Toward a cognitive science of poetics: **Ānandavardhana**, Abhinavagupta, and the theory of literature

Patrick Colm Hogan

Roy Bhaskar has remarked that realists are unblushingly falliblist and historicist about science.* They feel no need to be uncritical and 'complimentary' about everything that passes for knowledge or is done in science's name… no reason to 'buy in' to shoddy science… Nor do they feel under any imperative to write the story of science Whiggishly as one long continuous success story-without blemishes or periods of stasis and even regression" (156). The point holds for the physical sciences, but is most obvious with respect to the human sciences and the study of the arts. Indeed, it is particularly clear to those of us who have rejected what Noam Chomsky (following Donald Hockney) calls the "bifurcation thesis" (roughly, the thesis that objects in physics, chemistry, etc., are real, but objects in linguistics, psychology, etc., are not [see Chomsky 16-22]), those of us who believe that there are ascertainable facts about the human mind, society, and art, facts open to systematic scientific investigation.

There are a number of problems with "Whiggish" history in any discipline. First of all, there are economic, political, and ideological factors which frequently vitiate our evaluations of theories. These factors are particularly evident when the indigenous sciences of a colonized people (e.g., classical Sanskrit poetics) are at issue, though they operate at all times and at all places. Secondly, even in evaluation that is motivated by and focused on genuinely scientific concerns, it is not always possible to understand or appreciate the explanatory power of a theory or

---

* This essay was originally published in *College Literature* 23.1 (February 1996): 164 (15 pages), © 1996 West Chester University. It has been reproduced here with the kind permission of the author and the original publisher. Hogan's Introduction to the collection of essays on comparative literature that constitutes that number of *College Literature* has also been included in this volume.

hypothesis. Sometimes, a full appreciation of the value of a theory in one area must await the development of our understanding in other areas. This is what happens when we refer to a theory as “ahead of its time”. Perhaps the best-known cases of this come from linguistics. Chomsky has repeatedly argued that the rationalist linguistics of the 17th century was far superior to the behavioral linguistics of the mid 20th century. However, this superiority was unrecognizable as the theoretical insights of the earlier authors far exceeded the state of psychological and related understanding in Europe during that period. Rationalist linguistic theory reached a sort of dead end in then-contemporary psychology, anthropology (which provided inadequate knowledge of non-European languages), etc. It was only after the intervening developments (not only in linguistic psychology, anthropology, etc., but also in mathematics and elsewhere) that the research project adumbrated in this work could really begin. More germane to our concerns, Paul Kiparsky has shown that generative grammar was first developed in India by the Sanskrit grammarians, preeminently Panini, in the fifth century B.C.E. (It is generally acknowledged that classical Indian linguistic science was vastly superior to the science of the West until very recently.) It was, however, subsequently forgotten, as there was no broader scientific context-and, one might add, no broader political context-in which its generative principles could be adequately appreciated. Its value only came to be recognized after the development of Chomskyan linguistics in the late 1950’s and 1960’s.

In the following pages, I would like to consider Sanskrit poetic theory in a similar context. Starting more than two millennia ago, and extending over a millennium, Sanskrit theorists developed an elaborate theory of poetics which was closely allied with linguistic study and, in my view, equally brilliant-though also equally ignored in the West. I would attribute our ignorance of this theory principally to imperialism and racism, but there are also factors relating to the theory itself. Specifically, this theory reached a culmination and a sort of theoretical impasse in the writings of Abhinavagupta. However, in developing the theory, Abhinavagupta anticipated current principles of cognitive science, much as Panini anticipated current principles of generative grammar. Now, with the aid of cognitive science, it is possible to recognize the value of Sanskrit poetics and to develop them beyond the impasse they had reached in the 11th century. There is, however, a difference between this situation and the cases of Panini or the Rationalist linguists. In the case of Panini and the Rationalists, current theories, in effect, rediscovered the insights of

the earlier writers, so that by the time these insights were rediscovered they had already become part of contemporary thought. Indeed, it was because these early principles had been independently rediscovered that the early theory came to be appreciated as an historical precursor. The situation with Sanskrit poetics is, in this respect, substantially different. In my view, we still have no adequate or even productive and plausible theory of aesthetic creation and response. What I hope to indicate, then, is not that the insights of Ānandavardhana, Abhinavagupta, and others have been rediscovered in cognitive science. Rather, I hope to indicate that, when redeveloped in the context of cognitive science as a broad theory of the human mind, the work of these writers provides us with a plausible and productive cognitive theory of poetics. In other words, the theory of Abhinavagupta does not anticipate a currently developed sub-field within cognitive science, but rather might serve to guide the development of such a sub-field.

