Stephen Prince

The Discourse of Pictures
Iconicity and Film Studies

In its analysis of images, film theory since the 1970s has been deeply indebted to structuralist and Saussurean-derived linguistic models. Indeed, it would be difficult to overstate the depth and importance of this relationship. As Robert Stam has noted, “Semiotics in general, and film semiotics in particular, must be seen... as local manifestations of a more widespread linguistic turn.” To speak, for example, about “reading” a film, as many film analysts now do, irrespective of the critical methodology employed to generate the reading, is to index and emphasize this lineage. Like books, films are regarded as texts for reading by viewers or critics, with the concomitant implication that such reading activates similar processes of semiotic decoding.

But does it? To what extent are linguistic models appropriate for an understanding of how images communicate? Film theory since the 1970s has tended to place great emphasis upon what is regarded as the arbitrary nature of the signifier-signified relationship, that is, upon the purely conventional and symbolic aspect of signs. What this focus has tended to displace is an appreciation of the iconic and mimetic aspect of certain categories of signs, namely pictorial signs, those most relevant to an understanding of the cinema. This stress upon the arbitrary nature of semiotic coding has had enormous consequences for the way film studies as a discipline has tended to frame questions about visual meaning and communication.

Our purpose here is to examine some of these consequences and to see how well they square with the observable evidence about how viewers perceive and comprehend cinematic sequences. We will see that current film theory, tracing its lineage from Saussure, Althusser, and Lacan, has constructed accounts of the ways in which film transmits meaning that are, in certain important respects, counter to the observable skills, abilities, and reactions of real-world viewers. The viewer, as theorized in these accounts, differs substantially from his/her real-world counterpart. We shall emphasize some of these discrepancies in order to suggest a reorientation of theoretical focus. In short, a renewed attention to the iconic, mimetic nature of pictorial signs is warranted, so that our theories might become more sensitive to the unique, constitutive features of pictorial—as opposed to linguistic—modes of communication. (It is important to note that other currents exist in contemporary film theory. Pier Paolo Pasolini and Peter Wollen, for example, have offered analyses of cinematic signs that are inflected rather differently than the Saussurean-inspired accounts that this essay examines.) More than 20 years ago, Wollen stressed the importance of paying attention to the iconic aspect of pictorial signs. Borrowing from Charles
Peirce’s triadic model of the sign, he argued that in cinema, iconic and indexical aspects are more powerful than symbolic, and he pointed out that semiotologists have neglected the subject of iconic signs because they are biased in favor of conceptions of signs as arbitrary and symbolic.

Let me state very clearly at the outset that this suggested reorientation is not intended as a substitute for the very real and necessary work on the role that culture plays in film spectatorship and interpretation. Culture-bound attitudes do indeed reflect the content of film narratives, along with their stylistic visualization, at the point of production and, again, through the inferences viewers draw from those narratives. The position this essay takes is not intended as a replacement either for the critical work of ideological analysis and interpretation, or for inquiries into how viewers’ judgements about the nature of a social world depicted in film may be shaped by elements of cinematic structure. But, in the interest of conducting this kind of interpretive work, it is important to have a clear sense of where cultural variables do and do not enter into the production of meaning by the cinematic image. Only when we have a good understanding of those aspects of visual signification where culture plays a less determinative role are we likely to be able to construct a clear portrait of where it does play a role. Thus, this essay certainly does not argue against the role of cultural analysis in film interpretation, but it does attempt to suggest some reasons for caution when employing linguistic categories and models of cultural relativism in film analysis. It is not that there are no cases in which they might profitably be applied, but rather that there are some significant aspects of visual communication to which they probably should not be applied.

**The Linguistic Turn in Contemporary Film Studies**

To begin, it will help to briefly trace the linguistic influence on contemporary film theory, after which we may question whether it supplies an adequate account of images or even, in certain crucial respects, of language itself. It is well known that Saussure’s account of the sign as having an arbitrary and unmotivated connection between its structural components has been directly taken over by many film theorists. As Philip Rosen has pointed out, “One effect of the argument for the basic conventionality of cinematic images was to open the way for a utilization of the ideal of difference in cinematic signification.” Stressing the signifier as a differential construction enabled film theory to emphasize communication as discourse, as a culture-bound activity, relative to and differentially patterned by the unique social worlds of diverse groups of interactants. Signs, whether linguistic or cinematic, were viewed as culturally instantiated: “Sign systems don’t produce meaning outside of the social and cultural context from which they have developed.”

This concept of the sign enabled theorists to explicate many aspects of cinematic coding, from discrete optical devices like dissolves or wipes to more complex structures such as shot/reverse-shot cutting, subjective images, and other aspects of point-of-view editing. Viewing these devices as symbolic codes permitted theorists to emphasize the construction of cinematic discourse, that is, the deployment in film of an elaborate semiotic system whose address, and effects, could be comprehensible in Althusserian-Lacanian terms as the interpellation of subjects. Using a symbol system like language, in this view, entails being positioned—socially, ideologically—in and by the categories which that system has helped create. Thus film, like language, could be comprehensible as discourse, as the creation of apparent meaning where only true relations of difference prevail (due to the arbitrary nature of the sign and the consequent need for it to receive definition only in relation to what it is not, i.e., to all other signs). As Althusser and Lacan (and the film theory they inspired) emphasized, these deceptively real constellations of apparently fixed meaning could be an excellent site for ideology and for imaginary conceptions of the self to take root. Viewed as discourse, cinema assumed a symbiotic relationship with ideology, becoming an effective vehicle for its transmission. The work of film theory became increasingly focused on deciphering the ideology at work inside the cinema’s deceptive and transparent appearance of reality. That appearance of reality was, furthermore, suspect for having ideological effects (e.g., naturalizing that which is historical or cultural, etc.) and for creating ideal and false subject unities. Viewed in these terms, film history is the history of discourse, and the relation between film and the world is a matter of representational convention. As Stephen Heath has written, “That reality, the match of film and world, is a matter of representation, and representation is in turn a matter of discourse. . . . [I] In this sense at least, film is a series of languages, a history of codes.”

