

NARRATIVE COMPREHENSION and FILM



EDWARD BRANIGAN

L
SIGHT
N
E
S

NARRATIVE SCHEMA

PSYCHOLOGICAL USE VALUE

Narrative has existed in every known human society. Like metaphor, it seems to be everywhere: sometimes active and obvious, at other times fragmentary, dormant, and tacit. We encounter it not just in novels and conversation but also as we look around a room, wonder about an event, or think about what to do next week. One of the important ways we perceive our environment is by anticipating and telling ourselves mini-stories about that environment based on stories already told. Making narratives is a strategy for making our world of experiences and desires intelligible. It is a fundamental way of organizing data.¹

Recently narrative principles have been found in the work of a wide range of professionals, including attorneys,² historians,³ biographers, educators, psychiatrists, and journalists.⁴ This demonstrates that narrative should not be seen as exclusively *fictional* but instead should merely be contrasted to other (*nonnarrative*) ways of assembling and understanding data. The following kinds of document exemplify some *nonnarrative* ways of organizing data: lyric poetry, essay, chronology, inventory, classification, syllogism, declaration, sermon, prayer, letter, dialectic, summary, index, dictionary, diagram, map, recipe, advertisement, charity solicitation, instruction manual, laundry list, telephone directory, birth announcement, credit history, medical statement, job description, application form, wedding invitation, stock market report, administrative rules, and legal contract. The relevance and connection of narrative, or nonnarrative, to our world – how it may be used in that world to accomplish a goal – is a separate issue concerning its “mode of reference” as either fiction or nonfiction.

As a starting point and for simplicity, then, I will divide texts into four basic types: narrative fiction (e.g., a novel); narrative nonfiction (e.g., history); nonnarrative fiction (e.g., many kinds of poetry); and nonnarrative nonfiction (e.g., essay). The boundaries among these types are not absolute but relative to the questions one wishes to ask

NARRATIVE COMPREHENSION AND FILM

about the data that has been organized. The fact that certain poetry, for example, is nonnarrative does not mean that, considered at a fine grain, it may not also exhibit some aspects of narrative organization (e.g., defining a scene of action and temporal progression, dramatizing an observer of events). One should not allow the usefulness of broad categorizations (poetry, novel) to obscure the ways in which a narrative strategy may be applied successfully by a reader in comprehending certain aspects of some texts; or, for that matter, to obscure the ways in which nonnarrative reading strategies may penetrate narrative texts at certain levels.

It is also important to distinguish between two broad fields in which a given narrative may function. In one context, it can be said that a narrative must be consumed as a material and social object, and must respond to an agenda of community issues. In this context, a narrative acquires labels of immense variety in order to arouse the interests of community members. These labels are the pathways on which it moves through society by being bought and sold, or exchanged. In a second context, however, a narrative can be said to exist for only one person at a time. Engaging intimately with a perceiver, narrative enters thought itself, competing and jostling with other ways of reacting to the world. Thus narrative, at least initially, may be analyzed in two different ways. From one angle it appears as a social and political object with an *exchange* value arising from its manufacture as an object for a community; from another angle it appears as a psychological object with a *use* value arising from perceptual labor – from the exercise of the particular skills possessed by a member of that community. Ultimately, of course, these two values are not independent. One may study the psychological dimension of exchange (e.g., commodity fetishism) and the social dimension of use (e.g., propaganda). The particular social ground which defines an individual's language and horizon of action cannot be completely divorced from that individual's language competence and abilities. Narrative depends on an unspoken, permanent agenda of topics in a community which, in turn, justifies the community activities for which abilities must be found and developed in individuals.⁵ In studying how a narrative is assigned labels in order to be exchanged and used, one is studying basic human proficiencies: skills employed in manufacturing and selling a material object as well as perceptual skills employed in realizing a use for the object.

In spite of the copresence of exchange value and use value, I will tentatively separate the two contexts in order to better highlight the nature of a *relative* autonomy where each value provides a ground for the other. This will also enable me to limit the terms of discussion so as to begin to talk about how narrative functions in our world. It is the aim of this book to examine the use value of narrative, specifically the

NARRATIVE SCHEMA

psychological dimension of use. I wish to examine how we come to know that something is a narrative and how a narrative is able to make intelligible our experiences and feelings. I will argue that it is more than a way of classifying texts: *narrative is a perceptual activity that organizes data into a special pattern which represents and explains experience*. More specifically, narrative is a way of organizing spatial and temporal data into a cause-effect chain of events with a beginning, middle, and end that embodies a judgment about the nature of the events as well as demonstrates how it is possible to know, and hence to narrate, the events.

Although it will often be convenient to use the word "narrative" to refer to an end result, or goal, one should not forget that this final product ("here is a narrative") arises from a particular and ongoing (narrative) method of organizing data. Thus the word "narrative" may refer to either the product of storytelling/comprehending or to its process of construction. The first four chapters will begin to specify narrative in both these senses while chapter 7 will consider how a narrative relates to the real world in a "fictional" or "nonfictional" manner.

If narrative is to be considered as a way of perceiving, one still needs to specify the way. Further, one needs to specify what is meant by "perceiving." In general, my approach will be to allow the notion of "perceiving" to remain quite broad and elastic, capable of referring to any one of a range of distinct mental activities. When sharp lines must be drawn, I will use special concepts. Thus the word "perception" will be used in this book to point toward any of the following: a "percept" derived from reality; a preconscious assumption being made about reality; or an acknowledged fact of physical reality. The word "perception" may also be used to refer to an intuition (e.g., perceiving that color seems to be intrinsic and permanent to an object while sound appears to come from an object, to be created and contingent); or, it may refer to a propositional conclusion that a perceiver has reached about sensory perception through a process of reasoning; or, it may simply refer to an *attitude* we adopt when confronted by something that is a *representation* of something else. Some theories would classify the latter as cognition rather than perception. As we shall see, particular theories of narrative will divide up the operations of human consciousness in various ways to emphasize different abilities. Thus the word "perception" in this book will earn its exactness only through the finer discriminations made by particular theories. In the next chapter, for example, I will begin to refine the notion of "perception" by introducing a fundamental distinction between "top-down" and "bottom-up" modes of perceiving.

LOGICAL TRANSFORMATIONS IN NARRATIVE

What way of arranging data is characteristic of narrative perception? We readily distinguish narrative from other experiences even if we cannot say how the judgment is being made, just as we may not be able to say why something counts as a "game" or a "grammatical" sentence.⁶ Intuitively we believe that a narrative is more than a mere description of place or time, and more even than events in a logical or causal sequence. For example, an account of the placement of objects in a room is not a narrative. Similarly, though a recipe involves temporal duration and progression ("bake until golden brown . . ."), it is not normally thought of as a narrative (the story of a pie). Nor does a sequence of actions become a narrative by being causal, completed, or well-delineated; for example, a planet orbiting the sun, the construction of a syllogism, the recitation of an alphabet, or the actions of departing, traveling, and arriving do not by themselves form a narrative. Instead, narrative can be seen as an organization of experience which draws together many aspects of our spatial, temporal, and causal perception.

In a narrative, some person, object, or situation undergoes a particular type of change and this change is measured by a sequence of attributions which apply to the thing at different times. Narrative is a way of experiencing a group of sentences or pictures (or gestures or dance movements, etc.) which together attribute a beginning, middle, and end to something. The beginning, middle, and end are not contained in the discrete elements, say, the individual sentences of a novel but signified in the overall relationships established among the totality of the elements, or sentences. For example, the first sentence of a novel is not itself "the beginning." It acquires that status in relationship to certain other sentences. Although being "physically" first in some particular way may be necessary for a "beginning," it is not sufficient since a beginning must also be judged to be a proper part of an ordered sequence or pattern of *other* elements; the elements themselves are *not* the pattern. Narrative is thus a global interpretation of changing data measured through sets of relationships. We must now consider the nature of this overall pattern of relationships.

Tzvetan Todorov argues that narrative in its most basic form is a causal "transformation" of a situation through five stages:

- 1 a state of equilibrium at the outset;
- 2 a disruption of the equilibrium by some action;
- 3 a recognition that there has been a disruption;
- 4 an attempt to repair the disruption;
- 5 a reinstatement of the initial equilibrium.⁷

These changes of state are not random but are produced according to

NARRATIVE SCHEMA

principles of cause and effect (e.g., principles which describe possibility, probability, impossibility, and necessity among the actions that occur). This suggests that there are two fundamental kinds of predication in narrative: *existents*, which assert the existence of something (in the mode of the verb "to be"), and *processes*, which stipulate a change or process under a causal formula (in the mode of such verbs as "to go, to do, to happen").⁸ Typical existents are characters and settings while typical processes are actions of persons and forces of nature. But there is more: the changes of state create an *overall pattern* or "transformation" whereby Todorov's third stage is seen as the "inverse" of the first and fifth stages, and the fourth stage the "inverse" of the second (since it attempts to reverse the effects of the disruption).⁹ The five stages may be symbolized as follows: A, B, -A, -B, A. This amounts to a large-scale pattern (repetition, antithesis, symmetry, gradation) among the causal relationships and is temporal in a new way; in fact, some theorists refer to such patterns as a "spatial" form of narrative.¹⁰ This emergent form, or transformation, is a necessary feature of narrative because, as Christian Metz observes, "A narrative is not a sequence of closed events but a closed sequence of events."¹¹

Consider as an example the following limerick:

There was a young lady of Niger
Who smiled as she rode on a tiger.
They returned from the ride
With the lady inside
And the smile on the face of the tiger.¹²

Analyzing the limerick as a narrative using Todorov's transformations, results in the following global structure:

There was [once upon a time]:
A smile
B ride
-A [swallowed: a horrible pleasure?]
-B return
A smile
[which goes to show that . . .]

The limerick illustrates several important points about Todorov's transformations. First, the structure does not represent directly the actual processing of the narrative by a perceiver but only its conceptual or logical form after it has been interpreted. The reader discovers that the narrative did not begin with "lady," or "youth," or the place of "Niger," as its initial term ("A") because none of those beginnings will yield a macro-description of the required kind. Taking "smile" as an initial term, however, produces a sequence of transformations that will

NARRATIVE COMPREHENSION AND FILM

embrace the limerick as a whole (A, B, -A, -B, A). Nevertheless, this does not yet explain why a reader may smile at the limerick. The humor of the limerick resides in the sudden realization of what must have happened, of what was *omitted* from its proper sequence *in the telling*. The absence of the woman at the end answers to a gap in the chronological structure of the telling of the event. Todorov's middle stage - a "recognition" of the disruption - already hints that the *actual process* of moving from ignorance to knowledge will be of central importance to our experience of narrative. Not only characters and narrators, but readers are caught up in ways of perceiving and knowing. These crucial issues will need to be addressed in more detail and will be the topic of chapters 3 and 4.

Second, Todorov's structure does not represent the entirety of our comprehension of the narrative aspects of the limerick. The reader must supply an epilogue or moral to the story which justifies its being told (which goes to show that . . .). This involves a rereading and a reassignment of some of the meanings - a process facilitated in the first line by assuming that reference will be partially indeterminate in the manner of a fiction (once there was *a* time . . .). Eventually the reader must rationalize how he or she might know such an exotic world within his or her preconceptions of an ordinary world.

Finally, although this narrative is arranged to focus attention on what Todorov calls the inversion of the initial equilibrium (the middle cause which is the opposite of smiling, i.e., being swallowed, -A), the logical structure cannot account for all of the inferences that the reader must draw in discovering the nature of the "inversion" which turns out to have an unexpected literal dimension (ingestion) as well as a number of metaphorical dimensions. What qualifies the inversion *as* an inversion? The reader must make inferences in spite of (and also because of) being misled by the verse. Consider, for instance, the deception of the phrase "they returned" in line 3; and the fact that the lady's ride is enlarged by the word "returned" to mean that she had departed on a *trip*, even if only a short trip; and the semantic play with the preposition "on" and with the definite article of "the smile": at the end only her smile "rides on" the tiger and the smile is not hers but a smile of the tiger. (We will examine more fully the significance of deception in relation to perceiving and knowing a narrative in chapters 3 and 4.) The implications of the use of causation and metaphor in the narrative extend at least to the reader's knowledge and beliefs about female sexuality, pleasure, oral gratification, desire, risk and trust; and perhaps also to the consequences of being "away from the home."¹³ It is far from clear how the logical form of the limerick is able to summon these forms of knowledge. Would a reader, for example, be able to list all possible "inversions" of a given initial state? Or is there instead a

NARRATIVE SCHEMA

sense in which an inversion must be *discovered* to be appropriate through the operation of processes which are not all "logical" in the same ways?

Before expanding our idea of narrative form, it may be useful to contrast the above limerick with a poem which is nonnarrative:

Roses are red
Violets are blue
Sugar is sweet
And so are you.

For reasons to be made clear shortly, I will refer to the structure of this poem as a *catalogue*, not a narrative. For now it is enough to notice that the verb "to be" has been used four times in an attributive and atemporal sense (as in the extreme case of identity, "a rose is a rose"). The reader does not interpret the poem as implying one or more temporal adverbial complements, such as, "Roses are red at noon, violets are blue at two." No temporal logic connects the redness of roses with the blueness of violets and the sweetness of sugar and "you." Instead the reader constructs a pair of categories which have no "tense": one which contains two flowers, and another which contains both sugar and the reader himself or herself ("you"). The "causation" at work in the poem – producing the conclusion signaled by "and so" – is asserted to be as logical, natural, and timeless as grouping roses and violets together as flowers (or, perhaps, as objects having color). The rhyme (blue-you) brings together the two categories and implies that the logic of forming the flower category is as certain as the logic of grouping sugar and "you." Thus although both the poem and the limerick compare a person's desirable qualities to something which may be tasted or eaten, the poem is not a narrative because its conceptual structure does not depend on a definite temporal progression which ultimately reveals a global pattern (e.g., A, B, -A, -B, A). Instead the poem is based on forming simple pairs of things with the final intimation – an epilogue of sorts – that "you," the reader, and an implicit "I," the author, should also form a pair.

I would now like to imagine for a moment something incredible. Suppose that the limerick that tells the story of the woman riding on the tiger contains an interlude where the tiger sings for the woman the poem, "Roses are red." In one sense, the narrative has been interrupted by a nonnarrative, catalogue sequence. In another sense, however, there has been no real interruption, for both the narrative limerick and the nonnarrative poem develop a connection between taste and beauty in which the sexual drive is represented as an appetite that devours. Is the limerick-poem then a hybrid? Does the narrative dominate the catalogue, or is the narrative merely an excuse for a clever song? I

NARRATIVE COMPREHENSION AND FILM

believe that there is no definitive answer to what it really is. Rather, the answer will depend upon the purpose in asking the question: within what context must an answer be framed, how narrowly must the text be construed, which meanings are most important, and so forth. Recognizing the complexity and dynamism of a text is usually more important than assigning a final, decisive label to it.

Rick Altman has drawn attention to the importance of certain catalogue systems within narrative texts. He speaks of narrative as possessing a "dual focus" where one focus is composed of a chronological and causal progression (the "syntagmatic") while the other is composed of a multitude of binary oppositions among elements that are "static" and that exist outside the time of the causal progression (the "paradigmatic"). A textual element (shot, scene, aspect of style, character attribute, theme, etc.) that is functioning paradigmatically makes a pair not by calling forth its "effect" in a linear fashion, but by suggesting a *parallel* with something else, a similarity or contrast. Paradigmatic pairing (or, what I have described as a "catalogue") creates collections of objects organized according to "conceptual" principles. Altman finds that in the genre of the American musical film, a special kind of paradigmatic focus, designed to show that opposed sets of categories are not mutually exclusive, overwhelms the causal, frame story.¹⁴

For present purposes, I am less interested in reaching a definitive judgment about the precise nature of a text than in describing the different types of organization that underlie a reader's experience moment by moment. Accordingly, I will construe Altman's notions of "duality" and "focus" more narrowly and shift them to a new realm. I will also introduce new terms that divide up the field of study in a somewhat different way, allowing for finer distinctions.¹⁵ As we shall see, the reason for such a shift in terminology is correlated with a change in the object of study: an attempt to specify the formal logic of narrative gives way to an examination of the interaction of narrative with a perceiver – a pragmatics of comprehension.