More exactly, I shall begin with a broad outline of Sanskrit poetics from its beginnings through Anandavardbana (for a more detailed discussion, the reader may wish to consult a history of Sanskrit poetics, for example that of Gerow). Roughly speaking, Ānandavardhana sought to develop and systematize previous ideas in Sanskrit theory in order to provide an adequate description of poetic effect. Abhinavagupta took Ānandavardhana's descriptive ideas and sought to provide an explanatory framework for them; the second section is consequently devoted to the explanatory views of Abhinavagupta. In the third section, I turn to contemporary cognitive science, first presenting what I take to be a plausible, partial theory of the internal "lexicon" (or mental dictionary/encyclopedia), then going on to reformulate Abhinavagupta's views in terms of this theory.

Ānandavardhana and Suggestion

Histories of Sanskrit poetics generally divide early indian theories into two broad traditions: *alaṁkāra* and *rasa*. "Alaṁkāra" or "ornamentation" refers to a range of poetic devices and rhetorical figures from alliteration to metaphor. In itself, the *alaṁkāra* tradition is of limited interest, for it hardly extends beyond taxonomy (e.g., listing and describing the varieties and sub-varieties of figures such as simile). On the other hand, without this often tedious program of analysis and categorization, including its rudimentary literary semantics, Ānandavardhana could never have

formulated his seminal theory of *dhvani* or "suggestion", to which we shall turn in a moment

Ānandavardhana's work is, however, a culmination and synthesis, not only of *alaṁkāra* analysis, but of *rasa* theory as well. Rasa is usually translated as "sentiment." It is distinguished from *bhāva* or "emotion," to which it is, nonetheless, closely related. Specifically, *bhāva* is what we feel in ordinary life-love, sorrow, happiness, anger, etc. Rasa, in contrast, is what we feel in experiencing a work of art. It is akin to emotion, but not identical with it. (Hereafter, I shall use "emotion" rather than *bhāva* as the concepts are pretty much identical; I shall, however, most often use *rasa* rather than "sentiment" as *rasa* is in this context a technical term and thus has no precise ordinary language equivalent—just as "gravity" or "quark" or "morpheme" has no precise ordinary language equivalent.) Specifically, when I watch a romantic play, I do not actually love the hero or heroine (as I might love my spouse), but I do experience some sort of feeling. Moreover, this feeling is somehow related to love (in a way that it is not related to sorrow or anger, for example). Thus the Sanskrit theorists say that I am experiencing the "erotic *rasa*," not the emotion of love per se.

The earliest extant development of this theory is in the Naṭyaśāstra or Treatise on Dramaturgy attributed to Bharata-muni, but composed by a number of authors between the second century B.C.E. and the sixth century C.E. This seminal volume lists eight primary emotions and corresponding *rasa*s: love/the erotic, sorrow/the pathetic, and so on. (Lest this seem too restrictive, they acknowledge a wide range of ancillary feelings also.) Each literary work, in order to be aesthetically effective, was required to have one dominant *rasa*. This is not to say that other *rasa*s could not enter: they not only could, but must. However, any additional *rasa*s must function to further the dominant *rasa*. For example, suppose that the dominant *rasa* of a work is the erotic. Then it makes perfect sense to bring in the pathetic *rasa*. The pathetic may be part of the erotic, just as sorrow may be part of love. (Indeed, another way of thinking about the relation between *rasa* and *bhāva* is that the characters experience the *bhāva*s [say, love and sorrow] while the readers/spectators experience the *rasa*s [the erotic and the pathetic]). But the pathetic *rasa* in the work must operate to contribute to the erotic *rasa*. The same constraint holds if the

dominant *rasa* is the pathetic and the erotic is subsidiary. If the subsidiary *rasa* does not contribute to the dominant *rasa*, the experience of the reader/spectator will push in different directions and the overall aesthetic experience will be weakened; it will not be a satisfactory experience of the erotic or of the pathetic.

Again, Ānandavardhana combined the two strains of Sanskrit poetics developing them into a unified theory of aesthetic response, in part, based on the concept of *dhvani* or "suggestion." By the time of Ānandavardhana, Sanskrit theorists of "figures of thought" had isolated several varieties of meaning, literal and metaphorical. At this level, the distinctions with which they were operating were roughly the same as those operative in western poetics/semantics today. They had concepts of literal meaning, idiomatic meaning, and various metaphorical and related meanings (encompassing simile and other figures); they had also isolated a number of ways in which non-literal meanings could be manifested and interpreted (e.g., through some explicit marker, such as English "like" and "as" in similes). For example, "Varanasi is on the Ganges" does not literally mean that Varanasi is on the Ganges; it means (idiomatically) that Varanasi is on the bank of the Ganges; "Moon-faced beauty" does not literally mean that the woman's face was the moon, but that it was rounded and fair, and so on. In addition, Sanskrit theorists had isolated a variety of poetic and other implications (or, rather, non-logical "implicatures," to use Paul Grice's term), both literal and non-literal. For example, a message sent by a woman to her lover, "The lion, they tell me, does not prowl at the riverbank," would involve a complex metaphor/implicature to the effect that, if the woman and her lover meet at the riverbank, they will not be caught.