Furthermore, an emphasis upon differential and relational signifiers has entailed a denial, or, at best a
suspicion, of reference. If signs are arbitrarily related to what they represent, then meaning is at best provisional, at worst illusory or ideological. Representation which is non-arbitrary tends to be construed by film theory in terms of a relationship of identity between sign and referent, thus generating a dichotomy of arbitrariness/identity. Semiotic representation is either a matter of arbitrary coding or of identity and transparency, with the latter condition being construed negatively in terms of illusion and error. Identity is regarded as a deceptively false universal, as a part of what is sometimes called the ideology of the visual. "Reading encounters the text as a relation of difference not identity." Take a step further, this view regards empirical knowledge about the world as being thoroughly mediated by signs, as itself discourse. "The empirical . . . is not 'the real' but the product of the discourses of the dominant ideology." Since the "real" embodies the false universals of ideology, the relations among cinematic signifiers tend to be seen as cultural (and therefore symbolic), rather than as iconic or mimetic. The spectator's understanding of the cinema is therefore explained as a matter of cultural conditioning and learning. "The spectator chains together the film's signifiers on a cultural grid of intelligibility—an ensemble of assumptions and presuppositions about the 'real'—into an account that makes the film socially intelligible."

Note the assumptions here. Language imposes a system of relational distinctions upon the world, creating culture from the real. One makes entry into language, or, in Lacanian terms, into the Symbolic order, learning the culturally patterned distinctions and, in the process, being interpellated as a subject. Film is akin to language by virtue of employing relational, differential signifiers. Comprehension of the cinema, then, should likewise be predicated upon cultural conditioning, upon the apprehension of "cultural grids of intelligibility." But, as we will see, the spectator's understanding of cinematic images seems more immediately exploitable in terms of mimetic, referential coding rather than via the chains of displaced, arbitrary, and relational meaning in prevailing theories. In other words, this understanding is more a matter of recognition than translation. We will return to this point. Moreover, iconic representation is appropriately understood in terms of degrees of resemblance rather than the all-or-nothing terms of arbitrariness or identicality. A photograph, for example, exhibits a higher degree of iconicity than a line drawing.

Transposed in film theory to units of cinematic structure, the Saussurean view of the sign as an arbitrary and, taken in isolation, meaningless unit has allowed current theory to construct an extended analogy with language. Although no one any longer searches for the filmic equivalent of such linguistic features as the period or the comma, the debt to Saussurean linguistics is apparent in the axioms that the connections between cinematic representation and the world are, in all important respects, a matter of historical or cultural coding and convention, that is, that filmic representation is a matter of symbolic rather than iconic coding and that a viewer, rather than perceiving a film, "reads" it.

Problems with Linguistic Relativity

We shall explore some discrepancies between these axioms and the observable evidence about how viewers perceive and process cinematic images. First, however, it will be helpful to discuss some reservations regarding the alleged arbitrariness of the linguistic sign and the uses to which film studies has put this concept. As Devitt and Sterelny have recently pointed out, emphasizing the sign as a relational entity, defined through relations of difference, presents a problem for meaning. This emphasis fails to specify how meaning may arise or even how lexical borrowing from other language systems may occur. When speakers borrow terms from another language, they are often doing so in response to a perception that some thing or condition exists that needs a name, although at present it has none or is insufficiently labeled. In such a case, linguistic skills are deployed in response to non- or extra-linguistic perceptions, a condition which models of linguistic determinism have a hard time accounting for. Furthermore, rejecting reference, as Saussurean models do, mystifies language acquisition.

In a Saussurean view, where meaning is defined through relations of difference (e.g., Jonathan Culler's well-known view that understanding the color brown entails grasping the relation between brown and what it is not, i.e., all other colors), it is difficult to see how a child ever learns language. As Devitt and Sterelny note, "We want to say that a child begins by learning a minimal vocabulary and a few rudimentary syntactic rules. The child continues by extending these rules. On a structuralist picture of language, we cannot say this. Vocabulary does not remain constant across changes in the system. Each time the child changes the system, everything changes. Language learning cannot be represented as a cumulative process.'"
Part of the problem here is that to specify meaning, one at times has to step outside the language system, and poststructural methodologies have been most unwilling to posit the possibility of doing this. In many accounts, one cannot get outside representation at all because it is, by definition, determinative of human thought and experience. It is commonly maintained that "there is no absolute moment without signs, without language, without, in other words, a whole host of mediations between seeing, experience, and knowledge." Stephen Heath has noted that the referents of cinematic images exist only in discourse, as representations already constituted by history and culture. "The represented a discourse produces is grasped, realized, exists as such in the particular discursive process of representation, and it is this that needs first and foremost to be interrogated." This view—that pictorial representations and, beyond them, patterns of human and social organization are the expressions of discursive relations and positions—is a reformulation of the famous Whorf-Sapir hypothesis which suggested that radically different language systems might organize the world in unique ways for their users. The Whorf-Sapir hypothesis is a statement of extreme linguistic relativity, arguing for the influence of features of linguistic organization upon the perceptual habits of a community. It is a relativistic view because it argues against the possibility of semantic or perceptual universals. As languages vary, so do the realities they construct. The Whorf-Sapir hypothesis has been a very seductive and influential one for film studies. Bill Nichols has noted the relevance of this hypothesis for poststructural film theory. "Post-structuralist work does not regard language or, by extension, film as the neutral means by which we understand ourselves, others, and our world. Rather, it draws on versions of the Whorf-Sapir hypothesis, which describes a world constructed in, by, and through language."