PRAGMATIC FORMS IN NARRATIVE

The notion of narrative as a sequence of logical "transformations" brings together two concerns: an awareness of pattern as well as purpose. These concerns may be seen in the double meaning of the English word "design," which may signify either a formal composition, an "arrangement" of elements (e.g., "The *design* utilized bright colors"), or an "intention" (e.g., "Her letter ended in mid-sentence by *design*," "He has *designs* on her property"). The importance of the transformations for Todorov would seem to be the suggestion that some (designing) forces have *intervened* in the five stages of narrative to shape the

NARRATIVE SCHEMA

final pattern (design) which turns out to be a reshaping of the initial state. Thus something more than describing categories, and more even than labeling cause and effect, is needed to create a narrative; however attenuated, an element of choice, probability, or purpose must be seen by the reader to promise through its transformations an answer as to "why" or "when" something is or could be other, and "how" it returns to being the "same." One might say that the reader's discovery of this overall process at work in narrative is a mode of causal reasoning about human affairs which is distinct from merely labeling a cause, or assessing the probabilities of a local action. In this way, one may think of narrative as a mechanism that systematically *tests* certain combinations and transformations of a set of basic elements and propositions about events ("A" and "B" in my examples). The aim is not simply to enumerate causes, but to discover the causal efficacy of an element – its possibility for being, and for being other, as the reader may desire.

Many writers have argued that the logic underlying narrative is more complex than Todorov's pattern (A, B, -A, -B, A). A central concept like "transformation" may be understood in different ways, or new concepts may be developed in an attempt to interrelate narrative pattern and purpose.¹⁶ Claude Lévi-Strauss, Claude Bremond, and A.J. Greimas claim to be extending and refining the insights of Vladimir Propp, who defined the logic of the Russian wonder tale in terms of seven basic "spheres of action" (character roles), thirty-one "functions" (types of action), certain "moves" (fixed strings of functions), and "auxiliaries" (transitions).¹⁷ Lévi-Strauss defines pairs of opposed "mythemes" while Greimas tightens narrative logic even further by defining its elements in terms of the "square of opposition" used in traditional logic to classify categorical propositions. For Greimas, narrative becomes a special working through of contraries, subcontraries, converses, and contradictories.¹⁸ Like Todorov's five-part scheme, the goal of these methods is to describe the large-scale symmetries that draw together and unify the parts of narrative.

All of these approaches have been influential and have produced important results with certain texts. Nevertheless, the linguistic theories from which they have drawn many concepts have in the intervening years been modified or superseded. Also, formal logic has been shown to have limitations as a descriptive model for human thought. More recent models of human language emphasize the dynamics of a perceiver's interaction with a text – i.e., pragmatic situations – by studying a perceiver's use of "fuzzy" concepts, metaphorical reasoning, and "frame-arrays" of knowledge. Correspondingly, there has been a general tendency to move away from the linguistics of Ferdinand de Saussure as well as away from an exclusive reliance upon formal and logical schemes, such as Noam Chomsky's deductive rules which in many

cases do not seem flexible enough to capture the wide-ranging, often speculative aspect of interpretation.¹⁹ The stakes remain high, however, as Wallace Martin reminds us: "Identification of universal narrative patterns would seem to tell us not just about literature but about the nature of the mind and/or universal features of culture."²⁰ The goal of a pragmatics of narrative is to achieve a psychological description that can explain how a perceiver is able to interpret a text as a narrative moment by moment.²¹

One might begin to relax a strictly logical definition of narrative so as to include pragmatic aspects by pursuing the definition offered by Stephen Heath in his analysis of Orson Welles's film, *Touch of Evil* (1958):

A narrative action is a series of elements held in a relation of transformation such that their consecution determines a state S' different to an initial state S; thus: S-x-x-x-x-x-x-x-S'. . . . A beginning, therefore, is always a *violence*, the violation or interruption of the homogeneity of S. . . . The narrative transformation is the resolution of the violence, its containment – its *replacing* – in a new homogeneity. "Replacement" there has a double edge: on the one hand, the narrative produces something new, replaces S with S'; on the other, this production is the return of the same, S' re-replaces S, is the reinvestment of its elements. Hence the constraint of the need for exhaustion: every element must be used up in the resolution; the dispersion the violence provoked must be turned into a re-convergence – which is the action of the transformation, its activity. Ideally, a narrative is the perfect symmetry of this movement.²²

In *Touch of Evil* the initial violence is literal as a car explodes in flames interrupting a kiss between lovers. Heath notes that when those lovers kiss at the end of the film, it is "the same kiss, but delayed, *narrativised*."²³ For Heath, narrative is a precise series of displacements, often driven by the logic of the disclosure of an enigma that acts to replace an initial situation by returning to it. For Raymond Bellour the search for such a "perfect symmetry" in the form of repetitions and near repetitions ("rhymes") in the text becomes almost obsessive, extending from global patterns (where one is reminded of Todorov's precise transformations) down to the smallest micro-sequence of action.²⁴

But symmetries are not Heath's primary concern in defining narrative. He is anxious to show how some elements inevitably *escape* the tight narrative structure and become a residuum, an "excess," revealing hidden psychic and ideological processes at work in the text. Narrative exists because of these hidden processes and is an explicit attempt to master them. For Heath, the causality of narrative events in a plot is

NARRATIVE SCHEMA

merely a pretext for larger transformations which point to our *everyday beliefs* about ourselves and our world, and the ways in which we formulate (or repress) those beliefs. Heath is less interested in discovering a stable logical structure than in uncovering symptoms of belief, modes of persuasion, and values which are not at all logical in the way conceived by Todorov and Greimas. Narrative thus acquires the form of an *argument*, leading to such definitions of it as the following:

a connected sequence of . . . *statements*, where "statement" is quite independent of the particular expressive medium.²⁵

(Seymour Chatman)

A closed *discourse* [i.e., a sequence of predicative statements] that proceeds by *unrealizing* a temporal sequence of events.²⁶

(Christian Metz)

a . . . recounting [of] a chrono-logical sequence, where sequence is taken to be a group of non-simultaneous topic-comment structures the last one of which constitutes a modification of the first.²⁷

(Gerald Prince)

Still more generally, Sergei Eisenstein envisioned an "intellectual cinema" in which filmic "reasoning" would enrich narrative and produce a synthesis of art and science.²⁸

Prince's notion above of "modification" is quite broad and seems to include spatial, temporal, causal, and "zero" modifications as well as operations of inversion, negation, repetition, manner, and/or modality. The notion of narrative as a series of argumentative "statements" (i.e., propositions analyzable as a comment on a topic) that are suitably modified and independent of their manifestation in words, pictures, gestures, or other materials does capture something important about the phenomenon. Still, the notion of a narrative "statement" may have relinquished important detail for a generality bordering on vagueness. What, for example, are the limits of a "modification" to a "proposition"?

A similar problem of vagueness attends the almost obligatory discussion of the so-called "minimal narrative" which takes as its starting point E.M. Forster's distinction between chronology and causality. The following sentences illustrate the grounds of the debate, though theorists give different reasons for their conclusions:²⁹

These do not qualify as narratives:

(1a) The king died and then the queen died [chronology].

(2a) Mary ate an apple.

These are narratives:

(1b) The king died, and then the queen died of grief [causality].

(2b) Shirley was good then she drifted into a life of crime.

NARRATIVE COMPREHENSION AND FILM

In this debate there is an implicit belief that narrative is built up from a small set of basic units, or particles (e.g., topics, comments, and modifications), by addition and subtraction. The approach is reminiscent of the attempt by "analytic structuralism" in the field of psychology to account for human perception by positing certain basic "sensations" together with simple laws of combination.³⁰ The idea of narrative has become so impoverished by the search for minimal, logical conditions in a single sentence that it is unclear what qualities might attach to the more typical narratives which are exchanged and used in social arenas. Some writers, perhaps impressed by the pervasiveness of narrative thinking in everyday life and despairing of the attempt to find a bright line between narrative and nonnarrative, conclude that virtually everything is narrative. For instance, the following is deemed by one writer to be a narrative:

(3) Once upon a time there was a person. The End.³¹

Another writer concludes that "even mathematical proofs, with one step following another toward an inevitable conclusion, exhibit something of the dynamics of plot and closure."³²

I believe that what is needed is a description of narrative which avoids a strictly "logical" definition of minimal conditions even if supplemented by more expansive mechanisms like Todorov's transformations. Such a new description must also be more precise than discovering a set of "statements" which reveal pragmatic beliefs, or make arguments. One way to accomplish this goal is to concentrate on the cognitive processes active in a perceiver during his or her comprehension of narrative in an actual situation. The issue then focuses on how an overall narrative pattern may be discovered, or imposed, in the very act of perceiving. How do we manage to learn from narratives, moment by moment, and how do we learn to make our own narratives?

For Dan Lloyd the study of narrative comprehension is the study of a primary mode of thought quite distinct from other modes, such as "rational logic." He argues that the use of a "narrative logic" in solving problems explains why persons routinely fail certain tests of deductive and inductive reasoning. Thinking narratively has important advantages in the world and Lloyd calls for a new science - "psycho-narratology" - to examine the psychological foundations of narrative reasoning. The new science would be built upon concepts derived from the general study of narrative - "narratology" - and would include the work of such writers as Todorov, Bremond, Greimas, and Prince.³³

NARRATIVE SCHEMA

COGNITIVE SCHEMAS AND OTHER WAYS OF ASSOCIATING DATA

In order to focus on mental processes working in real time, one must begin with the fact that there are rather severe capacity limitations both on an individual's transient memory, which registers sensory information, and on his or her short-term memory, which is able to sort and classify only recent information. Short-term memory can manipulate only about five to nine "chunks" of data. (The word "red" will count as one chunk of data whereas the letters "rde" will count as three.) Thus it is primarily intermediate-term memory (sometimes called "working" memory) and long-term memory that must be carefully studied, for these are the sites of special mental operations that play decisive roles in redescribing data and recognizing global relationships, whether narrative or otherwise. Moreover, these special operations of working and long-term memory are not directly experienced by a perceiver, since "consciousness" has many of the limitations of short-term memory.

The use of working and long-term memory by a perceiver are notable examples of the fact that sensory perception (transient memory) cannot be considered apart from other types of mental processing. Experiments have demonstrated that what perceivers remember from a narrative, as well as what they forget, is not random but dictated by the specific method used in searching for global properties. This method of search guides the acts of encoding, comprehending, storing, retrieving, and "remembering" the features of narrative. These experiments support a basic premise of cognitive psychology, namely, that the classifications which a person imposes on material at the time of its processing will limit the ways in which the material can be subsequently accessed and used in problem-solving.³⁴ (Much of a person's childhood experience is lost because it is classified in ways that are incompatible with the classifications used by an adult to sort and retrieve experience.) I will refer to the specific method which searches for a narrative pattern as a narrative *schema*.

The notion of a schema is basic to much of cognitive psychology. A schema is an arrangement of knowledge *already possessed* by a perceiver that is used to predict and classify new sensory data. The assumption underlying this concept is simply that people's knowledge is organized. The fact that one often knows immediately what one *does not know* testifies to the structured nature of our knowledge. As Jean Mandler states, "when we know something about a given domain our knowledge does not consist of a list of unconnected facts, but coheres in specifiable ways."³⁵ A schema assigns probabilities to events and to parts of events. It may be thought of as a graded set of expectations about experience in a given domain. What we implicitly know about a

"room," for example, is much more than either the "connotations" of that word or the properties of an actual room that we may remember. We know still more about a "living room." The vague sort of mental pictures that we may summon of a "room" or a "living room" are not unlike the operation of a schema in representing an ordered set of associations and expectations that are used to judge certain experiences. A schema, of course, is more complex than a given word because it interacts with the environment. A schema tests and refines sensory data at the same time that the data is testing the adequacy of the (implicit) criteria embodied in the schema. The interaction of schema and data creates a perceiver's recognition of global patterns characteristic of that data. "Meaning" is said to exist when pattern is achieved.

A schema does not determine its object through necessary and sufficient conditions. It is a hierarchical arrangement which ranges from tentative and contingent conclusions about data (including "default" specifications) at one extreme to increasingly general and invariant specifications governing a class of data at the other extreme. Thus when "meaning" has been attributed to something through the use of a schema, the meaning has a probabilistic quality which incorporates assumptions and expectations rather than an absolute quality defined by necessary and sufficient conditions.

A schema is only one type of mental structure and a narrative schema is only one of many types of schema used to solve a wide range of everyday problems. Nonnarrative types of schema (some of which will be discussed shortly) may be applied to a narrative text; conversely, a narrative schema may be applied quite generally to process data and (as we shall see) may even be used to generate sense from "nonsense" data.

What sort of schema is responsible for the recognition of narrative patterns?³⁶ Nearly all researchers agree that a narrative schema has the following format:

- 1 introduction of setting and characters;
- 2 explanation of a state of affairs;
- 3 initiating event;
- 4 emotional response or statement of a goal by the protagonist;
- 5 complicating actions;
- 6 outcome;
- 7 reactions to the outcome.³⁷

Such a schema helps to explain some remarkable facts about narrative comprehension.

One of the most important yet least appreciated facts about narrative is that perceivers tend to *remember* a story in terms of *categories of information stated as propositions, interpretations, and summaries* rather than

NARRATIVE SCHEMA

remembering the way the story is actually presented or its surface features. It requires great effort to recall the exact words used in a novel or the exact sequence of shots, angles, lighting, etc. used in a film. The reason is that features of the "surface structure" of texts are typically stored only by recency in so-called "push-down" stacks where new elements are continually being added at the boundary, pushing the older elements farther away.³⁸ When we say we remember a film, we do not normally mean that we remember the angle from which it was viewed in the movie theater, or the exact angles assumed by the camera in a scene. Rather, when we speak of comprehending something, we mean that our knowledge of it may be stated in several equivalent ways; that is, our knowledge has achieved a certain independence from initial stimuli. In comprehending a visual object in film, for example, our knowledge of the object is such that we might imagine moving about within the space and assuming various angles of view, without thereby altering the object known. We know the object when we know how it may be seen regardless of the position from which it was actually seen. The object thus acquires an "ideal" or "abstract" quality. It should be mentioned that knowing how the object may be seen is very nearly imagining an object that is not in view at all.³⁹ This suggests that a theory of narrative comprehension will be incomplete without parallel theories of metaphor (because something new may be standing in for an original experience), and of fiction (because what is new may refer initially to the nonexistent).⁴⁰

There are many other remarkable facts about narrative comprehension. Information from a text is sorted and measured by a schema against other kinds of knowledge base. The result is that certain information in a narrative is elaborately processed and assigned to a hierarchy in working memory according to relative importance while much else is discarded. The "value" of information increases according to its improbability so that typical and probable elements - so-called "unmarked" elements of a paradigm - carry the least amount of information. The more typical the information is for a perceiver, the less well it is recalled for it is already implicit in a guiding schema. Events in a text are therefore marked as salient and acquire special significance because of expectations defined by the internal order of a schema.