Ānandavardhana distinguished a number of varieties of *dhvani*, covering these non-literal meanings. However, he maintained that there was one sort of meaning which was not part of this typology and which is most appropriately the referent of the term *dhvani*. This *dhvani* proper, is not an idiom or metaphor (or simile, metonymy, etc.); nor is it an implicature (cf. Amaladass 92–93). Developing and systematizing Ānandavardhana somewhat, we may say that it is, rather, a non-paraphrasable "suggestion" of a word, phrase, sentence, topic, or (linguistically constructed) situation (see Amaladass 105). (Hereafter, I will use text, as a term ambiguous between word, phrase, etc.) To say that the *dhvani* of a given text is non-paraphraseable is not to say that one cannot say anything about it (as Ānandavardhana emphasizes [671]). Quite the contrary. One can say many things

about it. It is not unparaphraseable because of being ineffable. There are, rather, three reasons why *dhvani* is unparaphraseable. The first is that it is infinitely ramified. We can never enumerate all the suggestions, even all the relevant suggestions, of a given text (cf. Abhinavagupta *Locana*, 206.) This is not, however, a distinctive property of *dhvani*. For example, some metaphors are, roughly, paraphraseable in the sense that one can list all the relevant information conveyed by the metaphor (especially for those metaphors which are close to idioms). But many metaphors are not fully paraphraseable in this sense. And at least for larger texts, even the relations between literal meanings are not fully paraphraseable.

The second reason that *dhvani* is not paraphraseable is more important and more distinctive. Indeed, it indicates that, in a sense, *dhvani* is not even partially paraphraseable. In the language of analytic philosophy, the *dhvani* of a text cannot be substituted for the text with a preservation of truth value (cf. Abhinavagupta *Locana*, 81). Indeed, *dhvani* is usually not statable in the form of a proposition which might have or not have a truth value. Consider a metaphor partially explicit in Derek Walcott's Dream on Monkey Mountain: "Makak is a lion." As a metaphor, this means that Makak is brave, ferocious to his foes, etc. (For the purposes of the example, we can assume that it is not fully paraphraseable and thus that this list will not come to an end.) But note that each of these interpretations is substitutable for a lion, with a preservation of truth-value. "Makak is brave" and "Makak is ferocious to his foes" are just as true as "Makak is a lion." (I leave aside the issue of how to specify "metaphorical truth." Clearly, at least at a certain point in the play, this sentence is [ambiguously] true in a way that, say, "Makak is a gazelle" or "Makak is a unicorn" is not.) In contrast, consider the standard example of "a village on the Ganges." This phrase, the Sanskrit commentators emphasize, size, suggests holiness due to the holiness of the sacred river Ganges. However, there is no way in which the word "holiness" can be substituted for the phrase or for part of the phrase. Moreover, there is no way that this suggestion can be turned into a proposition. For example, the suggestion of holiness does not imply or implicate that the village itself is holy or that any of the villagers is holy. Returning to Walcott, we might say that the name "Makak" suggests a whole range of west African mythological motifs concerning the divinity of the monkey, and at the same time a range of contrary European attitudes. Again, these are not substitutable for "Makak." Nor do they imply or implicate any propositions.

Even at this level, the notion of *dhvani* seems to be substantially different from anything in the Western tradition. For example, even though "connotations" are not conceived as truth-preserving, they are typically conceived of as involving some sort of assertion. Western theorists may well say that "a village on the Ganges" connotes "holiness" but I suspect that they would unreflectively conceive of this as part of a (perhaps ironic) assertion—that the village is or should be holy, that the villagers have a special duty to be holy, etc. For this reason, it would seem somewhat odd to say that a range of Yoruba beliefs are a "connotation" of the name "Makak," for, again, none of these beliefs is asserted (or denied) by this "connotation."

But this is not all that distinguishes *dhvani* from common Western notions—and this brings us to the final and perhaps most important reason why *dhvani* is not paraphraseable. Crucially, *dhvani* is not purely semantic. it is affective as well. Specifically, in the strictest sense, Ānandavardhana tells us, *dhvani* is *rasadhvani*, the *dhvani* of *rasa*—not the intellectual implication of some sentiment, but the "suggestion" of a *rasa* as an affective experience (see Abhinavagupta's comments on Ānandavardhana in *Locana* 70). in other words, *dhvani* is not paraphraseable, most importantly because it is not solely, or perhaps even primarily, a meaning; as *rasadhvani*, the "truest" form of *dhvani*, it is an experience—along the lines of what we would call "a moment of tenderness" or a pang of sadness." It is, in short, an experience of *rasa*. A literary portrait of a village on the Ganges involves a full *rasadhvani* of holiness (or sacred peace) only if it gives us a feeling of that holiness (or sacred peace) along with the more narrowly semantic suggestion. Indeed, the feeling is all that is crucial. The name "Makak" communicates not only the idea of Yoruba beliefs, but a feeling as well, or a complex feeling: sadness (or the pathetic *rasa*) over their loss, anger (or the furious *rasa*) over their colonial denigration, etc.