Benjamin Lee Whorf's study of Indian languages led him to emphasize how different languages may segment the world for their community of speakers in divergent ways. There are some data to support a limited view of this hypothesis. Language undoubtedly does modulate our experience of the world. Yet there is little evidence to support the extreme relativity of the Whorfian view (i.e., that language determines both thought and perception) and some clear evidence to counter it. Advocates of linguistic relativity, or discourse relativity, often support their position with reference to color terminology, i.e., to the fact that different cultures have varying numbers of color terms to designate locally important attributes, such as snow for an arctic community. In suggesting that representational images, like other texts, rely upon culturally determined codes, Nichols, for example, has employed this notion of linguistic relativity, writing that "to not know the perceptual codes maintained by a given culture is tantamount to being an illiterate infant wandering through an unintelligible world. (An example would be the utter inability of most members of non-Eskimo cultures to distinguish the dozens of different kinds of snow for which the Eskimo has separate words)." Analogously, Mas'ud Zavarzadeh has asserted that "we understand colors not because we respond to them directly through our sensory organs, but because the responses of our sensory organs are made meaningful for us by the language. Different languages make sense of this physical continuum in startlingly different and dissimilar ways. . . "

In contrast to the Whorf-Sapir hypothesis, however, a classic crosscultural study of color terminology found respondents plotting the same range of basic colors on a Munsell color chart irrespective of the labels furnished by their culture. Experimental data were obtained from speakers of 20 languages and were supplemented with historical information on 78 additional languages. The researchers found a universal inventory of 11 basic color categories across all languages studied and a fixed evolutionary sequence governing the order in which languages added new basic color categories. Writing that the allegation of arbitrariness in the way languages segment the color spectrum is "a gross overstatement," the researchers concluded that "the referents for the basic color terms of all languages appear to be drawn from a set of eleven universal perceptual categories, and these categories become encoded in the history of a given language in a partially fixed order. There appears to be no evidence to indicate that differences in complexity of basic color lexicons between one language and another reflect perceptual differences between the speakers of those languages." Different languages can make more or fewer distinctions according to environmental or cultural needs but cannot override the biological basis of color perception.

The evidence furnished by Berlin and Kay on color terminology fails to support the extreme linguistic relativity hypothesis that has been so influential for film studies, and suggests instead the importance of retaining a concept of referentiality outside language to which language may respond. This example can be a heuristic one for film theory. Reopening a space for principles of referentiality and iconicity could be a very
useful development, enabling our theories to move closer to observable real-world evidence about how viewers understand the cinema. Thus, with respect to pictures, film theory might ask if there is a nonlinguistic, even biological, basis on which visual communication might rest. Investigation of this question may help provide a more secure sense of the methodological limits constraining the equation of cinema with discourse. Julian Hochberg has recently made this point, noting, “only when we know where the line lies that separates . . . lower, mandatory cognitive processes from the higher, elective functions, can we sensibly formulate explanations—linguistic, psychoanalytic, and so on—that rest on cultural determination.”

One final point to note about the concepts of linguistic relativity and arbitrariness operative in current film theory is the extent to which accounts of the cinema as discourse, analogized with language, construct the analogy based on limited features rather than seeing language in more comprehensive terms. The linguist Charles Hockett, for example, seeking to distinguish human language from animal systems of communication, formulated a set of 13 design features uniquely characterizing human language. Arbitrariness of the sign was only one of them, and many of the others, such as prevarication, displacement, and abstractness, point to clear differences between the communicational capabilities of language and pictures, suggesting that analogies with language may not be the best way of explicating pictorial communication. The design feature of displacement, for example, denotes the capability of the language-system to create messages that refer to a time and space outside of the immediate communicational situation. Pictures, by contrast, lack tense and other aspects of syntax that can be used to establish remote temporal or spatial conditions (the dissolves, wipes, and odd music used to signal flashbacks in Hollywood narratives are a less powerful and flexible means of approximating this ability). Furthermore, as several scholars have pointed out, pictures cannot express negatives. These and other differences between pictorial and linguistic messages point toward distinctly different communicational modalities. Interestingly, Christian Metz also noted some of these fundamental distinctions between linguistic and pictorial modes of communication. He pointed out that cinema lacks the double level of structure in language (i.e., the morphemic and phonemic levels) and is a one-way system of communication, unlike language, where senders and receivers are interchangeable, and he concluded, therefore, that the concept of a language is probably inapplicable to film. Despite these objections, Metz nevertheless went on to explore film structure as parole, in terms of a taxonomy of segment types.

