connectionist model for the perception of tonal harmony, as also has Leman (1991), using a slightly different (but nonetheless self-organizing) approach. Both Bharucha's and Leman's models take tonal material (notes or chords) as input, and give tonal interpretations, in the form of a dynamically changing sense of key, as output. In both cases, the networks start out in an essentially undifferentiated state, and by exposure to tonal materials develop tonally specific characteristics. These changes take place not because the networks acquire representations, or increase their memory content, but because their patterns of connectivity and consequent behavior as systems change: in other words, they adapt.

This relationship between adaptation and memory is tricky. Some authors (e.g. Crowder 1993) resist making the distinction by adopting a functionalist perspective: any system that behaves differently by virtue of past experience or exposure can be said to display memory. In common with Gibson (e.g. Gibson 1966: 275–278), I want to distinguish between memory proper, which involves the encoding, storage, and retrieval of previous events, and perceptual learning or environmental shaping, which is a sensitivity to current events, brought about by adaptation of the perceptual system to environmental invariants. To invoke "memory" on every occasion that an organism demonstrates a response to the environment that has been shaped by previous exposure leads to absurd consequences: the curious shapes of trees and bushes that grow in windy places, for instance, would have to be seen as "memories" of earlier windy interactions. This is manifestly wrong, and the more appropriate and familiar way to talk about such trees and bushes is in terms of growth: having grown in a particular way, under the influence of prevailing winds, they now interact with the wind in a specific manner. Neural networks can be regarded in the same light: having been exposed to environmental shaping (such as tonal chord sequences), the network has "grown" in a certain manner with the consequence that it behaves in a specific and differentiated fashion when it again "feels the wind" of the same or similar sequences blowing upon it. In the brain, this adaptive growth is referred to as plasticity (see Gregory 1987), and it is increasingly recognized as a fundamental and defining feature of the brain's functioning (e.g. Hurley and Noë 2003).

Changing the connection weights in a network model cannot be directly equated with changes in the connections (synapses) between neurons in the brain, but it is a reasonable approximation—or at the very least a metaphor for it. The sense of key in a tonal environment, or the identification of a characteristic motivic/harmonic procedure, both of which are temporally distributed properties, are behaviors that networks of this kind can be shown to demonstrate after suitable exposure. Once a network starts to behave in this way, it has in an ecological sense become attuned to the environment. If it is then exposed to a sufficiently similar musical sequence, it will enter into more or less the same state as before.[4] Reaching the same state as a result of exposure to the same (or similar) material is what recognition is—but at no point is a representation involved. The perceptual system has entered into a state that is "attuned" to the particular characteristics of the environment, and the state is one that the system has been in before. Recognition is that kind of perception for which the system has become adapted (or tuned).

The environmental events that gave rise to the particular connections and weightings in the system (or the synaptic links in the brain) are manifest relationships in the concrete physical world. The subsequent "tuning" of the network (whether artificial model or actual brain) is the result of exposure to those real events—their trace, or residue. That trace, and its reactivation, is experienced as a dynamic state of the network and thus a state of mind—an awareness of real-world relationships. When part of the original event sequence is encountered again (for example just the first two or three chords of a longer tonal sequence), the rest of the original dynamic state may be activated to greater or lesser extent—a principle that in psychology is called perceptual facilitation, or priming (e.g. Bharucha 1987; Bharucha and Stoeckig 1987), and which has been the subject of research in language, music and visual perception (e.g. Swinney 1979). Only some of the original events are present again on this occasion, so the meaning of the event, understood as the response of the whole network to the particular events that are present on this occasion, is achieved by virtue of the facilitated connections within the network that result from previous exposure. But these connections are only activated because of previous exposure: it is as though the network "relives" the real events even in their absence. It is the actual exposure to the original conditions, however, that gave rise to the connections in the first place, and if nothing like those material conditions were ever to be encountered again, then eventually the facilitated connections would disappear too. The "noise" in a connectionist network (a constant background of random activation, exacerbated by the atrophying of weak connections by competitive learning) means that to be maintained, structured connections need to be reactivated or reinforced from time to time. Fundamentally, then, the whole system depends on, but is not reducible to, the effects of exposure to real-world events. As Gibson (1966) put it: "all knowledge rests on sensitivity" (26).