Furthermore, complex propositions tend to be formed in memory (and reading time slows) when "boundaries" in the text are perceived to correspond to the segmentation provided by a schema. The reason is that story comprehension involves the continuous generation of better-specified and more complicated expectations about what might be coming next and its place in a pattern. Thus a perceiver will strive to make "logical" connections among data in order to match the general categories of the schema. This will involve a mental rearrangement of

NARRATIVE COMPREHENSION AND FILM

temporal sequences in a text. These new macro-propositions concerning global relationships among data are stored in memory and represent the "gist" of the narrative. In this manner a perceiver uses a schema to automatically fill in any data that is deemed to be "missing" in the text.⁴¹

There are many ways that a text may disrupt a perceiver's expectations. Unclear character "goals" and "inverted" order in a text require increased processing time because of a necessity to experiment with various classifications of the data within a schematic framework. Also, unexpected information can cause a reorientation of the schema in order to reclaim the important from the superficial. Comprehension slows when explicit propositions constructed earlier must be reactivated (cf. the notion of "retrospective" temporal order in chapter 2); or when previous inferences are indirectly disconfirmed (e.g., by a *pattern* of events rather than by explicit statement); or when a perceiver must make novel inferences. Finally, the limitations of working memory may be exploited in order to accentuate the so-called "fluctuating existence" of diegetic off-screen space in film.⁴²

In short, it has been amply demonstrated through many psychological experiments that an individual's attention does not spread equally through a narrative text but works forward and backward in an uneven manner in constructing large-scale, hierarchical patterns which represent a particular story as an abstract grouping of knowledge based on an underlying schema. Furthermore, a narrative schema may be applied in many situations. It has been shown that

even with meaningless nonsense figures moving in abstract paths, viewers were able to describe and remember a much longer series of events, by generating a simple story, and attributing anthropomorphic qualities to the figures and the motions they perform, than they could handle in purely physical terms.⁴³

Especially important is the way a perceiver infers the purposes, intentions, and goals of the constructed anthropomorphic entities. Thus it would seem that a narrative schema is always an option in processing data even when there are no human characters or the events are essentially "nonsense" data.⁴⁴

Although narrative is a powerful and general way of organizing information, it is essential to realize that the concept of a schema addresses only some of the issues concerning narrative.⁴⁵ A narrative schema does not directly address such problems as a perceiver's fascination, emotional reaction, or participation in a story; the effect on a perceiver of manipulations of point of view; nor the effect of actually experiencing a story in a community setting. Also, presumably, the nature of the medium and the "style" of the story will exert pressure

NARRATIVE SCHEMA

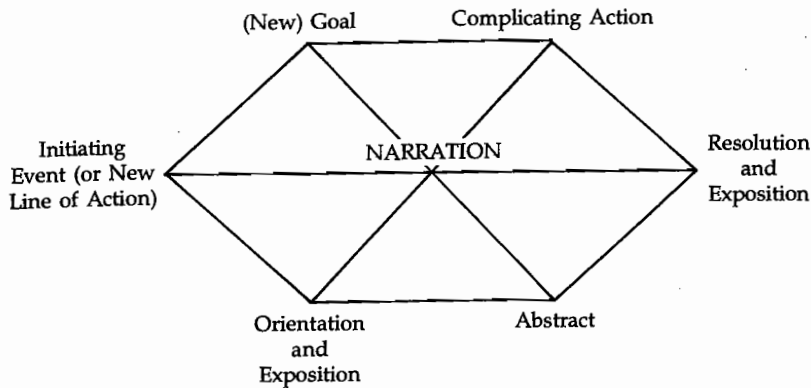


Figure 1 Narrative schema

Elements of a narrative schema displayed as a hexagon. A narrative is constructed by moving through the hexagon to create patterns at the levels of action, scene, episode, sequence, etc.

on comprehension. There is even a question about the nature of the "macro-propositions" generated through a schema: are they verbal, pictorial, or something else?⁴⁶ This range of issues is a reminder of the complexity of the narrative phenomenon. Still, it does not rule out addressing problems one at a time and then attempting to integrate various theories about human capability and performance.

I believe that by studying comprehension as a constructive activity (encompassing much more than the mere retrieval of images or words stored in the order received), one bypasses the now vexed film-theoretical question of whether film is like a "language."⁴⁷ Indeed, one almost bypasses the question of whether film is like a "communication." There is, in fact, good reason to believe that both film and natural language are special subsets of more general cognitive enterprises. One of these general enterprises is our ability to construct a narrative out of experience; that is to say, our ability to use a narrative schema to model a version of the world. In this view, both film narrative and written narrative express temporal relationships because both are mental constructions, not because film reduces to language.

A PROPOSAL FOR A NARRATIVE SCHEMA

It may be helpful to construct a narrative schema in somewhat more detail and to illustrate its application to a particular film. The elements of the schema I will present are derived primarily from Mary Louise Pratt's interpretation of the work of the sociolinguist, William Labov, who studied narrative patterns in the everyday conversation of inner

city minority groups.⁴⁸ I will represent the relationships among the elements as a hexagon (see fig. 1) because, in general, simple hierarchical tree diagrams, and other standard patterns, are not powerful enough to capture the complex semantic relationships generated by narrative.⁴⁹ Using a hexagon, then, is a way of leaving open the interrelationships among the elements of a narrative. It is important to note that one can apply the schema at many different levels – to a camera movement, composition, shot, sequence of shots, scene, sequence of scenes, etc. – depending on the size of the units that have been chosen for analysis. Narrative is a recursive organization of data; that is, its components may be embedded successively at various micro- and macro-levels of action.

The schema contains the following eight components, or functions, which may be repeated in various patterns to model our understanding of a given story; that is, one can move through the hexagon in a myriad of ways and any number of times.

- 1 An *abstract* is a title or compact summary of the situation which is to follow.⁵⁰ If an abstract is expanded, it becomes a *prologue*.
- 2 An *orientation* is a description of the present state of affairs (place, time, character) while an *exposition* gives information about past events which bear on the present.
- 3 An *initiating event* alters the present state of affairs. A narrative which delays orientation and exposition and begins with an initiating event, or a complicating action (see below), is said to begin *in medias res*.
- 4 A *goal* is a statement of intention or an emotional response to an initiating event by a protagonist.
- 5 A *complicating action* (linked to an antagonist) arises as a consequence of the initiating event and presents an obstacle to the attainment of the goal.
- 6 The *climax* and *resolution* end the conflict between goals and obstacles and establish a new equilibrium or state of affairs.
- 7 The *epilogue* is the moral lesson implicit in the history of these events and may include explicit character reactions to the resolution.
- 8 The *narration* is constantly at work seeking to justify implicitly or explicitly (1) why the narrator is competent and credible in arranging and reporting these events and (2) why the events are unusual, strange, or worthy of attention. In other words, how is it possible to possess the knowledge and why should it be possessed?

The study of how children acquire a narrative schema has been a fertile area of investigation and has greatly contributed to an understanding of mental schemas in general.⁵¹ The basic comprehension represented by a narrative schema is apparently not operative in most children until about the age of seven. Complex forms of narrative utilizing psychological

NARRATIVE SCHEMA

causation, point of view, multiple plots, and temporal complexity are not acquired until much later. Prior to the age of 7, there are partial realizations of the schema along with certain nonnarrative organizations of experience. In order to put the schema into context, I will briefly describe one researcher's attempt to sketch the stages of cognitive development leading up to the acquisition of narrative skills. The following stages of development may also be thought of as strategies by which to collect and associate data generally, ranging from creating a virtually random list through a chronology and finally to the causality and closure of a basic narrative. Although the essence of narrative is a presentation of systematic change through a cause and effect teleology, there is no reason that an actual narrative may not also contain some of these other ways of organizing data.⁵²

- 1 A *heap* is a virtually random collection of data or objects assembled largely by chance. Objects are linked to one another only through an immediacy of perception, a free-association of the moment.⁵³
- 2 A *catalogue* is created by collecting objects each of which is similarly related to a "center" or core. For example, a list of objects that belong in a particular room; or are used by a particular person; or are recorded in a particular time span (which yields a *chronology*). A list of personality traits in a novel helps define a "character." And, as shown in an earlier example, roses and violets may be collected together as flowers as well as placed in the category of what is only natural like sugar, sweetness, and "you."

One could perhaps think of the 180 degree rule in film as a catalogue of three-dimensional spatial fragments with a fixed relationship to a given hypothetical plane such that left-right orientation is preserved in all of the spatial fragments. The center also may be phonic,⁵⁴ or may be a particular action, such as, A does X to N; B does X to O; C does X to P (which yields the catalogue: A, B, C, N, O, P, with X as the "center" which justifies the list).⁵⁵

- 3 An *episode* is created by collecting together the *consequences* of a central situation: for example, collecting everything that happens to a particular character in a particular setting as well as everything that the character does in that setting. Unlike a heap or a catalogue, an episode does not simply grow longer, it shows change; it develops and progresses. Because the parts of an episode are defined through cause and effect, it is easier to *remember* an episode than to remember the miscellaneous parts of a heap or a catalogue.⁵⁶
- 4 An *unfocused chain* is a series of cause and effects but with no continuing center. For example, character A is followed for a time, then character B, then character C. (Consider, for example, Max Ophuls's *La Ronde*, 1950; Luis Buñuel's *The Phantom of Liberty*, 1974; and Jim

NARRATIVE COMPREHENSION AND FILM

Jarmusch's *Mystery Train*, 1989. On a smaller scale, consider certain elaborate camera movements by, for example, Ophuls, Renoir, Welles, Mizoguchi, Godard, and Jancsó.

- 5 A *focused chain* is a series of cause and effects with a continuing center. For example, the continuing adventures of a character, the events surrounding an object or place, or the elaboration of a theme.
- 6 A *simple narrative* is a series of episodes collected as a focused chain. Not only are the parts themselves in each episode linked by cause and effect, but the continuing center is allowed to develop, progress and interact from episode to episode. A narrative ends when its cause and effect chains are judged to be totally delineated. There is a reversibility in that the ending situation can be traced back to the beginning; or, to state it another way, the ending is seemingly entailed by the beginning. This is the feature of narrative often referred to as *closure*.⁵⁷

Without attempting to overly simplify these six ways of associating data, one can discern at least four different notions of time at work. The "heap" and "catalogue" primarily exploit two types of time: an atemporal, descriptive time⁵⁸ where elements are deemed to be simultaneous, and/or a chronological time of duration where elements are deemed to be merely consecutive. From "episode" to "simple narrative," however, time becomes increasingly consequential (implications, probabilities) and thus directional ("effects" never precede "causes").⁵⁹ And finally, time comes to exhibit a large-scale configuration (symmetry, reversal, parallelism, cycle, closure, unity) comparable, say, to the action of Todorov's transformations in creating principles of order out of local causation. The spectator's experience of duration and causality is forward in time while the spectator's experience of order may reverse the arrow of time, seemingly operating on the present from a point in the future; that is, earlier parts of a pattern have been arranged to fit with later parts.⁶⁰

I believe that simultaneity, duration, causality, and order are not simply four items in a taxonomy of time, but are the results of four specific ways of *processing* data. Since the actual processing of data is not available to awareness, our experience of these aspects of time is better described as a diffuse *affect* responsive to the complexity of juxtaposition allowable under a particular method of associating data.⁶¹

THE GIRL AND HER TRUST

As an extended illustration of some of the above ideas concerning a narrative schema and its creation of different types of temporal experience through different ways of associating data, I wish to consider

15-min
Griffit
narrat
ence c

The
the fa
or "st
adven
an int
graph
- whc
key. M
been t
trust.

promj
mana;
to orc
Grace
thrille
ence t
an ini
may b

Gr
'Nati
This i
will b
mess
allow
ecome
appr
Gulle
Brot
Botic

the
the
the

the
the
the

NARRATIVE SCHEMA

tain
oir,
ter.
ents
ain.
use
ress
its
is a
the
ngly
ften

15-minute film that spectators easily recognize as a narrative – D.W. Griffith's *The Girl and Her Trust* (1912). I will concentrate on how a narrative schema works with local data on the screen to produce coherence on a large-scale.

The title of Griffith's film – functioning as an "abstract" – alludes to the fact that the film will actually be *two*, intertwined focused chains, or "stories": a romance involving an unmarried girl coupled with an adventure which tests her trustworthiness. The first shot of the film is an intertitle which begins the process of orientation: "Grace, the telegraph operator, is admired by all." The next shot shows a young woman – whom we now assume is Grace – busily reading near her telegraph key. Not only do we see that she is charming and graceful but we have been told that she has an inner beauty: an integrity worthy of complete trust. A would-be admirer enters the room but his clumsiness soon prompts Grace to send him scurrying. Next, the handsome station manager enters, but his aggressiveness in stealing a kiss causes Grace to order him to leave the room. After he has left, however, we see Grace's reaction change from being deeply offended to being secretly thrilled. She smiles and presses two fingers to her lips as if to reexperience that sudden kiss. Within the romance story, there has clearly been an initiating event which has aroused the heroine to form a goal which may become a match for the explicit goal of the handsome protagonist.

Grace's reverie is interrupted by a telegraph message. It reads, "National Bank sending \$2000 on No. 7 for Simpson Construction Co."⁵⁹ This is the initiating event for a new line of action – a crime story. Grace will be entrusted with money arriving on the next train. The telegraph message, however, *causes* the romance story to take a new turn. Grace allows the hero to return to her office to help plan for the arrival of the money. The advent of the crime story has allowed him to again approach within range of stealing a kiss. He takes out a revolver and bullets. Grace is nervous and afraid of the gun and shrinks back as he enthusiastically loads the gun very near her body (see fig. 2). With this gesture he has *regained* the authority and initiative he lost when Grace first banished him from the room. Correspondingly, she has now lost some of her brashness and control over the situation. They are coming closer to being a match for each other; that is, to have matching goals.

When the train arrives, we see something that no one else sees. Two tramps sneak off the train and hide. They plan to steal the money. This complicating action represents a goal hostile to Grace's obligation to protect the money. Grace, however, is unaware of the complicating action. The hero offers the revolver to her but she declines, "Danger? Nothing ever happens here." These words illustrate the ever-present activity of narration for the words are placed in a context which allows the spectator to appreciate the heroine's mistake. We immediately



Figure 2 The Girl and Her Trust

reinterpret the words as saying to us, "Something dangerous (and interesting) will happen here!" The narration has demonstrated a power to know events by creating suspense (when will Grace realize the danger?), and has promised to repay our attention by exciting action. By operating from outside the diegetic world, the narration regulates our access to that world and thus produces effects based on knowledge not available to the characters.

The hero leaves the station on an errand and the train also leaves. Grace is alone. We have already seen the tramps watching the money and spying through a window at the hero. An intertitle announces "The tramps' opportunity." Why should the intertitle tell us something that is already obvious? The intertitle functions as an abstract for a new phase of the action and serves to explicitly mark the boundary between episodes. It aids the spectator in segmenting each of the two stories into abstract, orientation, initiating event, etc. The romance story is giving way again to the crime story and the spectator is being provided ample time to redirect his or her expectations. The overall narrative is being broken into component parts for the spectator and arranged into a hierarchy where general propositions about the events can be progressively modified and filled in by details. The crime story can now be elaborated.

We see the tramps spy on Grace through a window. The composition

NARRATIVE SCHEMA

of this shot has been carefully contrived by the narration to show the tramps in the background peering through the window while Grace is in the foreground where her reactions are clearly visible and reveal that she has not yet seen the tramps. Since the spectator already knows about the presence of the tramps, new and valuable information in the scene will come only when Grace discovers them, and the spectator is perfectly placed to witness *that* event of recognition. Like the intertitle ("The tramps' opportunity"), this shot functions as an abstract since it reveals in a concise manner the elements and relations that will generate the succeeding action: a drama of seeing and being seen followed by a drama of breaking down (various) barriers separating the tramps from Grace. This shows that a narrative schema may be applied to actions on a small-scale: an action is named (e.g., "robbery"), put into context, initiated, responded to, opposed by a new action, and so forth, until the ebb and flow of action, goal, and reaction results in a temporary equilibrium allowing the next phase of the action to commence (with a new abstract, orientation, initiating event, and so forth).

The narration is planning a surprise. We expect the tramps to threaten Grace, but if our predictions always turned out to be correct, we would lose interest in the story – its focused chains would become too obvious and trivial. The tramps will indeed attempt to reach Grace and the money; but later the unexpected will occur when it will be she who breaks down barriers to reach them. But first, a drama of vision unfolds. Grace senses something behind her. She spins around, but the tramps have ducked down. She turns back, relieved. The tramps look again. She spins around again. They have ducked down. But they pop up again unexpectedly while Grace is still watching and are finally seen. Notice that this is not a plausible account of a robbery since there is no reason that the tramps shouldn't quickly sneak into the station or that they should risk looking through the window more than once. Instead of plausibility the narration has lingered on the "goal" stage of a narrative schema, creating a subjective reality by drawing out both the emotional impact on the heroine and the suspense felt by the spectator. Here Griffith's film goes beyond a simple narrative by dramatizing the heroine's interior reality. Griffith treats psychology itself as an embedded form of story and, moreover, a form of story that seems to dictate what is plausible.

A series of complicating actions occurs. The tramps rush to get into the station but Grace runs from her office to barricade the door. They break down the door but Grace runs back into her office and barricades that door. She sends a telegraph message for help. The tramps react by cutting the telegraph lines and return to batter her door. She repulses them with a bullet placed in the keyhole and detonated with the point of a pair of scissors. We are surprised by Grace's unusual solution to

NARRATIVE COMPREHENSION AND FILM

her problem which creates an effect on us something like the opposite of suspense; a character has been shown to have knowledge superior to the spectator. (Consider the change in effect if the narration had earlier presented the hero explaining to Grace that in an emergency she could use the keyhole as a gun barrel. These effects of narration will be considered in more detail in chapter 3.) Grace's ingenuity partially makes up for her failure to keep the revolver when it was offered to her by the hero. She is therefore not humiliated and will remain a match for the hero when that phase of the story resumes.