If cinematic meaning is to be developed from a theoretical basis originating with a Saussurean conception of the relativity of the sign, film theory sometimes seems a little imprecise about the functional and at times invariant nature of communicational rules. As noted, film theory tends to view cutting patterns, camera positions, and even perceptually based images as culturally relative yet syntactically precise conventions. But, with respect to language (and perceptual images), Noam Chomsky points out that “having acquired the system of language, the person can (in principle) choose to use it or not, as he can choose to keep to or disregard his judgements concerning the position of objects in space. He cannot choose to have sentences mean other than what they do, any more than he can choose to have objects distributed in perceptual space otherwise than the way they are.” In other words, if the assumptions of the Saussurian view, as applied to pictures, are correct, we should find more cross-cultural variation on basic picture recognition tasks than we in fact do. We will explore this point in more detail shortly.

**Visual Grammar or Narrative Context?**

Chomsky’s observation raises the issue of grammaticality, which should be central to questions about whether film structure operates like a language. As noted, Metz’s presentation is somewhat ambiguous with respect to grammaticality, but this is an issue that cannot easily be avoided for accounts that seek to analogize film and language. Film theory has identified a number of syntactic elements (e.g., point-of-view editing employing perspectively based images, suture) which are thought to operate as codes with discernible effects across a range of individual films, thus constituting a kind of cinematic grammar. But the test of ungrammaticality is not often applied (that is, judgements about the syntactic correctness of given filmic constructions). The most thorough and complete exploration of the application of linguistic principles to problems of film structure is found in the work of John Carroll, who has sought to apply Chomskian transformational-generative grammar to the cinema. Carroll identifies a series of principles underlying cinema grammar violations of which ostensibly constitute cinematic equivalents of grammatical errors and
allegedly give rise to sequences that viewers will judge to be confusing or unfilmic. These are essentially continuity editing rules, such as the injunction that if an actor looks and reacts to anything off-frame, the next shot will be interpreted by viewers as a subjective shot. Hitchcock describes a sequence from *The Birds* where Hitchcock seems to violate this principle, and he remarks that viewers will find this confusing. It is the scene showing Tippi Hedren waiting outside the schoolhouse while, behind her, the birds gather en masse on the playground jungle gym. Hitchcock cuts between a shot of Hedren looking off-frame and a shot of the gathering flock, as if to imply (by virtue of the cutting pattern) that she must see it.

In a recent and thorough review of the issue of visual literacy in film and television, Messaris points out that, while Carroll is technically correct, the salient point is that the confusion (not shared by all viewers) is only momentary, because the surrounding contextual-narrative information makes it clear that the subsequent shot is not a subjective one. In other words, given the narrative context, it would be implausible to assume the character sees the birds without reacting because their deadly nature has already been established in the story. Messaris points out that one reason why notions of grammaticality are difficult to apply to film is that narrative context often overrules code. "The role of formal conventions in conveying a movie’s meaning is generally subordinate to conventional standards of plausibility or probability (‘conventional’ in the sense that the film-maker must be able to assume them of her/his audience). . . . [T]he viewer’s interpretation of edited sequences is largely a matter of cross-referencing possible interpretations against a broader context (i.e., the larger story in the movie itself, together with corresponding situations from real life and other movies), rather than a matter of ‘decoding’ formal devices (e.g., an off-screen look followed by a cut to a new shot). . . . Interpretation is driven by the narrative context, not the code."28

Another example should make this clear, although this instance does not involve an alleged violation of grammar. In *The Searchers*, after Ethan and Marty discover that Lucy has been killed by her Indian abductors and Brad, her lover, is killed while charging into the Indian camp, the searchers ride into a snowy landscape where Marty despairs of ever finding Debbie. Ethan, in reply, makes his famous speech about being a critter that just keeps coming. Fade out and fade in on the Jorgensens’ ranch as Ethan and Marty ride up. As a symbolic pictorial device that has no clear analogy in real-life visual experience, the fade operates in accord with its conventional usage to bracket and separate sections of the narrative, assisting the viewer’s parsing operations. But it is the contextual narrative information that provides the salient cues about the precise nature of the narrative shift. At the Jorgensens’, Ethan and Marty inhabit a different landscape. The absence of snow indicates a seasonal shift, and Jorgensen tells Ethan that his letter about Brad’s death arrived the previous year. The viewer’s extra-filmic knowledge of seasonal patterns and, most importantly, the dialogue between Ethan and Jorgensen establish the precise nature of the time shift in a way that the purely visual code—the fade—cannot. The dominance of narrative context over code explains why it would be equally permissible—though not in a film of the 1950s—to signal the temporal shift with a straight cut. With redundant informational cues available in the narrative, the viewer should be able to follow the flow of events, permitting a degree of interchangeability in the use of codes. Thus, image patterns may not need to be as rigidly sequenced as the concept of a “grammar” might imply, and the ostensive violation of grammatical rules may carry less significance (for the viewer’s interpretational abilities) than is sometimes thought, as the violations of the 180-degree rule in contemporary productions should indicate.

**Empirical Evidence about Image Comprehension**

The foregoing discussion has suggested that the linguistic orientation of film theory may not be the most efficacious for dealing with visual meaning or cinema structure. Currently, the discipline is undergoing some profound shifts as the Saussurean and Lacanian accounts are contested by alternative formulations advocating perceptually or cognitively based approaches. The cognitive turn in contemporary theory has opened a space for renewed appraisal of perceptual evidence that runs counter to notions of film as language and their logical consequence—that viewers must learn to interpret pictorial displays. Reviewing the results of empirical research on the perceptual and cognitive processing of visual images and narratives will help clarify some of the problems inherent in using concepts of arbitrary or relational signifiers to explain principles of cinematic meaning and communication.