Invariants in Perception

From the point of view of adaptation, an organism's most pressing need is to know "what is going on" in the environment. As a consequence, the ecological approach emphasizes the critical importance of information as information *for* something (objects and events). Perception is not a process of taking in "raw sensations" and then interpreting them, and the purely sensory character of perception is usually not at all evident to a perceiver. It is the objects and events that are specified in perception that are important: whether people notice a fire because they see the flames, hear the crackle, smell the smoke, or feel the heat is of little importance compared to the fact that they detect the event (the fire). When perception proceeds in an unproblematic way, we are usually unaware of the sensory aspect of the stimulus information, and are only attuned to the events that are specified by stimulus structure. But when that relationship is problematic, the stimulus structure itself can become more evident.

This can be the case with ambiguous or degraded perceptual information, as figure 1.1 is intended to illustrate. People who look at this image and see no recognizable scene (look at it now), tend

Figure 1.1 What do you see here? (Photographer: R. C. James).

to be much more aware of the array of differently shaped patches of black and white than those who immediately see a recognizable scene. The characteristics of the visual array (patches of black and white with certain shapes and orientations) are much more visible when you cannot see what the array specifies than when you see it as a picture of a Dalmatian dog. Similarly, a piece of music which presents sampled everyday sounds in a transformed, or radically de-contextualized, fashion may encourage a listener to detect the structure of the stimulus information (what might be called "purely sonorous" structures) by virtue of a disruption of the normal relationship of source specification (see Dibben 2001). Paradoxically, when a person hears what a sound means (i.e. understands the sound in relation to its source), it becomes more difficult to detect the sound's distinctive features. Speech perception provides a striking case of this: it is a common experience when listening to the sounds of an unfamiliar foreign language to notice the huge variety and specific qualities of the sounds that make up the language—to be quite acutely aware, in other words, of continuously variable acoustical features but to understand nothing. To a native speaker/listener, however, these are paradoxically far more difficult to detect even though they are the critical features that enable the language to function as a communicative medium at all.

Speech also demonstrates another very general characteristic of perception: the environment is usually perceived as comparatively stable despite widespread and continual physical variations. A native speaker/listener perceives the speech of others as being identifiable and comprehensible despite dramatic differences in the physical signals (vocal range, speed, accent, loudness, etc). How are the stability and constancy of the perceived environment to be explained? The key to this lies in the principle of invariance—the idea that within the continuous changes to which a perceiver is exposed there are also invariant properties. As the ecological approach emphasizes, these invariant properties are those of the stimulus information itself—not a representational projection by the perceiver. They are relationships between stimulus properties that remain unchanged despite transformations of the stimulus array as a whole. For example, a person hearing a passing motorbike will be exposed to a continuously changing array of acoustical information, but within that array there will be invariant acoustical properties, in a specific pattern of relationships, which together identify the motorbike and which remain constant under transformation (pitch changes due to the Doppler effect, amplitude change due to distance, etc). Warren and Verbrugge (1984), for example, showed that the sounds of objects bouncing and breaking could be distinguished from one another on the basis of the temporal properties of the impact sequences, and showed that listeners could still reliably distinguish between bouncing and breaking in artificially generated examples that used only a highly simplified simulation of the temporal properties of real impact sequences. The acoustical invariants that specify bouncing and breaking are, in other words, two different temporal patterns of impacts.

Music offers a particularly clear example of invariance in the perceived identity of material under transposition and other kinds of transformation. A theme or motif in music can be regarded as an invariant (a pattern of temporal proportions and pitch intervals) that is left intact, and hence retains its identity, under transformations such as pitch transposition or changes in global tempo. As Dowling and Harwood (1986) point out, these invariants can be of different orders, from local and specific to more general:

> Some invariants are specific to a certain piece, such as the pitch and rhythm contour of the initial theme of Beethoven's Fifth Symphony. Other invariants heard in a particular piece are common to a large family of similar pieces, for example, the characteristic repeated rhythmic pattern of certain dances such as beguine and tango. In terms of scale-structure invariants, a piece may exhibit a particular shift between keys in the middle. . . . Such a pattern involves variation within the single piece, but if the listener has heard many such pieces with the same pattern of modulation, then that pattern constitutes an invariant that the listener can perceive in each piece he or she hears, even pieces not heard before. Such invariants across sets of pieces constitute what we mean by a style. (Dowling and Harwood 1986: 160–161)

Notice the implicit reference to perceptual learning: listeners become more attuned to the invariants that specify a style, or a particular harmonic invariant (e.g. tonic/dominant alternation) through exposure to a particular repertoire, whether that exposure is accompanied by direct instruction or not. But notice also that these higher order invariants are no more abstract than the most specific and local invariant that is unique to one particular context, even if some of them may be more extended in time: in every case the invariant is a set of relationships that is available in the stimulus information. The ecological approach resists the cognitive tendency to explain constancy and invariance in terms of internal processes and points to the environmental and "given" nature of the phenomenon.