In sum, aesthetic response is a matter of the experience of *rasa*s. These *rasa*s are evoked in a reader by words, sentences, topics, etc., presented in a literary work, but not through their literal meaning, or even through their various secondary meanings, as such. Rather they are evoked through the clouds of non-denumerable, non-substitutable, non-propositional suggestions which surround these texts. Finally, it is important to add that, in the overwhelming majority of cases, they are evoked without our having any explicit awareness of suggested meaning. We do not, in other

words, selfconsciously infer some semantic suggestion, then feel the *rasa*. Rather, we watch the play or read the poem without reflecting on either the meaning or associated feeling. Indeed, if we have to stop and reflect, we may not experience the *rasa*.

Abhinavagupta, Memory Traces, and Aesthetic Feeling

Abhinavagupta turned his attention away from the linguistic and related abstractions which had preoccupied even Ānandavardhana, focusing his attention instead on the human mind, specifically the mind of the reader or viewer of a literary work. The first step in Abhinavagupta's project involved the at least tacit recognition that the theory of *rasadhvani*, could not be understood as a theory of abstract linguistic structure. Rather, it only made sense as a theory of the way people respond to literature. In other words, *rasadhvani* had to be conceived in psychological terms.

Broadly speaking, Abhinavagupta was a transcendental realist whose philosophy involved a strong empiricist component. (A Hindu theologian himself, he was one of the harshest opponents of the Vedantist view that perceptual reality is illusion or maya; see, for example, Pandit 23.) His theory of mind was, consequently, realist (implicitly eschewing the "bifurcation thesis"). Predictably, it was broadly similar to other (Eastern and Western) theories of mind in isolating perception, memory, and other components or faculties. For our purposes, the most important part of Abhinavagupta's theory of mind is his theory of memory, both storage and recollection. For Abhinavagupta, all experiences—perceptual, cognitive, emotional, etc.—leave "traces" in the mind (see Aesthetic Experience 79). Drawing out the implications of Abhinavagupta's analysis, we may understand these traces as having two components: one representational, one emotive. The representational part may be perceptual (e.g., a visual image) or propositional (e.g., a certain fact). The emotive part is not the abstract recollection of one's having had an emotion—such a recollection would be propositional. It is, rather, a sort of re-experiencing of the emotion. It is not, in other words, remembering that one was sad or happy or frightened at a given time and place, but actually feeling again, in some degree, that sadness or happiness or fright. The point is most obvious with respect to strong emotions. For example, if one recalls a deceased friend or relative, one will have certain visual and other impressions, one will remember certain facts ("He had a

passion for *samosas*;" "Her favorite authors were Shakespeare and Tagore"), and one will probably experience again, in a more or less attenuated form, the sense of loss, the sorrow which one felt at his/her death.

Now these traces, Abhinavagupta tells us, are usually latent in our minds. At times they are fully activated—that is, it times we recall these memories, re-experience the emotions, etc. This is all commonsensical enough. But, Abhinavagupta continues, there are also times when these traces are neither latent nor fully activated—and these are the most crucial. In other words, there are times when these traces are in some sense activated, but are not brought into self-reflective consciousness. Or, more exactly, there are times when we are not self-reflectively aware of the representational content of the trace, and yet feel some hint of associated affect. This is, I think, common in our experience, though we may not immediately recognize the fact. I enter a building and am suddenly sad; I ask myself why and then recall an embarrassing or unpleasant event which occurred the last time I entered the building. Clearly, the memory had in some sense been activated (i.e., it was not fully latent) and clearly the associated affect had bled into my conscious experience (i.e., I felt sad), but the representational content of the memory was initially not conscious.

This sort of analysis is what allows Abhinavagupta to explain the phenomena isolated and described by Ānandavardhana. Specifically, Abhinavagupta indicates that *rasadhvani* operates in the following manner. Through the semantic *dhvani*—which, again, is not explicitly brought to consciousness—the literary work activates traces in the mind of the reader (see Abhinavagupta *Locana*, 81), but does not bring them into consciousness. Again, these traces may be activated by words, phrases, topics, etc.; thus stories of suffering will activate memories of suffering, stories of romantic love will activate memories of romantic love, and so on, both at a rather general level, and in various specific details (e.g., a wedding day, an estrangement and reconciliation), etc. Once these traces are activated, the associated emotions seep into consciousness (again, not as ideas, but directly as feelings). The experience of the *rasa* of a literary work is precisely the experience of these feelings. More exactly, developing Abhinavagupta's ideas, we may say that all speech involves the activation of traces, the consequent seeping of affect, etc. However, most often, these traces are activated in a haphazard and non-cumulative manner, or else they are fully recalled. What makes aesthetic experience distinctive is that such