The tramps fail to capture Grace, and decide simply to take the locked strongbox and open it elsewhere without the key she possesses. This does not end Grace's ordeal, however, for it causes her to be tormented by a sense of duty. She decides that she must stop them. However, there is an obstacle: she discovers that she is now imprisoned in the office and so must break *out* to pursue the tramps. This effectively reverses the initial situation with the tramps breaking *in*. If the story were to end here, it would be construed as one about Grace's failure to respond (paralyzed and trapped in her office) or the failure to achieve a goal (she runs after the tramps but fails to catch them). She would still be a heroine for her valiant struggle and still be admired by all (and perhaps one person in particular). The difficulty would be that the chain of events of the crime story, focused through her psychological drama, would be incomplete with respect to expectations raised in the spectator that Grace's self-imposed devotion must be tested and renewed by events (does she actually deserve the admiration she receives?). Accordingly, the story does not end but begins a new phase as the narration searches for a resolution to close off all major questions and possible lines of action.

Grace rushes after the tramps who are attempting to escape on a railroad handcar. They beat her senseless but she clings to the handcar as it speeds down the tracks. An intertitle has already evaluated her conduct: "She risks her life for her trust." Meanwhile her telegraph message has resulted in a train being sent to provide aid, and the hero returns to the station just in time to leap aboard the train and direct it in pursuit of Grace and the tramps. The train eventually catches the handcar. Grace embraces the hero while the train engineer beats one of the tramps senseless. This task falls to the engineer because the hero belongs to the romance story and his aggressiveness (e.g., in stealing kisses) must now be moderated; a minor character can dispense the prosaic details of justice. The film ends as the engineer returns to the train while Grace and the hero jump onto the front of the locomotive above the cowcatcher (where their emotional reactions can be clearly seen by the spectator). There is no indication what has happened to

NARRATIVE SCHEMA

the tramps; or rather, enough has happened to them to resolve the crime story and now the romance must be resolved.

The final shot of the film begins as a medium two-shot of Grace and the hero in earnest and happy conversation. He pulls a sandwich from his pocket and offers it to her. She happily takes a bite as the train is backing away from the camera into an extreme long shot. In the opening scene of the film, Grace had offered the hero a drink of soda pop which he had pretended to accept in order to bend near her and steal a kiss. Now, over the sandwich, *she* steals a kiss from *him*. A symmetry is completed in which the ending situation balances the opening.⁶² They can dine together and indulge their appetites for he has proven his intentions were good and she has proven her worthiness as a recipient of those intentions. The locomotive suddenly sends out jets of white steam that swirl in front of the couple as the image fades to black. The archetypal extreme long shot which closes most classical films announces that we cannot see more details because the causal chain has played itself out.

In one sense the above action of the final two shots of the film, following the rescue and the beating of the tramps, seems trivial and redundant. In a deeper sense it allows the spectator time to formulate a response to the entire sequence of events and to appreciate that the sequence is in fact a powerfully focused chain of episodes which, like every narrative, reenacts "cultural beliefs about success."⁶³ The crime and romance stories have been *merged* in the epilogue in such a way that the heroine's place in an economic order (as trustee of capital) converges on her place in a social order (as an unmarried woman who must learn to trust in a man). The tramps exist as a catalyst to bring the man and woman together, to transform the woman as guardian of capital into property to be guarded. Grace herself has actually done nothing to delay the tramps in their flight or aid in their capture on the handcar. Her gesture has been one of pure devotion and sacrifice. Like many other Griffith heroines, she earns a man through noble self-sacrifice, strength of will, and patience, not through challenging the preconceptions of her society. As for the tramps, they are condemned for threatening the values and property of the middle and upper classes. To survive they will need property and wives. As Sergei Eisenstein recognized, Griffith perceived society "only as a contrast between the *haves* and the *have-nots*" and this contrast went "no deeper than the image of an intricate race between two parallel lines" of rich and poor.⁶⁴ Griffith's message and the schematic form of his narrative are essentially conservative and familiar even if some of his methods were innovative.

CAUSALITY AND SCHEMA

I would like to examine the notion of narrative causality in somewhat more detail and link it to the notion of a mental schema. Although *The Girl and Her Trust* presents events in chronological order and inventories certain actions (e.g., eating, spying, trusting, beating, and kissing), we do not perceive the film as a catalogue. The reason, of course, is that the film's events are principally defined through cause and effect (event B *because of* event A; scene Y *because of* scene X). The film's events are linked together by probability whereas elements of a heap or a catalogue are all equally likely with no single element necessary. Since episodes and narratives require a perceiver continually to assess and evaluate *probabilities* moment by moment, one might define narrative broadly as "an accepted technique for discussing the chances of life."⁶⁵

Probability, however, comes in many degrees and types, and hence many sorts of causality may be appropriate in comprehending a narrative. Noël Carroll observes:

Since most film narratives involve a series of actions, it may seem natural to think that causation is the major connective between scenes in movies. However, it is implausible to suggest that scenes follow each other in most film narratives via a chain of causal *entailments*. I would guess that most succeeding narrative scenes are causally underdetermined by what precedes them. Rather the connection is weaker than a causal one.⁶⁶

Several questions arise: can events or scenes, which are themselves only moderately likely, be strung together to create another event or scene – say, a climax and resolution – which is obligatory? What kinds of judgments can be made about the *relative likelihood* of particular events appearing together?

Informally, one can imagine a spectator's causal evaluations as falling along a spectrum:⁶⁷

- 1 Elements are merely consecutive ("and"); their order is arbitrary or optional.
- 2 Elements are chronological ("then"); order is governed only by duration.
- 3 Elements appear together conventionally; order is set by familiar social or generic practices.
- 4 An element (the "remote" cause) appears together with another element but only through the mediation of many other elements ("intervening" causes), the last of which is the immediate or "proximate" cause.
- 5 An element is necessary for the appearance of another (an "enabling" cause).

NARRATIVE SCHEMA

- 6 An element is sufficient for the appearance of another (a "direct" cause). If two or more elements are present, each of which is sufficient to cause a particular outcome, the effect is "overdetermined."
- 7 An element is both necessary and sufficient for the appearance of another (a "unique" cause).

L.B. Cebik imagines a complex but not atypical example of causality in narrative:

If we let capital letters stand for events, objects, actions, and conditions, as appropriate, we might string together a perfectly understandable narrative in which A may cause B, B anticipate C, C accompany D, which may give E reason to go through F, G, and H as steps toward the achievement of I, where I forms the motive for J to do K, with the unintended consequence L, which marks the beginning of an M, an M that ends when N marks the shift to O, which along with P and Q comprises evidence for the occurrence of R, which causes S . . .⁶⁸

Our experience of causality, then, depends upon our assessment of various probabilities. On a small scale, the connections among events may be quite weak and indirect even though on a large scale, an overall pattern may be evident. Recall that for Todorov "causality," or the logic of mere succession, was not enough to define a narrative: change must also emerge on a large scale in the form of a "transformation" among events. This global aspect of change in narrative is obscured when we describe causality as a *chain* of causes and effects. Although convenient, the chain metaphor focuses too much attention on local determinants, tends to make connections too strong, and remains uncomfortably close to the methodology of Behaviorism which posits a simple linear sequence of stimuli and responses. More promising, I believe, would be a psychological approach to narrative that would give equal weight to "top-down" frames of reference for grouping elements, that is, to principles and criteria that are not determined solely by local conditions but instead are responsive to larger contexts.

It is evident that in many cases our assessment of the "probabilities" draws upon broad *cultural* knowledge in judging which actions and transactions are *acceptable* as belonging together (and hence likely to occur together). Thus our comprehension of narrative causality - of what may follow what - may depend upon our general knowledge of social interactions; that is, the connections we are predisposed to call "causal." Roland Barthes argues that

the logic to which the narrative refers is nothing other than a logic of the *already-read*: the stereotype (proceeding from a culture many centuries old) is the veritable ground of the narrative world, built

NARRATIVE COMPREHENSION AND FILM

altogether on the traces which experience (much more bookish than practical) has left in the reader's memory and which constitutes it. Hence we can say that the perfect sequence [of actions], the one which affords the reader the strongest logical certainty, is the most "cultural" sequence, in which are immediately recognized a whole *summa* of readings and conversations. . . . Narrative logic, it must be admitted, is nothing other than the development of the Aristotelian *probable* (common opinion and not scientific truth); hence it is normal that, when an attempt is made to legalize this logic (in the form of esthetic constraints and values), it should still be an Aristotelian notion which the first classical theoreticians of narrative have advanced: that of *verisimilitude*.⁶⁹

I wish to indicate briefly the great complexity of the "causes" at work even in a narrative as simple as *The Girl and Her Trust* and to illustrate that the "logic" involved is often based on what seems familiar and natural within a culture, within a way of life. Consider the variety of connections that we must imagine in justifying each of the following pairs of actions: the hero returns to the station from his errand *just as* an unscheduled train arrives; Grace is reading a book, *then* a suitor walks into the room; the suitor *offers* Grace a bottle of soda pop (intending to initiate a socially defined courtship ritual), *and then* offers a straw (a second act to solicit a response); stealing a kiss *causes* a man to be ordered out of the room; continued spying in a window *causes* two tramps to be seen; Simpson Construction Company's need for \$2000 *causes* Grace to be beaten senseless and to kiss a man; white steam rises around Grace *while* she rides on the front of a locomotive. How does a spectator establish the proper context within which to judge the relative pertinence of such pairs of actions occurring together? Steam necessarily rises because it is a hot gas (and perhaps also because of the exuberance of the train engineer) and yet there are still other, more important reasons why it curls around the triumphant, virtuous, and passionate couple as the image fades to black. Establishing relevant contexts within which to evaluate causation is partly a matter of *segmentation*: how is something to be divided into parts that can be seen to interact? To anticipate my answer, I will state that an individual segments an event according to an explicit, or implicit, theory (or theories) of experience. A theory in this sense need not be as rigorous as a physics or a philosophy; all that is required is that a "fact" be given under a description and that we know how to produce descriptions. One of the important ways we produce descriptions is by utilizing schematic forms of knowledge.

In making a judgment about what properly goes with what, and in what context something can be seen in terms of something else, a

NARRATIVE SCHEMA

perceiver is also implicitly making a choice about how elements should be selected and how they should be grouped on levels from the smallest to the largest scale of action. In understanding the world of the story, a perceiver may link two elements together to make a pair even if the two elements do *not* appear together on the screen, or in the plot. (Similarly, elements appearing together on the screen may not belong together. Thus a device like editing has a psychological dimension and cannot be defined in strictly material or formal terms.) Here is where a narrative schema becomes important. It helps direct our search for pertinent causes by proposing a segmentation applicable on many scales of action and then by "filling in" any connectives that are missing from the surface structure.⁷⁰ We discover and justify connections among narrative elements with respect to such schematic functions as goal, reaction, resolution, epilogue, and narration. Of course, a narrative schema does not provide all the answers: one still needs to weigh the evidence from the text as well as be acquainted with nonnarrative schemas and specific cultural knowledge (e.g., the routines of courtship, and texts about courtship like the limerick of a woman riding on a tiger or the poem, "Roses are red"). One cannot use a schema as a search procedure without searching for and through some domain of knowledge.

In thinking about the types of causation which make up a narrative, one is led toward a deeper interpretation of a narrative "schema." Just as it is people who refer, not sentences, so it is people who judge plausibility, realism, and causal connection. What is familiar and real to an individual depends upon the regularities in that individual's environment which are judged to be important. Causes and effects fit together when they are part of an individual's plans and goals. Actions that occur become trapped within a cultural lexicon of human thoughts and deeds: accident, opportunity, hindrance, aspiration, decision, attempt, defeat, success, and so forth. A narrative schema, together with a host of related schemas,⁷¹ encapsulates the interest we take in the world as humans. These schemas are a way of working through cultural assumptions and values. Thus "causes and effects" emerge, as it were, after the fact as explanatory labels for a sequence of actions viewed under a particular schematic description. (To have fallen in love, it was sufficient to. . . . In order to protect one's trust, it was necessary to. . . .) Our concept of narrative causation must be powerful enough to include these social and ethical factors so that when a statue accidentally falls and kills a man, we can also see that in another sense it was not an "accident" at all when the statue is later revealed to be that of a woman who was murdered by the man.⁷² In this sense the epilogue of a narrative merely makes explicit the social judgment already contained within the causal chain; or, perhaps it would be better to say, the

NARRATIVE COMPREHENSION AND FILM

spectator makes a judgment about probabilities based upon life as experienced through the probabilities of his or her society.⁷³

In *The Girl and Her Trust* two major causal chains involving crime and romance have been tightly wound together. Such a double causal structure is typical of the classical narrative film.⁷⁴ What is the relationship between such a pair of causal structures? Why not create interest only through a single focused chain of episodes?

One advantage of a double causal structure is that a narrative schema may actually operate more efficiently with two causal chains. Intertwining two stories provides the spectator with more ways to imagine causal connections and more opportunities for the overall story to advance. To clarify this idea consider an analogy with a character who has two motives to do something. Raymond Durnat has argued that certain genres use "double motivation" (and even greater multiples of motivation) to portray the relationship between character goals and events. Double motives *prevent* the spectator from making a precise psychological analysis because one motive may be operating without the other, or if both must operate, the spectator does not know their relative contributions. This allows one motive to appeal to individuals who wouldn't respond to the other.⁷⁵ Perhaps more importantly, whatever a spectator *first believes* may be enough to drive the story forward. Just as essential plot details are usually repeated several times to promote clarity,⁷⁶ so a variety of motivations circulating in the text may be useful options in filling out, and making definite, causal sequences. If the text can suggest enough "intervening" and "enabling" causes (see above), a narrative schema will tend to generate a resolution which can be imagined as the closure of a "unique" cause implicit in the opening of the story. This allows the story to be made "unique" in many different ways to many spectators!⁷⁷

Grace's reaction as the hero brings a gun near her illustrates double causation at work. The composition which shows her reaction is carefully arranged and held as a tableau in the film (see fig. 2). Because of the preceding events, it is not really possible to say in what proportion her fear is caused by the gun and/or by the man's approach near her. Earlier when he approached and abruptly kissed her, she reacted with great anger; after he had left, however, she expressed an emotion of secret thrill. This ambivalence leaks into the film on many levels, providing several pathways along which the narrative can be seen to reach a definite resolution and epilogue. More importantly, if the spectator applies this notion of psychological ambivalence to the double causal structure of the story itself, the resolution of the film can be seen as a subtle transformation of Grace's initial situation. A second advantage of a double causal structure, then, is that it facilitates a more complex ending and epilogue than, say, the mere defeat of a threat. In the

NARRATIVE SCHEMA

context of crime and romance, Grace may be seen as defending herself from thieves who seek to steal from her both money (the tramps) and kisses (the hero). Since the tramps provide a condition for the success of the hero (who also triumphs over another suitor), they can be seen as being totally unsuitable suitors who represent the ultimate reversal and perversity of courtship. (The threat of rape, though never made explicit in the film, is ever-present.) Thus the two causal lines allow complex metaphorical comparisons to form between, say, the goals of protecting hard-earned property, earning a mate, and being/becoming protected. Thus Grace's ambivalence (expressed well in fig. 2) functions as a "psychological" cause which finally unifies interior and exterior realities as she overcomes her initial fear(s) through both decisive action and submission. She is doubly rewarded at the end for her travail by a happy excitement caused by a job well done and a man well earned.

In summary, the double causal structure has two functions. It is *efficient* in producing an effect on a large number of spectators because the vagueness of multiple motivations grants individual spectators a limited freedom in singling out the local causes that bind a beginning, middle, and end. It is also *efficacious* in producing a desired effect because the expansive metaphors that are encouraged are aimed at matching a spectator's own goals and desires in watching and comprehending.

It is not enough, however, merely to juxtapose two motives and two causal chains, and hope for the best. The narration of the story must also *block* other possible combinations among goals and actions. When the tramps besiege Grace in her inner office, an intertitle explains, "The tramps want the key to the express box." The narration here seeks to *limit* the scope of metaphors based on property: we are assured that the tramps do not want to rape the woman nor attack her because of her economic class. Nonetheless, there is a sense in which such expansive metaphors cannot be entirely contained; they remain as symptoms of the underlying fears raised by the events of the narrative even if denied by an intertitle. Thus it would be better to think of narration not as a single process, but as several processes moving on different levels, proposing and abolishing contradictions with varying degrees of explicitness and success. These contradictions, anxieties, and vague metaphors are often important clues to our comprehension of a narrative and to our complicity in it. As we shall see in the next chapter, narration is able to create complex, even apparently "impossible," causal schemes which nevertheless exemplify the mapping of a powerful will and desire onto a world of objects. Narration, therefore, is not really on a par with the other elements of a narrative schema, but rather stands in for the operation of a still deeper schema that drives the story. Although the deeper schema may speak about a character and his or

NARRATIVE COMPREHENSION AND FILM

her goals, it must always be speaking to and for the multiple desires and fears of an author who must tell and a spectator who already knows.