Contemporary notions that film is analogous to a language are consistent with the speculations of film
theorists dating from virtually the beginning of cinema history. Boris Eichenbaum, for example, noted that “Film language is no less conventional than any other language. . . . Cinema has not only its ‘language,’ but also its ‘jargon,’ rather inaccessible to the uninitiated.” Béla Balázs suggested that, to be comprehended, the new “form-language” of cinema required of its viewers a new sensibility and understanding, without which novice viewers would be baffled. However, to posit language-based modes of cinematic communication is to implicitly raise the issue of visual literacy by implying that a period of tutoring would be logically necessary in order to gain interpretive mastery of the cinematic vocabulary (learning, for example, that a subjective shot represents the view of the absent character or that the portions of an event elided by continuity editing nevertheless still occurred within the narrative). The assumption that untutored or inexperienced viewers would encounter difficulties making sense of unfamiliar images finds some apparent confirmation in the existing anecdotal reports of viewers’ bewildering first-time encounters with motion and still pictures. However, these difficulties are often due less to an inherent inability of naive viewers to see the pictorial objects than to a first-time encounter with a novel recording surface (e.g., paper in the case of still pictures) or a culturally unfamiliar depth cue used to render an object in abstract terms (e.g., linear perspective suggesting a straight road in a line drawing). Moreover, even where anthropologists and field workers report difficulties with picture perception, respondents typically are able to integrate figure-ground relations rather quickly, resulting in correct picture perception. Viewers who find picture perception completely impossible under any condition are quite rare and probably anomalous. Convincing evidence about inherent human abilities to perceive pictures is available from a classic experiment in which the researchers prevented a child (their own) from birth from seeing pictures until the age of 19 months, when it began to actively seek them out. The child learned his vocabulary solely through the use of objects and received no training regarding pictorial meaning or content. He was nevertheless able to recognize, when tested with a series of 21 two-dimensional line drawings and photographs, the series of pictured objects (people and things familiar in his environment). The researchers concluded that the results indicate the existence of innate picture perception abilities and that, if there are allegations of cultures or viewers lacking these, it cannot be a matter of not yet having learned the language of pictures.

Empirical research indicates that one basis for these innate abilities probably lies in the activation by pictures of perceptual skills (object recognition, depth perspective, etc.) developed in real-world experience, skills which are then transferred to pictures. Scholars have commonly emphasized that pictures replicate a series of monocular depth cues used in real-world experience for inferring information about the positioning of objects in physical space (e.g., overlap, texture density gradients, shading, and, in motion pictures, motion parallax). Picture perception, then, seems clearly based on 3D spatial skills transferred to the 2D representation, and one would expect that the more complete the set of cues employed by the picture regarding the spatial layout of the depicted scene, and the more culturally familiar the depicted objects, the greater the ease of recognition. If naive viewers can perceive still pictures, there is no logical reason for inferring that they would have difficulty doing so with moving pictures, especially since movies will supply the additional real-world depth cue of motion parallax.

But what about the role of culture in visual perception? As we have seen, culture-bound determinations are very important in language-based film theories which rely on arbitrary, relational signifiers. The move to empirical, perceptually based evidence might be unwelcome for linguistically based accounts of pictorial meaning because it can be seen as replacing cultural categories with biological (and therefore ideologically suspect) ones, and because it may suggest that pictorial and linguistic modalities are distinct from each other. With respect to the role of culture in picture perception, some evidence does exist for varying cultural susceptibilities to different visual illusions (the Mueller-Lyer illusion and the Sander parallelogram), perhaps due to variations in the physical environments of different cultures, but these do not seem to impact basic picture perception abilities.

In a recent summary of the existing crosscultural research on picture perception, Deregowksi suggests that relationships of culture, perception, and viewing abilities be conceptualized in terms of sets of contiguous and, at times, overlapping skills. Real-world spatial experience may utilize 3D skills that 2D representations do not exploit, such as binocular disparity (using the differences in the images recorded by each eye as a means of inferring information about depth and distance). Many real-world perceptual skills, by contrast, do overlap with visual skills relevant for picture perception, and some may be influenced by varying patterns of cultural organization (e.g., as Segall et al. point out, respondents from plains or dense jungle
environments may demonstrate less sensitivity to linear perspective when used as a pictorial depth code). Finally, some pictures employ purely representational codes with no overlap in real-life experience (e.g., wipes in film or streaky lines used to represent speed in comics). Obviously, with this last category one might expect greater interpretational difficulties for naïve viewers, but not inevitably. As our example from The Searchers indicated, narrative contextual information may be used to connect separate images or events even when the optical transition employed is a more symbolic one.