activations are patterned; they are focused on traces of a specific type, which is to say, traces which bear to one another a certain similarity in both representation and affect (e.g., in being memories of death and feelings of sorrow). While any given activation is likely to produce only very limited, perhaps imperceptible, experiences of affect, this sort of repeated, patterned activation should result in a more pronounced and continuous experience. As Abhinavagupta puts it, "the basic emotion is put to use in the process of relishing [a work of art]: through a succession of memory-elements it adds together a thought-trend which one has already experienced in one's own life" (*Locana* 117; see also 182, "the relishing of beauty arises in us from our memory bank of mental states which are suitable, to the "basic emotions" of the characters; and Aesthetic Experience 112).

Abhinavagupta extends this idea somewhat further when he argues that aesthetic pleasure results from the "generalization" of emotion in *rasa*, which is to say, its removal from the self-interest which is part of the link between the affect and the representational content in memory traces. In other words, when we fully remember a trace, the emotion which we experience is tied to self-interest (e.g., to our own personal loss of a loved one). However, through literature, we experience a version of the affect removed from its direct link with any particular (egocentric) representation in memory, and thus at least partially removed from self-interest (see Abhinavagupta Aesthetic Experience, 86–87 and 96–97). In this way, ram may be re-defined as emotion isolated from such self-interest and may even be compared, in Abhinavagupta's view, with the experience of religious enlightenment or *mokṣa*, where such self-interest is entirely extinguished (see, for example, *Locana* 226). (Note that, for Abhinavagupta, the dominant *rasa* of a successful work always resolves itself into *śāntarasa*, the *rasa* of peace, a temporary and partial version of the endless and perfect peace which accompanies *mokṣa*.)

To some extent, I have explicated and developed Abhinavagupta's ideas in a manner coherent with and partially derivative from contemporary cognitive science. However, I have not, I believe, fundamentally altered his claims or basic concepts. Again, it is my view that current theoretical and empirical work allows us to understand those claims and concepts—as claims and concepts of a nascent science—more thoroughly than could Abhinavagupta himself. However, I have refrained thus far from giving a full blown, reductive account of Abhinavagupta's ideas within a broader theory of cognition. in the following section, I will attempt

something along these lines@ a translation of Abhinavagupta's theoretical premises into one version of a theory of mind which I take to be plausible in light of recent research in cognitive science. As I have already indicated, cognitive science has little if anything to add to Ānandavardhana's and Abhinavagupta's ideas about literary response. My concern, then, is merely to situate these ideas in a current theoretical framework, which is part of ongoing research programs (to advert to Lakatos's valuable concept). After undertaking this translation, I will consider briefly the degree to which we may reasonably take the cognitive model of the mind to reflect real psychological structures.

Cognitive Science and Rasadhvani

Cognitive science comprises a set of related, but competing theories of the structure and operation of the human mind. Due to the nature of Abhinavagupta's hypotheses, I will be concentrating on the storage and accessing or recollection of representational knowledge (i.e., knowledge about something, including memories, etc., as opposed to procedural knowledge, knowledge of how to do something). I will assume that all representational information is stored in and accessible through a single long-term memory "unit." This unit includes a broad range of information which, in everyday life, we would be inclined to divide between a dictionary (meanings of words), an encyclopedia (general facts about things), and a biography or personal archive (propositional and perceptual memories of individual history). (This is a controversial assumption. However, the following analyses could be rewritten—with some loss in elegance—in terms of three separate but closely interrelated units.) Following standard usage, I shall refer to this unit as the "lexicon." However, I should note that this term should not be taken to imply a sort of unidirectional organization and access: from dictionary/meaning through encyclopedia/belief to personal archive/memory. in fact, lexical "entries" (e.g., that for "monkey") are structured in such a way as to allow access not only from words to memories, but from perceptions (e.g., seeing or hearing a monkey) to words, from memories to beliefs, etc. indeed, just as there are words for which we have no beliefs or memories, and some for which we do not have even basic meanings ("I've heard that word before@ what does it mean?"), there are also uncategorized perceptions and perceptual memories ("I've seen that before@ what is it?"). Thus it is no more a lexicon in our ordinary sense than it is a "mnemonicon" or an

“aistheticon” (or an encyclopedia). Fortunately, however, these complications do not directly affect our concerns, which are with texts, and thus roughly fit the common sense model of organization from “meanings” to beliefs and memories. Were we to extend this study to aesthetic response in, say, the visual arts, however, we would have to be more careful not to allow the prejudicial nature of the term to mislead us.