STORY WORLD AND SCREEN

A PRELIMINARY DELINEATION OF NARRATIVE IN FILM

Narrative in film rests on our ability to create a three-dimensional world out of a two-dimensional wash of light and dark. A bare facticity of graphics on the screen – size, color, angle, line, shape, etc. – must be transformed into an array of solid objects; and a texture of noise must be transformed into speech, music, and the sounds made by solid objects. Light and sound in narrative film are thus experienced in two ways: virtually unshaped *on* a screen as well as apparently *moving within*, reflecting and issuing from, a world which contains solid objects making sounds. Every basic spatial and temporal relationship, such as position and duration, thus has a double interpretation. A green circle might be seen to the *left* of a square in the same plane, or alternatively, it might be seen to lie *behind* the square along a diagonal line to the left. In the latter interpretation, the circle may become a “sphere,” the square a “box,” and “size” and “color” will be adjusted according to our judgments about how distance and light are being represented in a given perspective system. Similarly, the green circle may appear for ten seconds on the screen but represent many hours of world time for the green sphere, especially if there is no other “action” by which to gauge duration. Rudolf Arnheim asserts:

It is one of the most important formal qualities of film that every object that is reproduced appears simultaneously in two entirely different frames of reference, namely the two-dimensional and the three-dimensional, and that as one identical object it fulfills two different functions in the two contexts.¹

The spectator, therefore, encounters at least two major frames of reference in film: the space and time of a *screen* as well as (a sample of) the space and time of a *story world*.² More than space and time, however, is at stake. Causality also has a double interpretation. Changes in light and sound patterns will be perceived in at least two ways: as motion

across a screen and as movement among objects in a story world. Causality on a screen will involve patterns of a purely visual, phenomenal logic where, for example, one blob smashes into another but the resulting transformations in motion and color may not be analogous to the interactions of three-dimensional objects like billiard balls; the blobs may even "pass through" each other on the screen. Bizarre pictorial compositions and animation are clear examples of on-screen causality. In short, light and sound create two fundamental systems of space, time, and causal interaction: on a screen and within a story world. One of the tasks of narrative is to reconcile these systems.

It seems likely, in fact, that more than two frames of reference are active in our comprehension of film. It has even been argued that there is a stage of visual processing located halfway between two and three dimensional perception which produces a $2\frac{1}{2}$ dimensional representation of space.³ Clearly, major changes occur during the conversion from phenomenal appearances on the screen to functions in a story world. One of the essential tasks of a narrative theory is to specify the various stages through which we represent and comprehend a film as a narrative. On-screen patterns of light, sound, and motion do not denote and hence cannot be true or false; they are fully present and neither narrative nor fictive. Moreover, the time in which these patterns are present on the screen is determined initially by the film projector. By contrast, a story builds complex spatial and temporal contexts, makes references to things which are not present (and may not exist), and allows broad conclusions to be drawn about sequences of actions. Moreover, time in the world of the story may be quite different than the time of the projection of the film. For example, in *Letter from an Unknown Woman* (Ophuls, 1948) screen time is ninety minutes while the story covers three hours of an early morning during which a letter is read, and the letter, in turn, dramatizes events spanning fifteen years at the turn of the century in the world of Vienna.

Many concepts have been proposed to help describe how on-screen data is transformed through various spatial, temporal, and causal schemes culminating in a perceived story world. The various stages have been described with concepts like script, set decoration, technology, technique, performance, material, shot, form, style, plot, diegesis, code, narration, and referent. Since nonnarrative ways of organizing data may coexist with narrative, one might also recognize a conflict among discursive schemes, an "excess" within the story.⁴ The processing of film data has an important effect on how a spectator feels about the conceptual structures which are ultimately constructed. Some of the metaphors offered by film theorists suggest that our comprehension of film proceeds only forward, one step at a time, and depends simply on local and immediate juxtapositions, but other metaphors are less

STORY WORLD AND SCREEN

restrictive. Rudolf Arnheim speaks of picture postcards in an album while Noël Burch speaks of picture postcards suspended in a void, "radically autonomous."⁵ Early in his career Eisenstein argued that shots are perceived not next to each other in a horizontal or vertical chain, but on top of each other in collision.⁶ Later he refined the idea to include layers of pictures "rushing towards the spectator," but not necessarily in a straight line. He proposed that film data might be perceived as arranged vertically in matrix form, exhibiting a multiplicity of criss-crossing relationships in an instant.⁷ Finally, the psychologist Julian Hochberg mentions three types of perceptual analysis in film: simple summation, directional patterns, and cognitive maps.⁸ With one exception, I will not explore these sorts of idea now, but rather will consider them in later chapters in the context of particular narrative theories.

It will be useful now, however, to separate the concept of "story world" into two parts: the diegetic and the nondiegetic. In talking about a "story," we often refer to certain events which surround a character, events which have already occurred, or might occur in a particular manner, in a certain sequence and time span, and so forth. We understand such events as occurring in a "world" governed by a particular set of laws. I will refer to that imagined world as the *diegesis*. The spectator presumes that the laws of such a world allow many events to occur (whether or not we see them), contains many objects and characters, contains other stories-about other persons, and indeed permits events to be organized and perceived in nonnarrative ways. (Later I will argue that a documentary film also creates a diegetic world for its events.) The diegetic world extends beyond what is seen in a given shot and beyond even what is seen in the entire film, for we do not imagine that a character may only see and hear what we observe him or her seeing and hearing. The diegesis, then, is the implied spatial, temporal, and causal system of a character – a collection of sense data which is represented as being at least potentially accessible to a character.⁹ A sound in a film, for example, is diegetic if the spectator judges that it has been, or could have been, heard by a character. However some on-screen elements (e.g., "mood" music) are *nondiegetic* and addressed only to the spectator. These elements are *about* the diegetic world of a character and are meant to aid the spectator in organizing and interpreting that world and its events. Nondiegetic elements are not accessible to any of the characters. The spectator's organization of information into diegetic and nondiegetic story worlds is a critical step in the comprehension of a narrative and in understanding the relationship of story events to our everyday world.

Let us now attempt a preliminary delineation of narrative in film. This definition will aid us in examining narrative comprehension more

NARRATIVE COMPREHENSION AND FILM

precisely and will also provide a basis in chapter 4 for outlining five recent types of narrative theory, each of which stresses and interprets a different aspect of the narrative process.

Narrative is a way of comprehending space, time, and causality. Since in film there are at least two important frames of reference for understanding space, time, and causality, narrative in film is the principle by which data is converted from the frame of the *screen* into a *diegesis* – a world – that frames a particular *story*, or sequence of actions, in that world; equally, it is the principle by which data is converted from story onto screen. To facilitate analysis, narrative may be divided into a series of relationships. For example:

1 The relationship of diegesis and story may be analyzed with such narratological concepts as Todorov's "transformations," or a *narrative schema*. Which kinds of action sequences occurring in what kind of world will qualify as a narrative? For example, a narrative schema ("abstract," "orientation," "initiating event," etc.) describes how a reader collects a series of episodes into a focused causal chain (as opposed to a "heap," "catalogue," "unfocused chain," etc.).

Causal chains are not just sequences of paired events, but also embody a desire for pairing events and the power to make pairs. Narrative causes ("remote," "intervening," "enabling," etc.) are thus principles of explanation which are derived from cultural knowledge as well as from physical laws. Narrative causality includes the human plans, goals, desires, and routines – realized in action sequences – which are encouraged, tolerated, or proscribed by a community.

2 The relationship of diegesis and screen may be analyzed with such concepts as script, set decoration, technology, technique, shot, form, style, material, and excess. The present chapter will demonstrate that these kinds of concept may be approached by measuring their effects on a spectator's judgments about the *ordering* of space, time, and causality on the screen and in the diegesis.

3 The relationship of diegesis and what is external to it – the nondiegetic – raises issues of *narration*: from what sort of "other world" has a diegesis been created, a character presented, events told? What has been concealed, or excluded? And, furthermore, how do we come to *believe* in a narrative diegesis and relate it to our own world; that is, what is the nature of *fictional and non-fictional* reference?

STORY WORLD AND SCREEN

The previous chapter introduced some of the issues involved in analyzing diegesis and story (i.e., 1 above). In later chapters, I will deal with the problems of narration and fiction (3). But first we must examine the relationship of diegesis and screen, namely, how is data on the screen transformed into a story world? In order to answer this question, a distinction will be made between two types of perception operative in watching a film. These two types of perception produce different kinds of hypotheses about space, time, and causality. Distinguishing between them will allow us to examine closely how a spectator makes separate use of judgments about space, time, and causality, as well as how a spectator may integrate these judgments to produce an overall narrative rendering of experience.

TOP-DOWN PERCEPTION

The movement from screen to story world does not proceed along a smooth path and in only one direction. Many of our abilities are brought to bear simultaneously on a film, producing at least some conflict and uncertainty. As a first step toward unravelling some of these abilities and specifying the kinds of conflicts that arise, I will use a fundamental cognitive psychological distinction to divide perception into two kinds of process according to the "direction" in which they work. Some perceptual processes operate upon data on the screen in a direct, "bottom-up" manner by examining the data in very brief periods of time (utilizing little or no associated memory) and organizing it automatically into such features as edge, color, depth, motion, aural pitch, and so on. Bottom-up perception is serial and "data-driven," and produces only short-range effects. Other perceptual processes, however, are based on acquired knowledge and schemas, are not constrained by stimulus time, and work "top-down" on the data, using a spectator's expectations and goals as principles of organization. Top-down processes are indirect in the sense that they may reframe data in alternative ways independently of the stimulus conditions which govern the initial appearance of the data. Top-down processes must be flexible and general in order to be effective across a wide range of situations while allowing for (unpredictable) variations among specific cases. Top-down processes often treat data as an inductive sample to be projected and tested within a variety of parallel frames of reference while bottom-up processes are highly specialized and atomistic (e.g., detecting motion). Both kinds of process operate simultaneously on the data creating a variety of representations with varying degrees of compatibility.¹⁰

Because top-down processes are active in watching a film, a spectator's cognitive activity is not restricted to the particular moment being viewed in a film. Instead the spectator is able to move forward and

backward through screen data in order to experiment with a variety of syntactical, semantic, and referential hypotheses; as Ian Jarvie notes, "We cannot *see* movies without *thinking* about them."¹¹ By experimenting with various methods for ordering data, the spectator creates spatial, temporal, and causal experiences which do not derive directly from screen time. Also critical in top-down processing are procedures which test the degree of "progress" which has been made toward solving a perceptual problem. Such procedures are active, for example, when we search for the "end" of a story. If we are unable to detect progress, we may begin to doubt the particular techniques we have been using, or even whether we have properly understood the goal. Because of the diversity of top-down and bottom-up processes which may be at work at a given moment in a text, perception as a whole is perhaps best thought of as a system which struggles to manage different and often conflicting interpretations of data.

In addition, the fact that comprehension may be divided into top-down and bottom-up kinds of activity helps explain some inconsistencies in the terminology employed by film writers. For example, some writers prefer to use the concept of "voice" in film as a means to identify the source of words that are actually heard by a spectator while other writers prefer to apply the concept more broadly in order to include a number of top-down factors that influence a spectator's perception. Bill Nichols argues for an expansive notion of "voice."

[I]n the evolution of documentary [as a genre] the contestation among forms has centered on the question of "voice." By "voice" I mean something narrower than style: that which conveys to us a sense of a text's social point of view, of how it is speaking to us and how it is organizing the materials it is presenting to us. In this sense "voice" is not restricted to any one code or feature, such as dialogue or spoken commentary. Voice is perhaps akin to that intangible, moire-like pattern formed by the unique interaction of *all* a film's codes, and it applies to *all* modes of documentary.¹²

For Nichols, the concept of "voice" is not confined to words literally spoken, or written, nor confined to fictional narrative; instead, "voice" includes powerful, nonverbal patterns even in nonfiction (documentary) films. Accordingly, the "person" whose voice is "heard" in a text may be a much more complex (invisible and inaudible) entity than a voice-over narrator or someone being interviewed. Thus Nichols's approach is well-suited to an analysis of narration. As I will emphasize in this book, film narration cannot be limited to, say, an explicit commentary, or defined by literal, material, purely formal, stylistic, technical, technological, or "bottom-up" kinds of categories. Narration, and narrative, are preeminently top-down phenomena that require for their

STORY WORLD AND SCREEN

analysis the use of wide-ranging, complex concepts like "point of view," or Nichols's "social point of view."

When we think of narrative as a general phenomenon that may appear in many physical forms (conversation, pictures, dance, music, etc.), we are thinking of it as a top-down cognitive effect. Wallace Martin may put it too strongly when he says that "narratives may be the source of the varied visual resources of the movies, rather than vice versa."¹³ Nevertheless, much can be learned by concentrating on top-down processes in an attempt to isolate the psychological conditions that allow narrative to be understood in all media.

I wish to examine some of the top-down processes which seem to be relevant to our comprehension of narrative. I will begin by considering how our top-down search for a coherent causal system helps to organize screen data into diegetic and nondiegetic story worlds, each with a coherent temporal system. Later in the chapter, I will consider how judgments about screen space are related to judgments about story space and the causality of a story world. In general, we will discover that conflicts arise between top-down and bottom-up processing, between story and screen, and between the diegesis and what seems external to it. Hence we will find that the ongoing process of constructing and understanding a narrative is perhaps best seen as the moment by moment regulation of conflicts among competing spatial, temporal, and causal hypotheses. "Narrative" in film is therefore the overall process as well as the result of searching among hypotheses for an equilibrium, however precarious.

TEMPORAL AND SPATIAL ORDER

There is a sequence in *The Lady from Shanghai* (Welles, 1948) where three distinct actions are intercut through fifteen shots in such a way that it appears that when a woman presses a button, a door flies open allowing a dying man to drag himself into a room; when she presses the button again, a car is sent speeding down a road as a truck pulls up to a stop sign; and, when she presses the button a final time, the car is sent crashing into the back of the truck as the two men in the car react with horror at their helpless condition. The problem for the spectator of this film is how to interpret these events which can have no causal connection and yet are presented *as if* they were causally connected so that it seems that pushing a button brings a dying man into a room and creates a car accident.¹⁴ In effect, we are being asked to accept a special fiction ("as if") within an already fictional mystery story.

In order to solve this causal problem the spectator must evaluate the temporal relationships posed by the sequence. Four important principles of causal reasoning are that a cause must precede an effect, an

effect cannot work backward in time to create a cause, certain patterns of *repetition* among events make a causal connection more likely (e.g., pushing a button three times . . .),¹⁵ and a prior event which is temporally or spatially more proximate to the outcome than others is more likely to be a cause of the outcome.¹⁶ Many different sorts of temporal situations, bearing on our judgments of causality, may be created through the juxtaposition of spatial fragments from different shots.¹⁷ As Arnheim emphasized, "the fact that two sequences follow each other on the screen does not indicate in itself that they should be understood as following each other in time."¹⁸ Thus before tackling the causal problem, we must briefly survey some of the possible temporal, and spatial, situations which may arise.

An extraordinary fact about the physical world is that virtually all phenomena can be explained in terms of interactions between parts *taken two at a time*. According to Marvin Minsky, "One could conceive of a universe in which whenever three stars formed an equilateral triangle, one of them would instantly disappear – but virtually no three-part interactions have ever been observed in the physical world."¹⁹ I will assume that explanations of phenomena are constructed on this basis; specifically, that the spectator constructs temporal, spatial, and causal situations by assembling parts two at a time. Thus in figure 3, temporal situation AB₁ in the story is created by imagining a particular relationship between the durations of two on-screen spaces, A and B, resulting in such story relationships as temporal continuity, ellipsis, overlap, simultaneity, reversal, or distortion.

More specifically, these temporal relationships in the story may be described as follows:

B₁ represents the time of A as *continuing*²⁰ into B such that the story order AB₁ is presented as identical to the screen order AB.