What, then, does all of this imply about the need for a period of tutoring implicit in the film-as-language model? Quite simply, such a period should not be necessary for inferring narrative relations in standard movies (i.e., films that do not create deliberate narrative enigmas, in contrast to movies like Last Year at Marienbad). Empirical research with naïve viewers (in most cases young children and, in one unique study, inexperienced adults) offers evidence that the use of specifically cinematic devices (or, as the film-as-language paradigm might term them, symbolic codes), such as montage, camera movement, or subjective shots, do not pose substantial interpretational obstacles for naïve viewers provided developmental requisites are met (i.e., that the viewers have developed sufficient real-world conceptual and cognitive skills that can be applied to the film or television medium). A brief review of some of this evidence will help clarify these points and will provide further evidence from which to question the theoretical and methodological efficacy of grounding film theory in conceptions of cinematic relationships involving arbitrary, unmotivated signs.

Young children’s comprehension of an array of cinematic techniques was explored in a recent study by Smith, Anderson, and Fisher.37 In one experiment, they showed three- and five-year-old children brief stop-motion animation video sequences that employed pans, zooms, and cuts to present a simple story. Another group of children saw the same stories in sequences with no editing or camera movement. When the children were invited to recreate the stories using the dolls and sets seen in the films, no differences in story comprehension were observed between the two groups, indicating that these film techniques seemed to pose few cognitive or interpretational problems for these young viewers. These were viewers who had seen television before. It is, however, the absence of age-related differences in their performance that is striking, especially in light of the hypothesis that these techniques should require medium-specific learning in order to be understood.

Because these sequences were relatively simple in structure (featuring only one pan, zoom, and cut), the researchers ran a second experiment with more complex visual presentations and longer visual narratives. Cinematic manipulations this time consisted of parallel editing to imply simultaneity of action, subjective shots representing a character’s literal viewpoint, ellipses achieved through editing that deleted portions of a continuous event or action, and editing to imply a layout of contiguous spaces (e.g., cutting from an establishing shot of two buildings to a shot of a character looking out of a window, then to a reverse-angle interior shot of the character at the window). This time the children were four and seven years old.

As before, they were asked to recreate character perspectives and spatial layouts using the dolls and sets that had appeared in the films. Comparisons across all classes of montage indicated that clear majorities in both age groups made correct inferences of space, ellipsis, and character perspective. Inferences of simultaneity proved difficult for the four-year-olds, but not for the seven-year-olds. Both groups proved especially skilled at reconstructing implied actions omitted in the edited narratives.

This study indicates good comprehension by young children of basic montage techniques, which is what one would expect if interpretation of the visual displays draws on a child’s developing real-world visual skills and experience. Emphasizing the status of cinematic images as iconic signs having a clear referential basis and inviting a transfer of real-world visual skills to the pictorial display does not deny the existence of medium-specific skills relevant for making sense of a motion picture, but it does reserve a more modest space for them than would the arbitrary-relational signifier model. Clearly, some film techniques are more symbolic than iconic and may be coded in a more arbitrary fashion. Such techniques would invite the viewer’s application of specific, medium-based competencies. For example, the Smith et. al study employed parallel editing as one category of montage. Drawing the correct inference from parallel editing is arguably a medium-specific skill, and, as one would expect, the largest age differences in the average percentage of correct responses showed up there. But, rather than needing to become proficient at manipulating an arbitrarily symbolic set of cinematic signifiers, children can generally interpret visual displays involving point-of-view editing by transferring to the display their developing real-world visual skills and experience. Evidence supporting a developmental view of the application of real-world cognitive and perceptual skills
to film and television displays involving point-of-view editing is found in Comuntzis-Page’s study of the relationship between children’s evolving real-world perspective-taking skills and their ability to infer meaning from an edited video sequence. She found that children who successfully demonstrated a knowledge of visual perspectives in actual 3-dimensional situations (e.g., children who knew that observers situated in different places can see different sides of a common object) were better able to understand the changing camera viewpoints of a television presentation. In fact, she found that skill at inferring character relationships in a 3-dimensional layout seemed to be a prerequisite for making the proper analogous inferences from the 2-dimensional video display. Only children who did well at the former task also did well on the latter, and clear age-related differences between perspective takers and nonperspective takers were found, supporting a developmental view.

The research cited thus far has used children to represent a population of relatively unskilled viewers. However, one unique study of motion picture perception and interpretation exists using adult viewers who had little familiarity with any mass media. Working with a seminomadic, nonliterate, pastoral tribe in Kenya, a community whose aesthetic expressions concentrated on personal adornment and performance arts, the researchers showed adult villagers two videotapes of a culturally familiar story. One version was unedited, the other featured 14 cuts with frequent alterations of close-ups, medium shots, long shots, and zooms. No significant differences in ability to recall story information were found between respondents who viewed the edited and unedited versions. Fragmentation of the visual scene through point-of-view editing did not hinder comprehension, nor did it seem to require the use of any medium-specific skills. The researchers concluded that, on the contrary, continuity editing codes that manipulate point of view seem to function as “analogs of perceptual processes.”

Levels of Iconicity

The evidence reviewed thus far bears on the capability of pictures to furnish cues regarding the spatial layout of a scene or situation that are analogous to sources of information commonly found in real-world visual experience. It also bears mentioning, however, that another level of iconic information typically exists in pictures which is central to a consider-

ation of the comprehensibility and emotional effects of the cinema. Photographic images of people necessarily reproduce that information about facial expression and gesture which Ray Birdwhistell has called “kinesics.” Birdwhistell and his colleagues studied the systematic patterning of body motion within cultures and designed an elaborate notational system to transcribe and study the distinctively patterned strings of body motion cues (called “kines”) that are situationally articulated as a communicative form in everyday life. Birdwhistell explicitly rejected the notion that language was the most important channel of communication, arguing instead for a conception of communication as a multi-modal, multi-sensory process which included visual and gestural as well as auditory channels. He demonstrated how body motions are culturally patterned in symbolically significant ways that are understood by members of a given community whose socialization processes have sensitized them to coded kinesic displays. While Birdwhistell emphasized the importance of cultural context in determining the encoding of specific kinesic patterns, other researchers have argued that some gestural expressions—those on the face, for example—may function as biologically based paracultural signals for emotion. While all cultures mandate display rules governing who may display which emotions and when, experimental evidence furnished by Paul Ekman and his colleagues has indicated that certain facial displays seem correlated with basic categories of human emotion and are capable of being recognized crossculturally.