In the present context, then, we may think of the lexicon as structured into “lexical entries,” which is to say, meanings, perceptions, and so on, clustering around lexical items, such as “monkey,” “*dhvani*,” “chant,” “compose” “saffron” etc. These entries are multiply cross-indexed, such that each entry is part of a number of sub-networks which allow access across entries. Thus “monkey” is linked with “ape” and “chimpanzee” in one network (of related species), “Africa” and “India” in another network (of habitats), etc. These networks may themselves be categorized by some lexical item (e.g., “primate”), but they need not be—though, of course, one can always construct a category for a given network (e.g., “habitats of monkeys”) and, indeed, we use such ad hoc categories to access relevant lexical information.

As should already be clear, our mental lexicon—a structure which exists only in individual people’s minds and thus varies somewhat from person to person—is different from a dictionary or encyclopedia, not only in content, but in structure. Perhaps, most importantly, it does not have a single strict organizing principle (e.g., alphabetical order). Rather, our mental lexicon can be re-ordered in a variety of ways to suit our accessing needs. For example, we may access words by first consonant (as in a dictionary, more or less) or by some other phonetic property, such as final syllable—if, say, we are writing an alliterative, rhymed poem. In most contexts, however, we are likely to access items by topic. One way of thinking about this is in terms of the sub-networks just mentioned. When discussing animals, we re-order our lexicon in such a way as to make, for example, “monkey” more directly or swiftly accessible than when discussing computer technology (see Garman 293). Indeed, this effect is probably cumulative. If we are discussing Africa and animals, then “monkey” will be more accessible than if we are discussing animals alone. This “extra” accessibility of related terms, concepts, etc., is standardly interpreted in terms of “priming.” In this view, the introduction of one item “primes” cross-indexed items (e.g., “monkey” will prime the entries for “ape,” “primate,” etc.). At this level, priming can be understood as a re-ordering of the lexicon, such that the primed,, items are those which are placed highest in the order of a lexical search. For example, when we read or hear the syllable “mon” at the beginning of a word, we

begin searching the lexicon for a "fit." When the topic is Wall Street, we will reach "money" first; when the topic is animals related to apes, we will reach "monkey" first. (In fact, the situation is somewhat more complicated than this, but the point is adequately valid for present purposes.)

On the other hand, "priming" has complex effects not only on access, but on other aspects of comprehension and response. In this way, it is not simply a matter of re-ordering the lexicon. There seems to be a change of status in the primed network. It is, in effect, brought out of long-term memory, though it is not accessed directly in consciousness. When speaking of "primates," I will have part of one lexical entry in consciousness (and parts of several lexical entries in "rehearsal memory," which may be understood as almost conscious or as circulating through consciousness). I also have a whole range of material stored in long-term memory (e.g., the meaning of "ambidextrous" visual images of my first day at school, and a vast range of other currently irrelevant stuff). A network of primed lexical entries is clearly in a different mental state from either the conscious/rehearsal material or the material stored in long-term memory. Consider the following sentence: "Animals commonly associated with Africa would include lions, tigers, elephants, and apes." When I am directly conscious of the word "ape," the rest of the sentence is in rehearsal memory, and various irrelevant entries (e.g., "ambidextrous") are in long-term memory. The lexical entry for "monkey," however, is primed and is therefore in a state different from "apes" and "lions" on the one hand, and "ambidextrous," on the other. The primed entries are, we might say, placed temporarily in a sort of "buffer" between long-term memory and consciousness (or between long-term memory and rehearsal memory), a buffer which operates in part to allow access with minimal search.

This, I think, gives us a basis for rearticulating and reunderstanding Abhinavagupta's notion that traces may be activated while not consciously recalled. Specifically, we may conceive of *dhvani* as lexical networks primed and stored temporarily in the memory buffer. Note that these networks may "decay" rapidly, which is to say, drop quickly out of the buffer if they are not repeatedly primed. (The rapid decay of priming effects is a well documented phenomenon: see, for example Garman 294.) However, when repeatedly primed, they would yield a pattern of unstated suggestions of precisely the sort we discussed above as defining *dhvani*.

To get a more detailed idea of how this might work, however, we need to consider not only the relations between lexical entries, but their internal structures as well. Here, as elsewhere, cognitive scientists are not unanimous in theories or even terminology. I will use the terms “schema” “prototype,” and “exemplum” to define three types of sub-structure within a lexical item. By “schema,” I mean a hierarchy of principles defining/explaining a lexical item. The hierarchy is based on “definitiveness” or “centrality” of the properties; the most central or definitive properties are at the top of the hierarchy, with increasingly peripheral properties listed in descending order. For example, being organic is more central to our conception of a human being than is having two arms—a person with one arm would count as a human being, but a statue with two arms would not. (This roughly recapitulates the distinction between the dictionary meaning and encyclopedic factual beliefs, but the distinction is one of degree, not of kind; some elements are “more definitive” than others, rather than one set of elements being “the definition” and another being “the empirical beliefs.”) As virtually all cognitive scientists emphasize, this hierarchy involves a number of “default” options. These are properties or relations which we assume to hold unless we are told otherwise. Thus “human” includes “having two arms” unless we are given information about a birth defect or amputation. By “prototype,” I mean a sort of concretization of the schema with all default values in place, including those which are relatively unimportant in our schematic hierarchy (cf. Johnson-Laird and Wason 342). Again, these vary somewhat from person to person, but probably all of us have a prototype of, say, “man” as having two arms, two legs, etc. My prototype, from what I can infer, has dark hair (i.e., neither bald nor grey-haired nor blond), is clean shaven, etc. Our prototype of a bird is generally recognized to be roughly equivalent to a robin (thus rather different from an eagle, and more different still from a penguin). By “exemplum,” I mean any specific instance of a category—for example, any man I know is an “exemplum” of man.