Affordance

The idea of invariants leads to another important concept that Gibson developed, that of "affordance," which relates directly to the central theme of this book—musical meaning. Here is the passage in *The Senses Considered as Perceptual Systems* where Gibson introduces the term:

> When the constant properties of constant objects are perceived (the shape, size, color, texture, composition, motion, animation, and position relative to other objects), the observer can go on to detect their *affordances*. I have coined this word as a substitute for *values*, a term which carries an old burden of philosophical meaning. I mean simply what things furnish, for good or ill. What they *afford* the observer, after all, depends on their properties. The simplest affordance, as food, for example, or as a predatory enemy, may well be detected without learning by the young of some animals, but in general learning is all-important for this kind of perception. The child learns what things are manipulable and how they can be manipulated, what things are hurtful, what things are edible, what things can be put together with other things or put inside other things—and so on without limit. He also learns what objects can be used as the means to obtain a goal, or to make other desirable objects, or to make people do what

> he wants them to do. In short, the human observer learns to detect what have been called the values or meanings of things, perceiving their distinctive features, putting them into categories and subcategories, noticing their similarities and differences and even studying them for their own sakes, apart from learning what to do about them. All this discrimination, wonderful to say, has to be based entirely on the education of his attention to the subtleties of invariant stimulus information. (Gibson 1966: 285)

In this definition of affordance, Gibson places considerable emphasis on the properties of objects themselves, and some authors (e.g. Noble 1991) have criticized Gibson for having a rigid and one-sided approach. Elsewhere in his writing, however, Gibson presents the concept in a much more dialectical or mutual fashion, pointing out that although affordances depend on the properties of the object they don't depend solely on them: affordances are the product both of objective properties and the capacities and needs of the organism that encounters them.

> The verb *to afford* is found in the dictionary, but the noun *affordance* is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment. (Gibson 1979: 122; emphasis in original)

As Flach and Smith (2000) point out, Gibson's noun ("affordance") threatens to reify an essentially dynamic concept, so there may be advantages in sticking with the verb. To a person, a wooden chair affords sitting, while to a termite it affords eating. Equally, the same chair affords self-defense to a person under attack—an illustration of the way in which an organism can notice different affordances according to its own changing needs. The relationship is neither a case of organisms imposing their needs on an indifferent environment, nor a fixed environment determining possibilities: to a person, a chair can afford sitting and self-defense, but simply cannot afford eating because of the relationship between the capabilities of the human digestive system and the properties of wood. Note that the principle of affordance does not imply that perception will always be obvious and unambiguous, since objects and events can give rise to more than one perceptual experience. If perceptual information "carries different or contradictory variables of information it will afford different or contradictory perceptual experiences" (Gibson 1966: 248).

Although Gibson writes here of perceptual experience as an affordance, elsewhere in his writing and in the writing of other ecological psychologists, affordances are primarily understood as the *action* consequences of encountering perceptual information in the world. A chair affords sitting, a stick affords throwing, raspberries afford eating, a sharp pencil affords writing. In many ways, music fits into this scheme unproblematically: music affords dancing, worship, co-ordinated working, persuasion, emotional catharsis, marching, foot-tapping, and a myriad other activities of a perfectly tangible kind. But in certain musical traditions (and the concert music of the West is an obvious example) listening to music has become somewhat divorced from overt action—has become apparently autonomous. The particular consequences of these specific circumstances are examined elsewhere in this book (chapter 5 and the conclusion), but the example highlights the social nature of affordances for human beings. A concentration on common or garden objects might lead to the erroneous conclusion that affordances are a simple matter of physical properties and perceptual capacities. But even the most cursory consideration of some more socially embedded objects demonstrates the importance of the social component. A violin, for example, affords burning, but social factors ensure that this is a rather remote affordance—which might only be realized in extreme circumstances or by an individual who had no regard for (or even deliberately disdained) the musical context which regulates its affordances.[5]