The obvious point to be made from this brief discussion of the work on facial and body motion communication is that the motion picture camera furnishes an excellent means of recording the expressive meanings carried by these channels. Birdwhistell, in fact, employed still photographs and motion picture footage to record and store the data used in his analyses. It should be an assertion of the obvious to point out that a major source of the appeal and power of the movies lies here, in film’s ability to capture the subtleties and nuances of socially resonant streams of kinesic expressions, and not just to passively capture them but, via close-ups and other expressive devices, to intensify and emphasize the most salient cues for the viewer’s understanding in cognitive and affective terms of the meaning of the scenes depicted on screen. (Recent studies of acting by James Naremore and Roberta Pearson have emphasized the way that different performance traditions code expressive behavior and how film technique may be used to frame and emphasize this coding.) The relational-arbitrary signifier hy-
The Discourse of Pictures

What do these studies and avenues of research tell us about cinematic codes and the appropriateness of defining them, as current film theory does, as a series of relational differences among arbitrary signs? The empirical evidence clearly suggests that pictorial identification skills do not develop from an extended period of exposure to signification and consequent learning, as do language skills, and that this is probably due to the fact that most realistic pictures are isomorphic with corresponding real-world visual displays, unlike symbolic signs, which have a more arbitrary relationship to what they represent. Furthermore, via the technologies of motion picture recording, the camera is able to reproduce in clearly recognizable and even intensified form the familiar streams of facial and body motion cues which symbolically encode the meanings of the social situations portrayed on screen. These cues should be readily understood by cinema viewers, just as they are in real-world visual experience. If the distinctions between iconic and symbolic modes that we have been emphasizing are really relevant to differences between pictures and language, then one would expect iconic modes to be processed more readily than symbolic ones, especially by young children, and existing research lends support to this idea.

Thus, it is not clear how the concept of the relational, arbitrary signifier might apply to cinematic images or, for that matter, to any iconic image. Certainly, as Hockett indicated, few communicational signs are completely iconic, since a state of total iconicity would imply that the sign was completely indistinguishable from its referent. Nevertheless, pictorial signs do bear clear structural similarities to their referents, with the attendant consequences for perception and comprehension that we have just reviewed. The linguistic model has proven to be a very powerful paradigm for the analysis of images and larger film structures (e.g., dramatic scenes) in part because that analysis is typically conducted via words. When film scholars analyze images, visual information is translated into verbal description, one modality is substituted for another. But, as Branigan has noted, one must be cautious in speaking about the discourse of pictures: "The analyst must recognize that the very fact of talking about narration requires a re-presentation of it by verbal or other means which may capture only some of its features." The concept of the relational, arbitrary signifier would seem most applicable to pictures here, in the language of analysis which is employed to intellectually manipulate pictorial representations. Furthermore, while empirical evidence indicates that viewers easily and readily make the required inferences necessary to sequence a series of pictures into a coherent narrative, even to the extent of inferring information about situations not directly pictured, several scholars have argued that these inferences are organized on a conceptual and propositional—but not necessarily a linguistic—basis. Pylyshyn, for example, suggested that "Such concepts and predicates may be perceptually well defined without having any explicit natural language label.

Thus we may have a concept corresponding to the equivalence class of certain sounds or visual patterns without an explicit verbal label for it. Such a view implies that we can have mental concepts or ways of abstracting from our sense data which are beyond the reach of our current stock of words, but for which we could develop a vocabulary if communicating such concepts became important." Linguistic models, in other words, are not requisite for explaining how we respond to and make sense out of pictorial information, nor even for describing how we encode visual information for longterm memory storage such that it can be subsequently reaccessed.

As noted, arbitrary-relational signifier models are most applicable where the discourse being examined is language-based rather than strictly pictorial. In this respect, it follows that theories about ideology and ideological effects are best formulated not with regard to pictures sui generis (e.g., as has been done with perspective-based images) or with such machines for seeing as the camera and projector, but rather as a matter of the content of a given film and its modulation via techniques of cinematic style. As Noël Carroll has remarked, ideological design needs to be examined film by film, on an individual case-by-case basis. Because theories about linguistic structure do not seem to be directly applicable to pictures, they therefore form an insufficient basis for grounding a critique of cinematic discourse or the place of ideology in pictorial images.