B₂ represents the time of A as continuing into B but with an initial *ellipsis* so that the screen order is interpreted as having omitted something from the story (which must be restored by the spectator's imagination); that is, the true order is: A, X, B₂, where X is not represented on the screen. If the ellipsis is large, but later disappears when completed by new screen events, then B₂ is a flashforward.

B₃ represents the time of A as continuing into B but only after an initial *overlap* in which there has been a partial replay of time already experienced in A.

B₄ represents a complete overlap with the time of A so that story event B is understood to be *simultaneous* with story event A even

STORY WORLD AND SCREEN

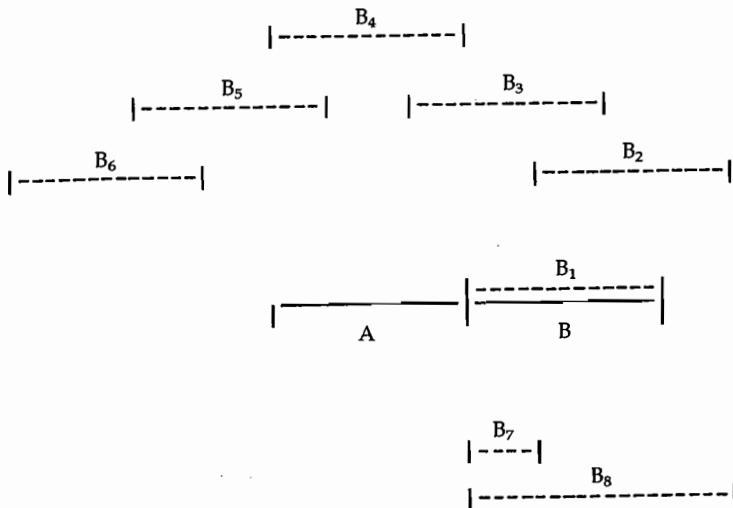


Figure 3 Story time

Graphic display of several varieties of story time created as a spectator relocates the on-screen time of spatial fragment B relative to the on-screen time of spatial fragment A resulting in a new and imaginary temporal order in the story, relationship AB_n . Some of the new relationships that may be created include temporal continuity, ellipsis, overlap, simultaneity, reversal, and distortion. The general principles illustrated here for time may also be applied to describe the ordering of space into such patterns as chains, gaps, reversals, and distortions (see text).

though B is seen to occur after A on the screen; that is, story time overrides the literal order on the screen.

B_5 represents an overlap with the time of A but with an initial brief jump back in time. This produces a fleeting but curious story time in which an effect (in A) has apparently been shown prior to its cause (in B). The spectator, in fact, is tempted to *mentally reverse* A and B_5 (creating a relation like A and B_3) in order to restore the forward arrow of time in which causes precede effects (i.e., prospective time). It is also possible that AB_5 may require the spectator to imagine an even earlier time (e.g., B_6) which is then taken as an explanation of A in the story – an implicit flashback – while B_5 continues to represent the “present time” of A. Using Noël Burch’s terms, I will refer to the AB_5 type of story order as a *retrospective* or *retroactive* story time.²¹ Our usual expectation is prospective time – A and so B. Less usual is retrospective time – A because of B.²²

Here is an example of retrospective time: Shot A shows an object

NARRATIVE COMPREHENSION AND FILM

from a certain position, but then shot B shows a person looking at the object from that previous position. In this way, we discover that the object we saw in A *had already been seen* by a character (and in fact, without knowing it at the time, we were seeing how that character saw the object). Thus with shot B we are forced to mentally readjust the order of events and reconceive shot A using the character as a new reference point, as a new condition for our seeing. We now conceive of the story event as composed of, first, a character who looks, followed by our view of what can be seen from that character's viewpoint. Part of shot B then comes to stand for either a literal, brief jump back in time, or else an approximation of what it would have looked like to have seen the character first looking. In any event, what is important is that the shots require the spectator to refigure the temporal scheme.²³

B₆ represents a time prior to A – a *past* time, or flashback, which requires the spectator to reorder story events and imagine other events which have been omitted and not seen on the screen between B₆ and A.

B₇ and B₈ represent temporal *distortions*. The on-screen duration of B (with respect to A!) is radically altered in such a way that it is not immediately clear what relationship with A is appropriate. For example, the duration of B may be compressed or expanded by running the film at a new speed, showing it backwards, repeating B, showing alternative takes, omitting frames by step printing, using freeze frames, and so forth.²⁴ In these situations A and B do not seem commensurate and hence we cannot immediately decide what story order is appropriate. Also included in these categories is indeterminate time. For example, in Jean-Luc Godard's *Weekend* (1967) there is a shot (involving hippie-guerrillas) that is so carefully arranged that its time cannot be ascertained. In fact, the shot is a flashforward but it cannot be recognized as such until much later in the film when the event being depicted actually occurs! The shot is thus a retrospective, nonsubjective flashforward.²⁵

The above scheme is a way of talking generally about principles of temporal ordering. However, it also applies to principles of *spatial ordering*. Although my discussion of the particular causal problem of the car crash in *The Lady from Shanghai* will center on time, it should be evident that, in general, space is just as relevant to the solution of causal problems. Therefore I wish to indicate briefly how the above scheme may be interpreted as an overview of some spatial principles of ordering. The scheme is not meant to be restrictive but merely to provide a

way of comparing various narrative theories each of which will use specialized terms to examine space and time in still finer detail.

In order to demonstrate how figure 3 may be applied to space, I will for the moment make an artificial, but simplifying assumption about space. I will assume that space comprises only two sectors: a foreground and a background. The question then becomes, how may we recognize a *change* in space? How may a given space "connect to" and be related to another space to form a new ordered whole? For convenience I will also assume – as in the case of time – that the change is effected through the editing of shots even though the scheme applies to changes effected in other ways (e.g., through camera or character movement, sound, changes in lighting level). The result of these assumptions is that space may evolve in only three basic ways: a new background may appear with the old foreground; a new foreground may appear with the old background; or, both a new foreground and a new background may appear. In the first two cases, what is new is introduced in conjunction with what has already been seen (an old foreground, or else an old background). This means that spaces are being connected into a *chain*. In the third case (a new foreground and a new background) the relationship of the new space to the spaces which have already been seen is *open* and not yet defined; that is, there is a "gap" of some type between new and old space. Such a gap is indicated in figure 3 by the gap between fragment A and fragment B₂. On the other hand, fragments B₁, B₃, and B₄ represent a new space which either adjoins, overlaps, or repeats an old space so as to compose a *chain* of spaces.

There is a special case of the open space (A–B₂) which must be mentioned. When the new foreground is simply the old background and the new background is the old foreground, there has been no real change: foreground and background have simply been interchanged across the two shots. Space has been reversed, or mirrored between the shots. Fragments B₅ and B₆ are meant to represent this general class of *reverse angles* in film. A typical example is shot/reverse-shot editing which depicts a conversation between two characters by alternating shots taken over each character's shoulder.²⁶

What can spatial "reversals" like B₅ and B₆ mean in terms of the story world being created by the spectator from a series of on-screen spaces? David Bordwell offers a proposal:

Shot/reverse-shot editing helps make narration covert by creating the sense that no important scenographic space remains unaccounted for. If shot two shows the important material outside shot one, there is no spatial point we can assign to the narration; the narration is always elsewhere, outside this shot but never visible in the next.²⁷

NARRATIVE COMPREHENSION AND FILM

In other words, when space is reversed we do not see a camera, sets, or technicians but only more diegetic space which seemingly is part of a consistent and unified group of spaces with no disturbing (causal) outside influences (e.g., by an "author"). The new and larger space being represented through a reversal is an imaginary space – a diegetic space of the characters that is seemingly like itself in every direction.

There are, of course, degrees and kinds of chains, gaps, and reversals of space; and our recognition of the kinds will depend on the nature of other conventions governing, say, camera placement (for example, whether spaces are oriented toward a 180 degree axis of action). Connecting screen spaces to a pattern of story space does not prohibit also using gaps (B_2 and B_6), or other distortions (cf. B_7 and B_8), to create a story space which is *not* the sum of spatial fragments on the screen.²⁸ Such a gap between screen and story space leads to degrees and kinds of "impossible" space; that is, *to space which can not be justified as existing wholly within the diegesis*. Impossible space leads to perceptual problems of a new kind that force the spectator to reconsider prior hypotheses about time and causality.

CAUSALITY AND METAPHOR

I will have more to say later about the narrative effects of spatial perception and the problem of "impossible" (nondiegetic) space, but for now I wish to return to the particular moment in *The Lady from Shanghai* in which it appears that a woman causes a car to crash into a truck by pressing a button. To work on this causal problem, the spectator must make a judgment about the temporal relationships of the three intercut actions. A number of factors point to some sort of continuity among the actions: no explicit motive is given for the woman's pressing of the button (hence we wonder what she is doing); she stares into space as if preoccupied by a thought (fear? determination?); it is clear from previous events that some sort of devious plan is being set in motion; the tempo of the editing increases to match the increasing speed of the diverse actions as if the actions form a group (a button is pressed, an indicator snaps on a display, a dying man quickly sits up, a car speeds along a road); and, most importantly, nondiegetic music which began with the first push of the button continues across all the spaces and rapidly builds to a loud and high-pitched climax coinciding, and merging, with the shrieking of automobile brakes and ending with the crash of glass (the final effect?). Furthermore, nothing about the representation of time in the scene rules out a causal link among the events: conjectured effects do not precede causes and the time of the causal interaction does not violate the presumed method of interaction. Pressing a button suggests the speed of light which is in accord with the

STORY WORLD AND SCREEN

"instantaneous" transitions suggested by the editing, but *not* in accord with transitions, conspicuously avoided in the sequence, based on character movement, camera movement, or optical effects, since these latter transitions would consume a screen time requiring additional explanation within the story.

It is worth emphasizing again that what I am describing is the spectator's interpretive process (here involving time) which works top-down on the data. The impetus for creating a *story* time does not derive in any simple way from the running time of the film – screen time – but rather from top-down processes seeking a new order which will be sensitive to other constraints on the data, e.g. presumed causality and event duration. Even our judgment of an event as temporally continuous is *not* based on the *necessity* of its being physically complete on the screen. Perceptual illusions and constancies demonstrate that we may easily see what is not present, or fail to see what is present.²⁹ Physical continuity on the screen depends on the use of equipment (camera, printer, projector) while the *perception* of physical continuity on the screen is a matter of *bottom-up* processing (i.e., the operation of sensory mechanisms). The perception of *narrative* continuity, however, is a very different matter. Indeed, it turns out that the best perception of narrative continuity is often obtained by certain violations of physical continuity where part of an action is left out or sometimes briefly repeated.³⁰ Moreover, even blatant "mismatches" between shots will be overlooked by a spectator.³¹ This demonstrates, again, that recognition of objects and actions in a setting through top-down processing takes precedence over involuntary, bottom-up processing, and certainly takes precedence over physical continuity on the screen.

There is no general set of necessary and sufficient conditions which determine how the results of bottom-up processing must be interpreted. The perception of continuous or discontinuous screen time may lead equally to judgments of either continuous or discontinuous story time. The fact of juxtaposition on the screen carries no necessary implication about temporal sequence or spatial relationship (cf. fig. 3), nor about causality (cf. temporal relationship AB_s in fig. 3). Even a repetition of the *same shot* need not signify that the *same time* is again being represented. The reason is that more complex events may be represented than what only occurs in front of the camera and/or occurs only once.³²

These ideas amount to saying that human comprehension does not proceed by progressively refining sensory data from lower to higher stages until a single thought is perfected and grasped by a singular Self. Rather, the human mind seems to be organized into modules that operate in parallel, are often too specialized to "communicate" with one another (or even to make use of "words"), and produce crisscrossing outputs which conflict as well as unify.³³ The Self in this view

is not stable and unified, but instead is "diverse, capable of being all those it will at one time be, a group acting together."³⁴ Hume reached a similar conclusion from a much different perspective. He argued in 1739 that "what we call a *mind*, is nothing but a heap or collection of different perceptions." These perceptions, thoughts, and motives "succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement." Hume could find no common feature, or continuing link, among the perceptions that could qualify as a unified Self; instead there are "the successive perceptions only."³⁵ In 1890 William James posited a "hierarchy" of selves to account for the felt presence in consciousness of an "ideal spectator," but he retained Hume's notions of multiplicity and conflict among the selves.³⁶ For Erving Goffman, "Self . . . is not an entity half-concealed behind events, but a changeable formula for managing oneself during them."³⁷ Marvin Minsky emphasizes that the Self also functions "to keep us from changing too rapidly."³⁸ Recently, similar conclusions about the "divided" self, or "split" subject, have been reached in certain film and literary theories inspired by post-structuralism, psychoanalysis, and ideological criticism.³⁹

I believe that the diverse strategies used by an individual for managing perceptual and behavioral problems in general include the creation of hierarchies and an "ideal" position of spectatorship, and that such strategies have counterparts in an individual's comprehension of narrative texts and in an individual's construction of an efficient self in relation to a text. I will return to these notions in later chapters where they will be reflected in some basic concepts, such as "levels of narration."

What is certain about comprehension is that the perceiver must search, compare, test, discriminate, remember, and speculate within many realms and imagined contexts. Evidence on the screen cannot "speak for itself," for how can we know what is *implied* by the evidence and what limits may exist on a given (top-down) method of seeing? Switching to a new method of interpreting – i.e., a different top-down schema – may reveal that the data under the old method was actually inconsistent, fragmentary, mistaken, deceptive, or ambiguous. Switching methods again may recast the evidence a third time. This is particularly true for fictions which, as we shall see in chapter 7, exploit a qualified indeterminateness of signification (or, according to some radical theories, exploit the fundamental indeterminacy of signification). The necessary incompleteness of data under alternative interpretation highlights a critical fact: there is a *persistence* to top-down processing. A spectator is willing to fill in some data or ignore other data in order to maintain a particular temporal and spatial context – a master "frame of reference or schema – for events and details. For example, we rou

tinely ignore the break in time signalled by a change of camera angle in favor of maintaining an interpretation of temporal continuity in a story (relationship A-B₁ above), even if, say, the editing of a match on to overlook other details. Indeed, using a word like "technical" to characterize the editing betrays the decision we've already made to preserve story time. The relative persistence and autonomy of certain top-down narrative processes in our comprehension will be a recurrent theme of this book.

Returning to our analysis of *The Lady from Shanghai*, it must be said that the spectator does not finally believe that the woman has directly caused the violent actions. The spectator's perception of the distances among the spaces, coupled with culturally-acquired knowledge of how cars operate and of how rich people summon servants at the push of a button, suggests that there is no immediate causal connection among the events. Moreover, this conclusion is reinforced by a knowledge of genre – the story is a mystery, not science fiction – and by prior events in the story which intimate that each of the men in the car, though in different ways and unknown to each other, are her accomplices in some scheme, not enemies who must be killed. Still, she does seem to be standing contemplative and motionless at the very center of several violent actions and to be linked somehow to death. (Notice that our interpretation might be quite different if she were seen running up a flight of stairs or chatting with a friend during these events. In that case the causal problem might never have arisen.) The usual form of the mystery story has been turned upside down: the mystery here is why we should see these three events as mysterious when apparently there can be no mystery about them because we know their causes (including who shot the man who is dying in a nearby room). Ultimately, this specially created, but then rejected, fiction where a woman apparently causes men to die by pushing a button, must itself be seen as embedded within a larger fiction – a mystery story. The truth of the embedded fiction must be sought by looking for a fit with the questions proposed by the plot and then, in a still larger sense, a fit with the pre-suppositions which may be held by a society about the concealed mystery of women, their danger to men, and their proper place.⁴⁰ That is, the woman character's desire is first analyzed through her different relationships with four men – what does she really want from them? – and then analyzed through the filter of what we are supposed to imagine generally that women must want. I do not wish to pursue this interpretation, but rather to extract from the analysis of this particular scene the following set of general proposals about narrative comprehension:

Interpretations of narrative depend in a crucial way upon the

judgments we make about space, time, and causation as we work top-down on screen data. Identifying an event as a "story event" is a matter of deciding where actions begin, how they break off, and which actions belong together. We must judge not only the temporal status of special cross-cut actions but the time implied by the juxtaposition of *any* two shots including ensuring that time has not stopped or otherwise shifted *within* a shot. (Note, again, that the physical material of film – such as the break between shots – does not guarantee *a priori* a specific temporal relation nor guarantee a change in temporal relations.) We use hypotheses about time to search for causation and, reciprocally, we use hypotheses about causality to establish temporal order. Identifying what counts as an event involves searching for an "equilibrium" among possible "values" for space, time, and causation. Although other sorts of knowledge are clearly relevant, judgments about space, time, and causation are fundamental because they give us the means with which to see and hear: a framework within which to perceive. Even when the pushing of a button is meant to cause a car to crash (say, on an automobile test track), it is necessary to imagine the event within appropriate conditions of space and time. As George Wilson says, our task as spectators is ultimately "to work out how our perceptual comprehension of the relevant film world is related to our normal modes of ordering and understanding perception in everyday visual experience."⁴¹

In order to connect our understanding of film to our understanding of the ordinary world, we have to be sensitive to the specific techniques available in film for representing space, time, and causal relations. In watching a film, for example, we must respond to camera movements, matches on action, perspective relations, attached and cast shadows, optical transitions, screen direction, sound overlap, off-screen sound, voice-over and voice-under,⁴² and an enormous number of other features of the medium. Although narrative may appear in any medium, the particular materials and techniques of a given medium partly determine when and how we apply our skills of spatial, temporal, and causal construction. In this sense, the break between shots does not usually "mean" anything; it is merely a catalyst for us to *proceed to generate hypotheses* using certain strategies. Many difficulties in narrative analysis and film theory have arisen through the failure to appreciate the difference between a "reference" – such as a denotation – and a "procedure" or instruction to be followed in discovering/assigning a reference.⁴³ The distinction between a reference that has been achieved and a procedure for referring is vital for any complete definition of narrative and will be addressed in chapters 3 and 4.