Nature and Culture

The perspective offered so far implies an environment consisting of sources of information that are all of the same general kind. But can the sound of horses' hooves, the sound of a radio commentary on a horse race, the sound of Tennyson's poem "The Charge of the Light Brigade," and the sound of Wagner's "Valkyries" leitmotiv really be regarded as equivalent sources of information about horses? Can a single approach deal with a whole range of more or less culturally mediated information sources? Gibson himself wrote of the need to avoid a sharp division between culture and nature—and in doing so made one of his rare references to music:

> In the study of anthropology and ecology, the 'natural' environment is often distinguished from the 'cultural' environment. As described here, there is no sharp division between them. Culture evolved out of natural opportunities. The cultural environment, however, is often divided into two parts, 'material' culture and 'non-material' culture. This is a seriously misleading distinction, for it seems to imply that language, tradition, art, music, law, and religion are immaterial, insubstantial, or intangible, whereas tools, shelters, clothing, vehicles, and books are not. Symbols are taken to be profoundly different from things. But let us be clear about this. There have to be modes of stimulation, or ways of conveying information, for any individual to perceive anything, however abstract. He must be sensitive to stimuli no matter how universal or fine-spun the thing he apprehends. No symbol exists except as it is realized in sound, projected light, mechanical contact, or the like. (Gibson 1966: 26)

Once a convention or tradition is established and is embodied in widespread and relatively permanent objects and practices, it becomes as much a part of the environment as any other feature. As

Heft (2001) and Windsor (1995; 2000) have pointed out, cultural regularities are as much a part of the environment as natural forces, and they exert their influence on the invariants of the world in just the same way.

It is important to recognize the cultural specificity of perception, but since for human beings *every* circumstance and experience is cultural, there is no basis on which to propose some kind of primary pre-cultural experience characterized by a spurious immediacy. The theoretically arbitrary nature of linguistic and other semiotic codes is largely irrelevant to the way in which they function once a system and community are established: once embedded in a system, they are subject to enormous systematic inertia and cannot simply be overturned at a moment's notice. Although arbitrary in principle, they take on a fixed character in practice.

> The association of sound and representation is the outcome of a collective training (for instance the learning of the French tongue); this association—which is the signification—is by no means arbitrary (for no French person is free to modify it), indeed it is, on the contrary, necessary. . . . We shall therefore say in general terms that in the language the link between the signifier and signified is contractual in its principle, but that this contract is collective, inscribed in a long temporality (Saussure says that 'a language is always a legacy'), and that consequently it is, as it were, *naturalized*; in the same way, Levi-Strauss specified that the linguistic sign is arbitrary *a priori* but non-arbitrary *a posteriori*. (Barthes 1968: 50–51; emphasis in original)

The same set of principles, therefore, can account for the ways in which perceivers pick up information from all parts of the environment—cultural and natural. When I hear someone explaining that the "Valkyries" leitmotiv is just one example of a "horse" topic in music, the vocal sounds may specify an adult male speaker from Scotland, who is animated and enthusiastic, is standing about two meters away from me and facing me, and is telling me about horses, rhythm, and the history of musical materials. The sounds specify the speaker's sex by virtue of pitch and timbral features that are the direct consequence of the size and shape of a man's vocal tract (a natural consequence), even as a portion of these same sounds also specify the word "topic," which in turn denotes a particular concept by virtue of a cultural (linguistic) convention.

Perception and Cognition

A consideration of language demonstrates the close relationship between perception and cognition. Because of its emphasis on understanding perception, the ecological approach in general, and Gibson in particular, have been accused of having no theory of cognition—or even of rejecting cognition altogether. As Reed (1988; 1991; 1996) has shown in Gibson's own writings, and has argued from more general ecological principles, nothing could be further from the truth. Gibson's own aim was to develop a cognitive psychology—but one which theorized perception in a radically different manner from the information-processing approach and also expressed the relationship between perception and cognition quite differently. Despite the incorporation of "top-down" and "bottom-up" interactions, overlapping stages of processing, and so on, the standard cognitive approach is to regard perception as simply the starting-point for a series of cognitive processes—the information-gathering that precedes the real business of sorting out and structuring the data into a representation of some kind. Perception starts when stimuli cause sensations, according to this view, and all the rest is cognitive processing of one sort or another.