It is extremely important to note that all three sub-structures operate in cognition. We do not merely interpret or respond to the world and to other people in terms of abstract schemas. Indeed, we are more likely to understand and respond by reference to prototypes or particularly salient exempla (on the latter, see, for example chapter three of Ross and Nisbett). Indeed, this central cognitive role of exempla is crucial to literary response for a number of reasons. Most importantly, it appears to be the case that exempla—which are regularly primed and accessed along

with schemas and prototypes—often have considerable affective force. Clearly, many such exempla (e.g., the memory of a close relative's death) cause one to feel strong emotion when one recollects them consciously.

This, then, gives us a way of translating and redeveloping Abhinavagupta's notion that the affective component of memory traces is the source of *rasa*. Specifically, the *dhvani* of a text may now be understood as the schemas, prototypes, and exempla primed or placed in a buffer between long term memory and rehearsal memory. The exempla include not only representational content, but affective force. When an exemplum is sustained in the buffer, its affective force should lead to precisely the sorts of effect hypothesized by Abhinavagupta when he explained *rasa* in terms of memory traces. Specifically, we have every reason to expect that the affective force of an exemplum would bleed into consciousness without our being aware of its associated representational content, which is to say, the perceptual or propositional aspect of the exemplum. Or, rather, we have every reason to expect this when a set of affectively and representationally related exempla (e.g., sorrowful exempla of love in separation) are maintained in the buffer through repeated priming due to the patterned *dhvani* of a text. (It is important that the exempla are related both affectively and representationally because the representation, so to speak, gives definition to the affect; for example, sorrow over death is not the same as sorrow over love in separation.)

Interestingly, there is some empirical research on literary response which would seem to support these hypotheses. Specifically, Uffe Seilman, Steen Larsen, Laszlo Halasz, Janos Laszlo, and others have conducted experiments designed to determine the degree to which literary and non-literary texts spontaneously elicit personal memories from readers (see Halasz; Larsen, Laszlo, and Seilman, and citations in both articles). This research is in some degree questionable on methodological grounds—primarily, some obvious variables (e.g., subject matter) are inadequately controlled. In addition, the researchers appear to assume that aesthetic response is purely a function of the type of text being read and has little or nothing to do with the attitude or approach of the reader. Nonetheless, this research is broadly consistent with the present analysis, and provides at least some prima facie support.

The Nature of Mind and the Claims of Theories

Having said all this, however, I should step back for a moment and say something about the precise nature of my claims. As mentioned at the outset, I am a realist in the sense of someone who believes that there is a fact of the matter about how the world is—and about how the mind is. However, I am not a realist in the sense of believing that the best theory available at any given moment is likely to be true or even likely to be talking about real things. Indeed, I do not even believe that the best theory should necessarily be confined to a discussion of entities which we can presume to be real. As Zenon Pylyshyn has indicated, every actual psychological theory is realist at some level; the only question is: at what level do we claim that our theoretical posits and principles represent real entities? Pylyshyn notes that our claims may range from the barest input/output realism to a full blown realism extending all the way to details concerning posits, etc. (see "Rules" 236 and Computation 87–89). Consider linguistic theory. In the input/output case, a theorist claims only that the theory will generate all and only grammatical sentences of a given language. He/she does not claim that the way the theory generates them is the same way in which they are produced by speakers; he/she does not claim that the rules of the theory have "psychological reality" in the minds of speakers of the language. In the "full blown" case, in contrast, the theorist does claim that the rules of the theory and the rules in speakers' minds are identical—down to specific notational details. Very few theorists make either of these extreme claims, though some are closer to the input/output extreme, and others are closer to the full blown extreme. I consider myself to fall pretty squarely in the middle.

Specifically, I believe that the mind should ultimately be understood as subjective/experiential (as I have argued at length in the first and second chapters of On Interpretation). Many theories impute to the mind a quasi-objectal character, most obviously when they posit mental items which are in principle inaccessible and thus in principle non-subjective (as is the case with Chomsky's underlying structures), but also when they speak of mind in terms of internal objects and forces. When theorists objectify the mind in this way, they base their theories on what I believe to be false premises. This is not to say that I think theorists should stop doing this. Daniel Dennett has pointed out that, when faced with extremely complex material phenomena, we frequently adopt an "intentional stance" toward the phenomena, attributing an intention in order to allow ourselves to think productively about the object or situation (see Dennett 236–39). Inverting Dennett, I

do not see any reason that we could not adopt an “objectal stance” toward subjectivity in order to construct models by which we may better understand this highly complex phenomenon.