Proponents of the application of the Saussurean lineage to cinema studies might reply that the poststructural attention to textuality and discourse as
culturally relative constructs provides a framework for investigating ideology and the cultural or political polyvalence of a given text. That the poststructural methodology works in this way is undoubtedly true. But, unfortunately, what tends to be displaced are issues of how cinema is able to communicate crossculturally (i.e., attain global popularity) and the even more basic questions of what makes the cinema intelligible to its viewers. Not all cultures organize libidinal and psychic energy in the same way, and one would not therefore wish to posit Lacanian categories as basic mechanisms explicating the means of cinematic communication. Yet all cultures currently studied do demonstrate clear pictorial and cinematic perception abilities. Moreover, these abilities are shared with a variety of animals. Picture recognition abilities have been demonstrated across a wide range of nonhuman subjects—primates, birds, fish, reptiles, even insects. These include unlearned, spontaneous responses to still and motion pictures as well as responses that are the result of conditioning. Recognition of objects by nonhuman subjects has been demonstrated across different classes of pictorial media—black-and-white and color photographs, high-contrast photographs, film and videotape, even line drawings. Comprehension of filmed events has been elicited, as well as the formation and pickup of object-class concepts from pictures (e.g., as in the ability of pigeons, trained to recognize human figures in slides of various environments, to transfer this ability and the class concept of “human being” to sets of slides they had never seen before). Positing arbitrary signifiers and unconscious processes of the mind seems to deflect the ability of cinema studies to grapple with some of the most basic questions about cinema, namely, how are viewers crossculturally able to make sense of the medium and what is the necessary biological basis of pictorial communication that makes it so effective for human and nonhuman subjects alike. As Noël Carroll has recently suggested, emphasizing the way that film structure works to secure the cognitive clarity of the experience for the viewer can give us a basis for explaining the attractiveness and appeal, and perhaps even the emotional power, of the medium.

Furthermore, emphasizing the cinematic code as an arbitrary, unmotivated sign places film theory in the uncomfortable position of implying certain consequences for the real-world viewer that are contrary to the observable evidence, namely, that cinematic images should not function in ways that are isomorphic with their real-world counterparts, that viewers should have to learn the symbolic meanings of basic cinematic structure, and that, logically, because discourse is culture-bound, the appeal of the medium should be too. Film theory has dealt with these problems by pointing to the transparency effect of the cinema or to the illusionism of the photographic image based on what are regarded as the culture-bound and symbolic conventions of perspective. But, as this article has suggested, some important gaps and contradictions are found between these formulations, the Saussurean lineage, and the empirical evidence regarding pictorial recognition and visual communication. Unless film theory can reconceptualize the cinema as an iconic, rather than as a purely symbolic, mode of communication, it is difficult to see how these serious theoretical gaps may be closed. We need to recover a recognition of the analogical component of pictorial signs. Rather than dealing with this component in metaphorical terms (e.g., via transparency effects, subject positioning in the discourse of the Imaginary, or illusionism), an appreciation of the isomorphic relations between pictorial signs and their referents, and attention to the differences between pictorial and linguistic modes of communication, can help invigorate film theory by reconnecting it to the observable experiences of real viewers.

As noted at the beginning of this essay, the intention here is not to remove cultural considerations from questions about film viewing and interpretation. Culture clearly enters into the inferences viewers draw from cinematic images and narratives where meaning may be constructed in terms of the intellectual horizons provided by class, race, gender, and similar variables. But it has been the contention of this essay that culture plays a less decisive role at the level of comprehension discussed herein. Furthermore, only by knowing where cultural considerations shade off into what may be more properly termed physiological or cognitive capabilities are we likely to be able to apply cultural analyses in fruitful and relevant ways. This approach can help theory explicate the intelligibility and power of the movies in ways that are responsive to, and grounded in, the observable experiences of actual viewers.

Stephen Prince teaches in the Communication Studies Department of Virginia Tech. He is the Book Review Editor of Film Quarterly.
NOTES

My thanks to Carl Plantinga for reading an earlier version of this essay and offering helpful suggestions.


5. In the influential analyses of Jean-Louis Baudry, these functions are a consequence of the machineries of camera and projector, irrespective of the content and subjects that are filmed. See “Ideological Effects of the Basic Cinematographic Apparatus” and “The Apparatus: Metapsychological Approaches to the Impression of Reality in the Cinema,” in Rosen, ed., Narrative, Apparatus, Ideology, pp. 286–318.


9. Ibid., p. 11.


11. Ibid., p. 216.

12. Burnett, Explorations in Film Theory, p. xvii.


15. The relevant essays explicating this position can be found in Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf, ed. John B. Carroll (Cambridge, MA: MIT Press, 1984).

16. See, for example, John B. Carroll and Joseph B. Casagrande, “The Function of Language Classifications in Behavior,” in Readings in Social Psychology, ed. Eleanor E. Maccoby, Theodore M. Newcomb, and Eugene L. Hartley (New York: Holt, Rinehart and Winston, 1958), pp. 18–31. It should be noted, however, that the results of this study, while offering some support of the Whorfian view, are nevertheless mixed and somewhat inconsistent.


28. Ibid.


32. Messaris also makes this point.


34. Discussing these cues, David Bordwell points out that cinematic images also routinely distort or alter them (e.g., through the effects of different focal length lenses, etc.). David Bordwell, Narration in the Fiction Film (Madison, WI: University of Wisconsin Press, 1985), pp. 107–10.


40. Ibid., p. 58.


46. Branigan, “Here Is a Picture of No Revolver,” p. 11.


Copying notice: Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by the Regents of the University of California for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of $1.00 per copy is paid directly to CCC, 27 Congress Street, Salem, MA 01970. For permission to photocopy materials for classroom use, call CCC’s APS Customer Service Department at (508) 744-3350 or fax them at (508) 745-9379.