The ecological approach presents the situation entirely differently because it rejects the whole idea of "stimuli" in perception. Perceiving organisms seek out and respond to perceptual information that specifies objects and events in the environment, and this perceiving is a continuous process that is both initiated by, and results in, action.[6] One consequence of recognizing perception as a process is that, while mainstream psychology presents the temporal aspect of perception as a stream of discrete stimuli, processed separately and "glued together" by memory, an ecological approach sees it as perceptual flow—the specification of objects and events over time. There is nothing more problematic in principle about temporal successiveness than spatial adjacency in the distribution of perceptual information. Some of Gibson's early research was concerned with investigating the perceptual flow that specified the horizon (or the point of touchdown) as pilots landed aeroplanes (Gibson 1958), and a considerable amount of subsequent research has shown the importance of both optic and acoustic flow in a variety of perceptual tasks with humans and other animals (e.g. Lee 1980; Warren and Verbrugge 1984).

The relationship between perception and cognition is, for Gibson and most other ecological psychologists, bound up with the distinction between direct and indirect forms of knowing. In his 1966 book, Gibson wrote:

> In this book, a distinction will be made between perceptual cognition, or knowledge of the environment, and symbolic cognition, or knowledge *about* the environment. The former is a direct response to things based on stimulus information, the latter is an indirect response to things based on stimulus sources produced by another human individual. The information in the latter case is *coded*; in the former case it cannot properly be called that. (Gibson 1966: 91; emphasis in original)

Representational systems have particular properties that go beyond their purely perceptual attributes: language, for instance, has the property of semantics that gives it the possibilities of predication and discursiveness—the capacity to articulate and communicate *about* something. It can be used to provide knowledge of abstract concepts (the idea of eternity), and of objects and events that are elsewhere (the layout of the surface of the moon), or in the future (tomorrow's weather)—none of which can be perceived (or not at this time and place). When we learn about something by virtue of a representational system of some kind (language, maps, road signs, etc.), we learn not by virtue of what the perceptual information directly specifies, but because what is specified in turn stands for something conceptual, and this conceptual content both adds to our experience and helps to guide our subsequent perception. If I read about the layout of the surface of the moon, this not only informs me about it: it also helps me to perceive and navigate it if I subsequently arrive there. Gibson writes of the ways in which "Perceiving helps talking, and talking fixes the gains of perceiving" (Gibson 1966: 282), though what people come to know about the world through representational systems can, of course, be completely at odds with what they discover through a direct perceptual encounter. I might read about a thrilling-sounding roller-coaster ride, only to find that I hate it when I actually have a go.

We live in a world permeated by representational systems, but it would be wrong to conclude that all of human experience therefore consists of symbolic cognition. Representational systems can guide perceptual information pick-up explicitly or tacitly, and can lead to the accumulation and transformation of knowledge, but every kind of knowing rests upon or involves a perceptual relationship with the environment. In the specific case of music, the relationship between these different ways of knowing has been widely debated (e.g. Cook 1990; 1994; 1998; Kerman 1985; Nattiez 1990; Scruton 1997). The construction of musical meaning through language and other forms of representation is undeniable, but it does not proceed independently of the affordances of musical materials. Ideologies and discourses, however powerful or persuasive they may seem to be, cannot simply impose themselves arbitrarily on the perceptual sensitivities of human beings, which are rooted in (though not defined by) the common ground of immediate experience.

Three More Sound Examples

Consider now three more imagined sound clips like those discussed at the start of the introduction—tracks on an unlabeled recording. Imagine that the first consists of the sounds of a violin being tuned. What do these sounds specify—or rather (and this amounts to the same question) what is their perceptual meaning? Presented like this, the question is unanswerable, since the perceiver is unspecified and, as already discussed, the ecological position rests on the premise that perceptual specification is a reciprocal relationship between the invariants of the environment and the particular capacities of the perceiver. Let us assume, therefore, that the listener presupposed in the discussion of the three following examples is a person enculturated in mainstream classical and pop music of the Western tradition, with roughly the aural awareness and technical understanding of a university music student. Each example is played over a hi-fi, the listener is asked to say what he or she hears, and let us suppose that the response to this first one is "someone tuning a violin." Notice that this description of the perceptual meaning of the sound (or what the sound specifies) refers to a number of different kinds of object or event: there is the instrument (violin), presumably specified by invariants such as the timbre, pitch height, and attack characteristics, which also specify the mode of activation (bowing) of the instrument and thus help to signal both its identity and the presence of a human being. Then there is the particular kind of event (*tuning* the instrument) that is specified in the stimulus invariants—just as the instrument itself is. In this case the invariants would include the irregular, nonmetrical rhythm of the bow strokes; the consistent sounding of only open strings (specified by their characteristic timbre) at intervals of approximately a perfect fifth, always in pairs sounded together; and the continuous pitch glides in just one of the two paired strings that bring the pitches nearer to, or further from, a perfect fifth. Changing any of these invariants has the potential to cause the sound to specify a quite different musical event: if the two pitches, for instance, both varied in a continuous gliding manner in perfect parallelism, the resulting sounds might be heard as specifying a person playing or practicing a passage of *music* of some kind. This is a significantly different kind of event from the culturally specific phenomenon of "tuning," which in the Western classical system has a particular cultural value, lying as it does outside the boundary of a "piece of music."[7] It would be a straightforward matter to investigate the precise nature of the invariants that specify "tuning a violin" by modifying each of the invariant dimensions (rhythmic irregularity, interval structure, pitch glides, timbre) mentioned here—and any others that turned out to be important. And it is also clear that what is specified is both material and concrete (an instrument, a body, a class of action) and also social: tuning an instrument is a socially defined practice, with a distinct place in the cultural system of Western concert music.