On the other hand, model—whether intentional or objectal—should in my view seek to preserve the broad structural relations of the original phenomena. it is this structural correlation which I would claim for the model of mind presented above. In other words, I would not claim that there is a lexicon per se, a buffer, etc. However, I would say that there is a subjective distinction which preserves the structural relations which define and distinguish these entities. Put simply, our subjectivity could be such that we have one thing under consideration and, after that, all our memories are equal. If this were so, then our subjectivity would not preserve the structural relations which distinguish long-term memory and the buffer. My contention, then, is that there are such structural correlates—and that the theories of Ānandavardhana and Abhinavagupta can be revised and developed in terms of these structures in such a way as to provide a fruitful and plausible account of aesthetic experience which can be part of an ongoing research program.

In any event, however we decide ultimately to construe theories of the mental lexicon, I hope that the preceding discussion of Sanskrit poetics and cognitive science has demonstrated the value of the former and its remarkable congruence with the latter. More importantly, I hope that the explication, synthesis, and extension of these ideas has led to a more adequate understanding of literary response and that the resulting theory might productively be incorporated into the empirical research programs now flourishing in the discipline of cognitive science.

Works Cited

Abhinavagupta. *The Aesthetic Experience According to Abhinavagupta*. Ed. and trans. Raniero Gnoli. 2nd ed. Varanasi: Chowkhamba Sanskrit Series, 1968.

_____. *The Dhvanyāloka of Ānandavardhana with the Locana of Abhinavagupta*. Trans. Daniel Ingalls, Jeffrey Masson, and M. V. Patwardhan. Cambridge: Harvard UP, 1990.

Amaladass, Anand. *Philosophical Implications of Dhvani: Experience of Symbol Language in Indian Aesthetics*. Vienna: De Nobili Research, 1984.

Ānandavardhana. *The Dhvanyāloka of Ānandavardhana with the Locana of Abhinavagupta*. Trans. Daniel Ingalls, Jeffrey Masson, and M. V. Patwardhan. Cambridge: Harvard UP, 1990.

Bharata-muni. *The Naṭyaśāstra*. Trans. Manomohan Ghosh. Calcutta: Manisha Granthalaya, 1967.

Bhaskar, Roy. *Reclaiming Reality: A Critical Introduction to Contemporary Philosophy*. London: Verso, 1989.

Chomsky, Noam. *Rules and Representations*. New York: Columbia UP, 1980.

Dennett, Daniel. *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge: MIT P, 1978.

Garman, Michael. *Psycholinguistics*. Cambridge: Cambridge UP, 1990.

Gerow, Edwin. *Indian Poetics*. Wiesbaden: Otto Harrassowitz, 1977.

Grice, Paul. "Logic and Conversation." *Studies in the Way of Words*. Cambridge: Harvard UP, 1989.

Halasz, Laszlo: "Emotional effect and reminding in literary processing." *Poetics* 20 (1991): 247–72.

Hogan, Patrick Colm. *On Interpretation: Meaning and Inference in Law, Psychoanalysis, and Literature*. Athens: U of Georgia P, 1995.

Holland, John, Holyoak, Keith, Nisbett, Richard, and Thagard, Paul. *Induction: Processes of Inference, Learning, and Discovery*. Cambridge: MIT P, 1986.

Johnson-Laird, Philip N., and P. C. Wason, eds. *Thinking.. Readings in Cognitive Science*. Cambridge: Cambridge UP, 1977.

Kasher, Asa, ed. *The Chomskyan Turn*. Oxford: Blackwell, 1991.

Lakatos, Imre. "Falsification and the Methodology of Scientific Research Programmes." *Criticism and the Growth of Knowledge*. Ed. Imre Lakatos and Alan Musgrave. Cambridge: Cambridge UP, 1970. Pp. 91–95.

Larsen, Steen, Janos Laszlo, and Uffe Seilman. "Across Time and Place: Cultural-Historical Knowledge and Personal Experience in Appreciation of Literature." *Empirical Studies of Literature*. Ed. E. Ibsch, D. Schram and G. Steen. Amsterdam: Rodopi, 1991.

Pandit, B. N. *Essence of the Exact Reality or Paramārthasāra of Abhinavagupta*. Delhi: Munshiram Manoharlal, 1991.

Pylyshyn, Zenon. *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge: MIT P, 1984.

_____. "Rules and Representations: Chomsky and Representational Realism." *The Chomskyan Turn*. Ed. Asa Kasher. Oxford: Blackwell, 1991. Pp. 231–251.