Thus, in summary, the spectator must find an interpretation of *The Lady from Shanghai* which assigns responsibility for the car crash to the woman but not on the basis of a button being pushed. The editing of

the film is not dismissed as simply false, misleading, or accidental, but instead may be viewed as creating a second fiction in which a realization of the truth of the apparent causal connection is merely *deferred* to a later time. The contradiction whereby the woman both causes and does not cause the car crash may be represented by the spectator's grouping of on-screen elements into two simultaneous worlds: the diegetic and the nondiegetic. Nondiegetic references are not taken to be part of the character's world, and hence not subject to its laws, but instead are taken to be *about* that world and are addressed only to the spectator. In this way the film allows the spectator to begin to see one thing (a car crash) in terms of another (a woman), but not in a literal (diegetic) way. Just as a statue is said to emerge "out of" clay, the dying man and the car crash are to emerge metaphorically from a presumed feminine state of being: "She killed males *out of* desperation."⁴⁴ The metaphor ("out of") functions to describe the nature of the causation. This sequence in the film is particularly complex because neither the woman nor the spectator fully understands as yet the nature of her responsibility or its consequences.

Selecting something to be seen in terms of something else – that is, substituting one thing for another to highlight a shared quality – demonstrates the close connection between creating metaphors, and discovering causes.⁴⁵ To discover a "cause" in this sense is to recognize that the logic of grouping certain "events" together on the basis of a shared quality may point to an underlying, formative process. Such a formative process may act to transform that quality, thus drawing together still more events.⁴⁶ In representing the car crash, *The Lady from Shanghai* is at the very threshold of stating a "likeness," or shared bond, sufficient to support the conclusion that a woman, or Woman, is a source of deadly threat to males. This may be expressed in another way by saying that although we must conclude that time is merely chronological in the car crash sequence, the ordering of the event nonetheless seems to betray a larger principle or pattern beginning to emerge in which a threat to males arises from the *difference* between male and female. The paradoxical causality whereby a woman both causes and does not cause a car crash reflects a general principle of *The Lady from Shanghai* that a *dissimilarity* between the sexes is the common link among a set of events. Indeed some writers have argued that narrative causality is essentially these larger principles of grouping (recall, for example, Todorov's notion of narrative reversals, or "transformations") which come to dominate a simple and irreversible kind of time like chronology. Roland Barthes asserts that

the mainspring of narrative is precisely the confusion of consecutiveness and consequence, what-comes-*after* being read in narrative

NARRATIVE COMPREHENSION AND FILM

as what-is-caused-by; in which case narrative would be a systematic application of the logical fallacy denounced by Scholasticism in the formula *post hoc, ergo propter hoc*.⁴⁷

The car crash sequence of *The Lady from Shanghai* reveals chronological time in the very process of being transformed into narrative time.

It is clear that we will need the distinctions introduced in the previous chapter that were used to isolate types of time and types of causality. Also, in understanding the causal structure of a story and its associated metaphors a spectator may need to create further subdivisions of the broad categories of diegetic and nondiegetic in order to hold together data of different sorts until it is needed to produce through juxtaposition more complex and subtle descriptions of the references (and "causation") put into play by a story. In chapter 4, I will describe several more of these subdivisions utilizing the concept of "levels of narration" in order to make more precise the notion of related, yet distinct story "worlds," or "levels of reality."

IMPOSSIBLE STORY SPACE

In fashioning diegetic and nondiegetic worlds we are constantly required to keep track of the ways in which our perceptions are related to the perceptions of characters within the story.⁴⁸ The nature of a car crash will change according to whether we believe a woman is dreaming about a car crash, remembering one, or unaware that one is occurring. In the first two cases, the status we accord the crash – its spatial, temporal, and causal implications – derives from explicit relationships with the character; in the third case, the event is determined "negatively" as an independent occurrence, as unrelated to character perception.⁴⁹

Our knowledge of the car crash is also affected if the woman is shown only to be *observing* it as a bystander. The reason is that our perception still coincides in some manner with that of the character; for example, we see the event at the same time as the woman. This fourth case falls between the first two and the third; the precise extent to which our perception coincides with the character has yet to be specified by the text. Nevertheless, the very existence of the character as she looks around her world and shares ideas with other characters testifies to our own perception of the diegetic world. Her act of perceiving seems to justify, and may even direct, the spectator's act of perceiving. Such a first-person account of space and time renders the "author" of the fiction invisible behind the character's experience. Like the apparent causation in *The Lady from Shanghai*, a first-person recounting of events is an illusion, but one which is bound up with the very conditions

STORY WORLD AND SCREEN

which allow us to make sense of the fictional world. The creation of a "character" has enormous implications because a character is understood to perceive in ways that we might imagine ourselves imitating if we were in a similar situation.

There are, of course, many ways to show characters perceiving, and consequently many relationships we may have to their perceptions. "First-person" and "third-person" modes of perceiving may be visualized as acting at ninety degrees to one another. Actual cases, however, will place our involvement with a character at intermediate angles. For example, an eyeline match differs from a point-of-view shot according to the inferences a spectator may draw about the object that is seen. The former shows us what a character sees and when the character sees it; the latter shows us, in addition, the perspective of the character. Thus the point-of-view shot represents on screen an additional "subjective" feature. It is more *restricted* than the eyeline match because its representation of space and distance is tied much closer to the presumed place of the character. These sorts of (semi-subjective) effect must be taken into account as we construct the space and time of a story world.

The development of space through the perceptions of characters, like the impossible causality of *The Lady from Shanghai*, may create problems for the spectator that lead toward nonliteral interpretations of events. Fritz Lang's *Dr Mabuse, the Gambler* (Part One, 1922) opens on a close-up of photographs of men spread out in someone's hand as if they were playing cards. New photographs are selected as if from a "deck." In the second shot we see the man who was holding the "cards," Dr Mabuse, now shuffling them, though we do not see his eyes. There are no other "players" seated around the table. The third shot repeats the close-up framing of the first shot. One photograph is selected. In the fourth shot we again see Dr Mabuse as he holds the "card" up over his right shoulder, and finally we see his eyes as he looks up and stares into the right foreground (shot 4, fig. 4). The next shot presents a spatial problem. We see a man standing in another part of the room looking left (shot 5, fig. 5). Where is he? If he is responding to Mabuse, he would appear to be located in the room near where the camera was positioned in the previous shot. However in the next shot Mabuse turns to look back over his right shoulder (shot 6, fig. 6). Intertitle: "You've been taking cocaine again." Mabuse continues to stare. Then we see the man again but he is now looking right (shot 9, fig. 7). We did not see him turn to face in the opposite direction. (Has he now turned away from Mabuse?) Intertitle: "If you fire me, I'll kill myself." We are surprised both by the dialogue and the space: If he is responding to Mabuse, he is not near the camera where we first thought him to be (i.e., near the camera position of shot 4, fig. 4), but in the background off left. The following shots make clear that our first belief was in fact

wrong. Initially neither man was looking at the other. The space seen in shots 5 and 9 (where the man is standing) must be relocated from the off-screen right foreground of shot 4 to the off-screen left background. It becomes clear that the man in shots 5 and 9 was avoiding Mabuse's gaze. Dominated by Mabuse, he responds instantly. Later in shot 12 he will enter obediently from behind Mabuse to take the photograph.

Shot 12 finally makes the organization of the space explicit. All of this strikes us as rather abrupt and confusing. If the man had been shown turning around, or Mabuse looking in the "proper" direction, we could have correctly understood the interior of the room. The space that we believed was created by the juxtaposition of shots 4 and 5 (figs 4 and 5) in the right foreground of shot 4 has been shown to be completely false. We did not in fact see it; we saw something else. At a thematic level, Mabuse is revealed to be at the very center of the event with the ability to create and destroy space through his powerful gaze (e.g., fig. 4). Moreover, we soon discover that Mabuse himself, as a master of disguises, will often be concealed. In this way we begin to see that appearances in the film may be misleading and real power concealed. We are perhaps reminded of the opening shot in which Mabuse – almost like the invisible maker of a fiction film – shuffles photographs of men as if he had an absolute power to control destiny or assume a new identity at will.

Lang has created these effects through editing and the use of angled glances. The space of the scene is developed through attention to various rules of continuity editing (e.g., the 180 degree axis of action, matches on movement, the 30 degree rule, etc.).⁵⁰ Lang has not used *mise-en-scène* or overlapping space to orient the spectator. We cannot, for example, decide from the background of shot 5 (fig. 5) where the man must be located with respect to Mabuse. Incredibly, not until shot 25 are we presented with an establishing shot of the room. Lang avoids techniques of the early cinema which would have presented the scene in a single shot as a distant tableau. Instead the spectator is initially placed "inside" the action. For this reason Noël Burch accords the film a special place in film history and praises its power to force the spectator to mentally create a continuous space and time out of a series of fragments and glances.⁵¹ The spectator is induced to overlook what is literally on the screen – compositions which cut up the room coupled with frequent changes of angle – in order to imagine a coherent story event.

But Lang has achieved more than a mastery of the conventions of so-called invisible or transparent editing. His use of two contradictory spaces (figs 4-5 as opposed to figs 6-7) demonstrates that character glance has the power to generate a discontinuous, or impossible, space as well as to generate the illusion of an integrated space. At a global level, too, many events in the story will prove initially baffling or

remain only partially comprehensible which spurs us to search for an underlying rationale, or better, an *underlying rationality*.⁵² In the same way that a certain photograph, or image on the screen, offered to us by Lang is diegetically unassimilable (shot 5, fig. 5) so also are other playing-card photographs, and characters, apparently shuffled and discarded in the story world. The measured confusion we feel between screen and story may drive us toward an understanding of the demonic forms of causality which increasingly are associated with Mabuse and which, like those of *The Lady from Shanghai*, hold us in their grip.

Why should a glance have so much power to generate a narrative? The reason is that unlike, say, a shot of a character wiping his forehead or a shot of a vase on a table, a glance bristles with implications about space, time, and causality. A glance leaps across space: its direction orients us to something nearby and hence enables us to build spatial relationships within a scene. A glance implies temporal relationships as well: an object seen is interpreted to exist in a time continuous, or simultaneous, with the act of seeing.⁵³ Also, a glance may be linked directly to a character's intention, or to a forthcoming act by the character, or to a reaction (when the character is acted upon). A glance implies an interaction with an object. In fact, glances are so important to narrating a story world that the only glance that is generally avoided is a glance into the lens of the camera.⁵⁴ A look into the camera breaks the diegesis because it makes the conventional reverse shot or eyeline match impossible. (Such a match would reveal the camera itself; its absence would be just as revealing.) In a deeper sense, a character's glance is an important measure of the *acquisition of knowledge* by character and spectator. As we shall see in chapters 3 and 4, knowing, and thus being able to tell, is a fundamental property of narration. Psychoanalytic and feminist theories go even further: they tie the glance to fundamental human drives and to scenarios of unconscious desire; to forms of visual pleasure, anxiety, and fantasy (e.g., voyeurism and fetishism); and to the very constitution of a self and a gender distinct from an Other.⁵⁵

The Lady from Shanghai illustrated that even impossible causation may affect our interpretation of diegetic events while *Dr Mabuse, the Gambler* illustrated that character glances may disrupt our perception of diegetic space. The impossible space created through Mabuse's false eyeline match (figs 4-5) has the effect of interrupting the flow of space around the spectator. In *Dr Mabuse* the frontality of early cinema is giving way to spatial articulations modeled on shot-reverse shot, creating more complex principles of spatial ordering as well as more complex forms of continuity and discontinuity. I want to turn now to screen space in order to consider how purely graphic patterns may also play a role in the spectator's experience of story space.

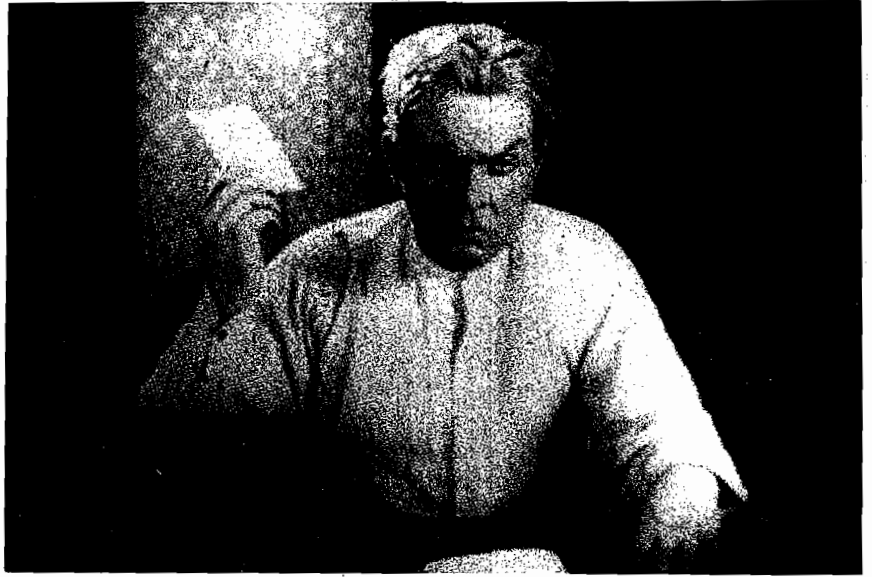


Figure 4 Dr Mabuse, the Gambler (shot 4)



Figure 5 Dr Mabuse, the Gambler (shot 5)

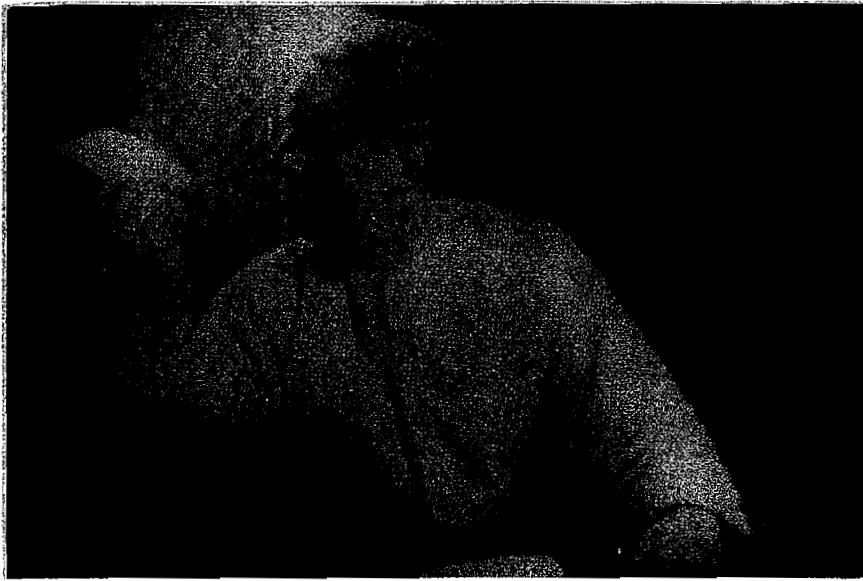


Figure 6 Dr Mabuse, the Gambler (shot 6)



Figure 7 Dr Mabuse, the Gambler (shot 9)

SCREEN SPACE AND STYLISTIC METAPHORS

In Alfred Hitchcock's *The 39 Steps* (1935) a Scottish farmer and his young wife agree to allow the hero, Hannay, to spend the night with them. The wife discovers that Hannay is fleeing from the police who falsely believe that he is a murderer. She decides to help him. Her jealous husband, however, becomes suspicious of the way both of them are behaving toward each other. During supper he says that he must go outside to lock the barn. Once outside, however, he walks around the house to the kitchen window where we watch from behind him as he spies on his wife through the window. Next we see a medium close-up of his face from inside the house framed by the window bars (fig. 8) followed by his point-of-view from outside looking back through the window, again framed by window bars (fig. 9). This shot discloses the wife and Hannay leaning toward each other in earnest conversation. The point-of-view position is reinforced by our inability to hear what is being said because we are outside with the farmer. Next we see the husband's face framed by the window bars from the same camera position as before (cf. fig. 8). The scene then ends with a fade-out. We realize that the husband has been deceived by appearances for he believes that his wife has a sexual interest in Hannay and that they are secretly plotting to be together.