Now imagine another example with the same general conditions—this time a recording of a perfect cadence in F played on the piano. Our assumed listener might give a variety of answers to the question, "What do you hear?" depending on his or her specific skills. Among such answers might be: "a musical ending"; "an extract from an aural test"; "a cadence played on the piano"; "a perfect cadence in F played on the piano." The response given would depend on the listener's descriptive competence, current preoccupations, and particular perceptual capacities: the last response would only be possible (given the circumstances) for a person with absolute pitch. But the four possibilities (and there are of course many more) demonstrate once again the direct pick-up of a very concrete material source (the piano) and an equally direct quality that is often regarded as far more abstract—a tonal function (closure, cadence), or a social function (the testing of aural skills). These are quite obviously socially defined events—a musical function that arises out of the operation of a musical system, or a training function that arises out of an educational system—but they are nonetheless directly specified in the sounds themselves to a suitably attuned perceiver. Again it is not difficult to envisage straightforward empirical studies that could determine what the invariants that specify "cadence" or "aural test" might be (harmonic, rhythmic, textural, etc.).

As a final example, imagine a one-second burst from the middle of the tenor aria "La donna e mobile" from Verdi's opera *Rigoletto*—the kind of sound clip that you might get if you were scanning the stations on an analogue radio, and happened to pass through a broadcast of that opera. Again a range of responses to the question, "What do you hear?" might be imagined from the type of listener defined above: "vocal music"; "a burst of opera"; "over-the-top singing"; "an extract from Verdi's *Rigoletto.*" These again represent different perceptual capacities in some sense, as well as different musical values and kinds of musical experience. "Over-the-top singing" might be the response of someone whose previous experience was primarily of genres other than opera, who had sufficient exposure to this kind of singing to be attuned to its general type, but who had little interest in it or sympathy with it. All of the responses however, are consistent with the idea that auditory information can specify what may be regarded as abstract events—and certainly events that are overwhelmingly culturally defined.

Summary

This chapter has challenged those information-processing accounts of music perception that imply, or assert, that the cultural and ideological components of music are more abstract and remote than are its basic sensory and perceptual attributes, and that it is to the latter that listeners primarily respond. The ecological approach to perception offers an alternative view that gives a coherent account of the directness of listeners' perceptual responses to a variety of environmental attributes, ranging from the spatial location and physical source of musical sounds, to their structural function and cultural and ideological value. This entails extending ecological theory into the cultural environment, based on the principle that the material objects and practices that constitute culture are just as directly specified in the auditory invariants of music as the events and objects of the natural environment are specified in their corresponding auditory information. The conventions of culture, arbitrary though they may be in principle, are in practice as binding as a natural law. The directness of our perception of the world is not an inexplicable or "magical" reciprocity between perceiver and environment: it is the consequence of adaptation, perceptual learning, and the interdependence of perception and action. The advantages of this approach as far as music is concerned are that it places the emphasis on an investigation of the invariants that specify all of the phenomena that music is able to afford in relation to the diversity of perceptual capacities of different listeners; and that it offers a framework within which attributes of music that have previously been regarded as poles apart (from physical sources and musical structures to cultural meaning and critical content) can be understood together. This last point is based on a principle that recognizes the distinctiveness of different phenomena and the manner in which they may be specified, as well as the reciprocity between listeners' capacities and environmental opportunities (affordances), while asserting the commonality of the perceptual principles on which a sensitivity to these phenomena depends.