We discovered in *Dr Mabuse, the Gambler* that a shot could represent a fragment of story space as itself a collection of points, one of which could become the point *from which* the camera reveals the *next* fragment of story space. In this way a master space may be constructed by connecting spatial fragments from point to point in a transitive series from *within* the space of the story: if A is to the left of B and B is to the left of C then A is also to the left of C. In *The 39 Steps* the camera has actually taken the place of a character at a point in space in order to show us how the character sees (figs 8-9). The bars on the window and our distance from them in these two shots play a crucial role in our recognition that *the camera angle has changed by 180 degrees* between the two shots and that the camera has, in fact, assumed a point in space which has already been seen. By thus *following the angle* we are able to reorient ourselves to the story space and develop it in new directions. In recognizing the point-of-view structure we have converted a place *on* a flat screen (marked by one horizontal and one vertical line) into a place *in* the story world (two bars on a window). A continuity of story space is preserved between the two shots. But this is not all. I want to show that it is not a matter of indifference where the bars of the window appear on the screen to help us in marking story space.

Let's consider four basic possibilities for positioning the window bars in the point-of-view shot (fig. 9; see figs 10-13). One possibility would

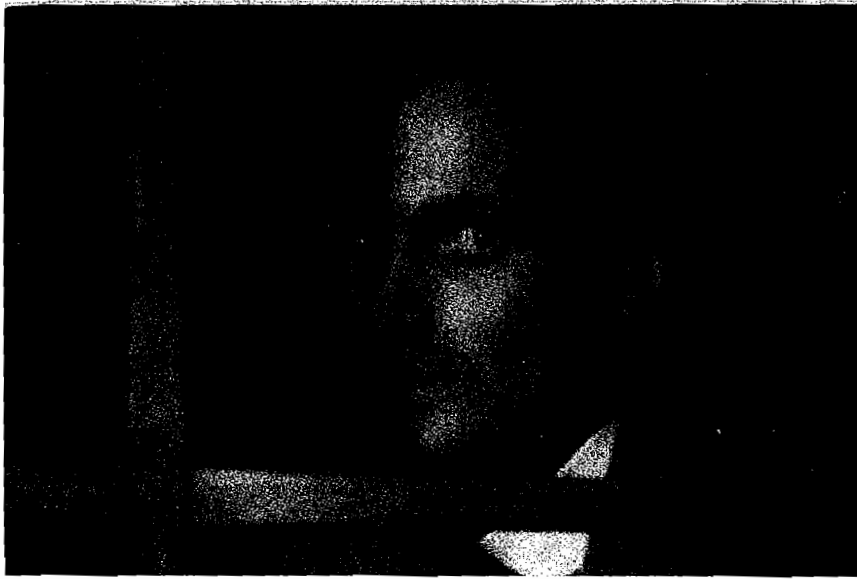


Figure 8 The 39 Steps (shot 1)

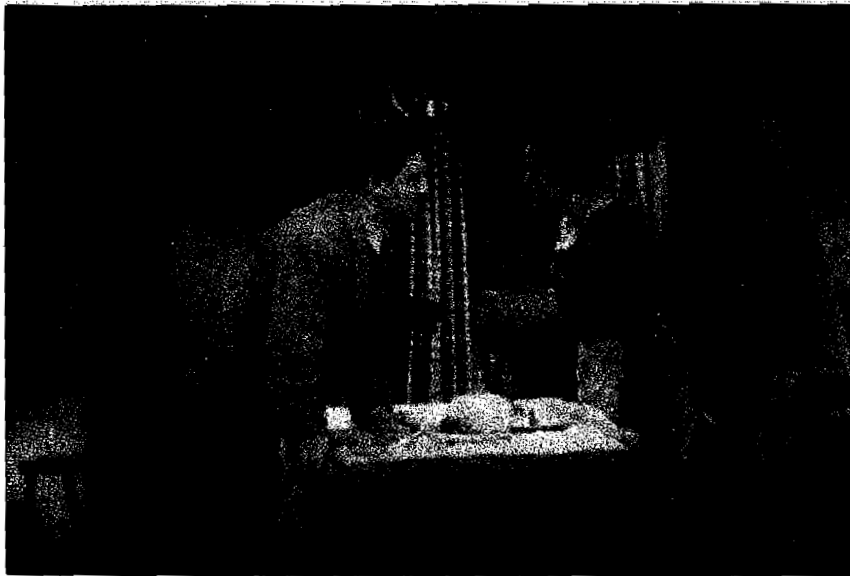


Figure 9 The 39 Steps (shot 2)

NARRATIVE COMPREHENSION AND FILM

be to show them exactly where they were in the previous shot of the husband's face looking in the window (fig. 8). Figure 10 shows this possibility using asterisks (* * *) to mark where the bars would be in the new shot. Through persistence of vision,⁵⁶ short-term memory, and the gestalt law of organization known as "proximity" (i.e., parts that are close together are grouped together and hence tend to be seen as one object, a whole) we would recognize an exact overlap with the previous shot. I will call this a "graphic match" between the two shots. It has an important disadvantage: it suggests that there has *not* been a 180 degree reversal of angle between the two shots. If the same window bars of the first shot are to be shown reversed 180 degrees in the second shot, then the vertical bar would switch from screen left to right while the horizontal bar would remain at the bottom of the screen. This *expectation* of what would result if the camera angle were reversed 180 degrees (i.e., if the story space were to be seen reversed) is shown in figures 10-13 as a solid line (—). The graphic match thus works *against* our perception of spatial continuity in the story world by suggesting that there has been *no* change of angle and hence no point-of-view from the spatial position of the husband. We can, of course, still see the second shot as a point-of-view shot but we must rely on other cues and we must "overturn" the evidence of the immobile window bars (by, for example, believing that after reversing the angle 180 degrees, Hitchcock shifted the framing to show *different* window bars).

A second possibility for using the window bars in the point-of-view shot would be to show them reversed exactly as we would expect them to be if the camera had reversed 180 degrees (fig. 11). In this case the graphic configuration reinforces our spatial hypothesis and supports the spatial continuity of a character's point of view. I will call this a "spatial match" between the two shots.

A third possibility would be to show the window bars as if they were reflected along a diagonal line upwards to the right (fig. 12). This bizarre sort of movement does not accord with the normal way in which characters move through and view their world. On the other hand, there has been a clear change from the view presented in the first shot which is not inconsistent with a 180 degree reversal of angle: the vertical bar has in fact switched to the right side. Perhaps the horizontal upper bar, then, is simply a different part of the window. This graphic configuration is merely inconclusive. It neither encourages nor rules out spatial continuity in the story world. I will call this an "open match" between the two shots since graphic space and story space are represented as simply "decoupled" from one another.

A final possibility would be to show the bars as if they were merely reflected upwards (fig. 13). This is the choice that Hitchcock actually makes in the film (fig. 9). It is a decision *not* to show the same window

STORY WORLD AND SCREEN

Figure 10 Graphic match

Graphic continuity opposes spatial continuity.

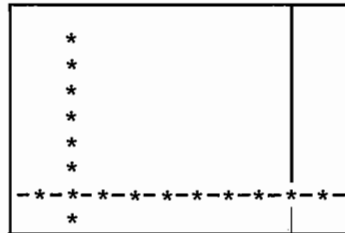


Figure 11 Spatial match

Graphics reinforce spatial continuity.

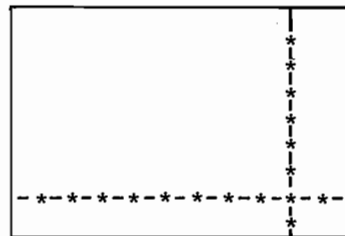


Figure 12 Open match

Graphics neither oppose nor reinforce spatial continuity.

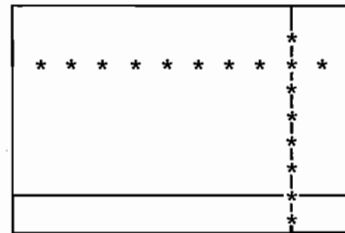
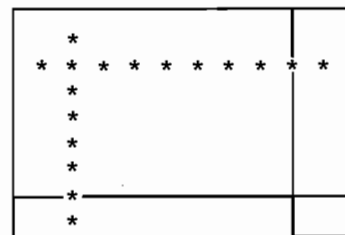


Figure 13 Integrated match

Hitchcock's choice: graphic pattern on the screen parallels a spatial pattern in the story. See figure 9.



Figures 10-13 Story and screen matching

Four possible framings of an object seen through a window in *The 39 Steps* representing the point-of-view of a character who is shown looking through the window in the previous shot (see figure 8). Asterisks (***) represent alternative locations for the position of the horizontal and vertical window bars. Solid lines (—) represent the position of the bars in the window as we would expect to see them if the camera angle were reversed 180 degrees from its position in the previous shot.

bars seen in the previous shot for at least one of the bars must be new. Hitchcock has given up a simple continuity (as in fig. 11) to obtain a most delicate effect. The new composition does not rule out spatial continuity since there has been a clear change in the graphic configuration between the two shots (the horizontal bar has shifted to the top). By keeping the vertical bar at the left, however, Hitchcock is able to *unify* a purely graphic pattern on the screen. It occurs in the following way: If the spectator perceives the old graphic configuration correctly reversed for the proper spatial continuity of the point-of-view shot *and* couples it with the new graphic configuration, the overall screen graphics of both shots form a closed inner frame within the frame of the shot. This unusual effect, I believe, results from exploiting two gestalt laws of organization: good continuation and closedness (closure). To put it another way, if one *anticipates* the 180 degree reversal of story space (by wondering what the husband will see if he looks in the window and how it will look to him), then one is rewarded by graphic continuation and closure *on the screen*. The graphics then attest to the pertinence of a question posed by the spectator of the narrative. The subtle power of this articulation is that screen space is allowed to close in on itself – apparently guaranteeing completeness in two dimensions – while story space reveals a maximum of information in three dimensions without breaking its essential continuity and completeness; that is, the two shots are linked through the shared space of the window pane while enabling us to see in two opposite directions in an instant.

In general, one can imagine other types of patterns besides an inner "rectangle" that may be fulfilled in a purely graphic way. Note especially that graphic patterns *may extend off-screen* (i.e., a spectator may imagine how a compositional pattern on screen may be continued off screen) and/or such patterns *may extend over several shots*. The important point is that the graphics of the screen complete a pattern in parallel with the completion of a coherent pattern in the story space. In such a situation, screen space neither directly opposes nor reinforces story space, nor is it neutral; instead, it complements story space. I will therefore call this an "integrated match" between the two shots. It is a moment in which our *top-down* processing of space – our expectation of what an invariant space in the story world should look like under a particular transformation – is smoothly integrated with our *bottom-up* perception of the new shapes on the screen.⁵⁷ The virtual space we anticipate in *The 39 Steps* is here amply confirmed whereas the virtual space of *Dr Mabuse, the Gambler* (figs 4–5) was overturned. Nevertheless, a spectator's conjectures about the space of *Dr Mabuse* are not wasted but rather, like the impossibilities of *The Lady from Shanghai*, are re-invested in a more powerful (nondiegetic) hypothesis *about* character and story world. The new hypothesis serves to organize and direct our

STORY WORLD AND SCREEN

search for entirely new sorts of data from the screen in succeeding shots.

The integrated match is also a moment when *stylistic metaphors* (i.e., metaphors joining style with story) become especially tempting for the spectator and critic. In the present case, for example, one might speculate that Hannay and the wife are "boxed in," *entrapped* by a husband's murderous jealousy; or that Hannay is being *doubly framed* for a murder he did not commit and a sexual liaison he does not intend; or that the camera's *double look* provides the spectator with greater knowledge than any of the characters but at the cost of making our perception congruent with the villain who places the man we sympathize with, Hannay, in great peril.⁵⁸ Moreover, Hitchcock later raises the stakes for there is a sense in which the false appearances turn out not to be false at all and the husband is right to be fearful (a reversal of what we had believed to be reversed): his wife is, in fact, trapped in a loveless marriage, he is inadequate as a husband, and Hannay is a "murderer" of women.⁵⁹ We will see in chapter 3 that these sorts of "problem" of perception, and especially the "reversals" of situation discussed in the previous chapter (e.g., Todorov's narrative "transformations"), are essential to any definition of narrative and narration. And we will see in chapters 4 and 5 that the point-of-view shot does not belong entirely to a character but rather is rendered logically possible for the character because of the simultaneous operation of more powerful, non-character narrations that are addressing the spectator by creating and manipulating "views."

One should not think that stylistic metaphors are merely a decorative use of language or a way of talking about what can be quickly verified by examining the screen. The story is not "objective" in this sense. These and other types of interpretive metaphor are being employed by a spectator in the process of comprehension. They provide a way of discovering and fashioning an appropriate story out of the material on the screen. They may be helpful or not helpful in focusing our thought; they may be consistent or inconsistent with other beliefs we hold about the story, or indeed with other beliefs we hold about the fundamental nature of film; but they cannot be right or wrong with respect to a story that is visible on the screen, for a story is created from the top down and is most certainly not visible.

Speaking broadly, then, we can say that screen and story may be related in four types of way. Graphics may exist simply (1) to reinforce the story, or (2) to complete a pattern analogous to a story pattern (as in figs 11 and 13, respectively). These are the ways commonly explored in the so-called classical Hollywood narrative which aims to create continuities between screen and story. By contrast, graphics may (3) be independent of story, or (4) actively oppose its coherence (figs 12 and 10, respectively). These are the ways associated with experimental

NARRATIVE COMPREHENSION AND FILM

cinema though they may also be systematically employed in certain narrative films.⁶⁰ Although the example from *The 39 Steps* emphasizes space, I believe that screen and story may be related in analogous types of way with respect to time and causality, and furthermore, that textures of sound may be integrated, or variously not integrated, with sounds heard in the story. It is thus only a first approximation to oppose "continuity" to "discontinuity," or even "spatial continuity" to "spatial discontinuity." What must also be kept in mind is the interplay between top-down and bottom-up processes as well as the effect of creating impossible space, impossible time, and/or impossible causality.

The four types of match reveal a basic tension in our comprehension of narrative in film. Our expectation that significant changes in camera angles will be correlated with changing events in the story world potentially conflicts with our actual perception of the two-dimensional plane of the screen which contains those light and dark shapes intended to represent the story world. In a similar way, temporal references in the story may conflict with our sense of the actual duration of the imagery appearing on the screen. Narrativity, or the narrative process, seeks to strike a balance between the demands of three-dimensional and two-dimensional space, between character time and spectator time. This suggests that narrative in general is a function which correlates imagined space-time with perceived space-time. Traces of such an activity may be found in the temporal anomalies and apparent causation of *The Lady from Shanghai*, the virtual space of *Dr Mabuse, the Gambler*, and the fitting together of story and screen in *The 39 Steps*. A narrative cannot avoid in some way telling the story of its own telling, just as the spectator cannot avoid retelling a story which exists less on the screen than in our predisposition to make sense, to apply what we already know from the top down.