FORMALISM AND INTERNAL EVIDENCE

MARY SIRRIDGE

Louisiana State University

IN THIS PAPER, I shall argue that it is incorrect to restrict the acceptable evidence for a critical interpretation of a work of fiction to elements "internal to the work." This is a restriction commonly associated with formalism of one variety or another,[1] but not all theories of criticism that have the restriction are, strictly speaking, formalist theories. The theories that I wish to consider are thus perhaps more correctly termed simply "internalist" theories, and I shall henceforth adopt this terminology. The most important manifestation of the internalist bias is the exclusion from criticism of factors surrounding the creation of the work, the so-called "intentional considerations."[2]

Interpretive criticism of fiction is criticism that undertakes to explore what might be called the "world depth" of the work of fiction and its relation to the actual world. The interpretive critic offers an explanation of the characters and their motivations and the interrelations of sequences of events vis-à-vis the kind of world situation presented in the work. It is quite possible that not all criticism of fiction is interpretive in this sense.

I will attempt to show that internalist theories of interpretation have a fundamental flaw that renders them unable to do the job for which they were designed, that is, to deliver a satisfactory interpretation or explanation of a work of fiction.[3] No matter how generously "internal evidence" is construed, the internalist who restricts himself solely to it finds himself faced with an unpleasant problem: either all explanatory criticism is illegitimate because it demands decisions that cannot be made on internal grounds alone; or any work has as many correct interpretations as it has consistent and complete explanations. On the first alternative, we lose most of literary criticism; on the second, we lose all sense of objectivity in literary interpretation. Noninternalist theories, I shall argue, are not faced with this particular difficulty.

I

Internalism is worth considering. This is not solely because there are still a number of critics and theorists who accept its tenets or who accept uncritically assumptions about critical methodology that are obviously warranted

Reason Papers No. 5 (Winter 1979) 27–39.

only if internalism is correct. Internalism has a certain independent intuitive plausibility. When a work of fiction leaves an author's hand, it seems to acquire a life of its own. We need not, it seems, know what Fielding intended or anything about 18th-century fiction to decide what happens in *Tom Jones* or whether the novel is hilarious. In addition, we want to be able to deal critically with works where our information about the author and the work's origin is radically deficient. Moreover, most critics are not prepared to defer to the author, even if the information is available. We may have good reason to suspect that the author is philosophically or psychologically obtuse, even when, qua artist, he has an unerring sense of detail. Why not adopt internalism then?

True, it is by no means clear initially what does and what does not count as "restricting oneself to the internal features of a work of fiction." Still, it seems this clarification need not be a major problem. Surely we want to admit that there is what we call "explicit information." For every declarative sentence of the work, then, let us admit that there is what we shall call a "corresponding sentence" of criticism true of the work. In addition, we are entitled, it seems, to the deductive consequences of this first set of critical sentences. Let us then construe "the internal elements of a work of fiction" as follows: (i) the "corresponding sentences" of all the fact-establishing sentences of the work,[4] and (ii) the logical entailments of the corresponding sentences. But the enlightened internalist will surely not stop here. He will want, in addition, those critical descriptions obtained by analyzing the meanings of the expressions of the work. This demands a competent grasp of the syntax of the language of the work.[5] We must concede to the internalist any inference based on meaning-preserving syntactical transformations. And we may as well concede inferences based on quasi-logical expressions such as modal and epistemic constructions. To forestall difficulties, let us also concede to the internalist whatever connections of expressions are pronounced analytic. We can also grant that our internalist is very sensitive to unusual syntax and to subtle meaning nuances. Naturally, he is also sensitive to unusual combinations of expressions that are structurally complex and significantly recurrent. But when the internalist has taken all these factors into account, he has exhausted his theoretically available evidence. He has dealt with the literary object qua object. In theory, he can go no further. He is not allowed recourse to what the author intended, what his other works were like, the spirit of the age in which the work was written. If the internalist adheres to this program, then, let us say that he is "attending exclusively to the internal features of the work." If internalism has difficulties on this generous formulation, then more restricted versions will have greater difficulties. I shall refer to the position outlined here as that of the "hypothetical internalist." If this position *per se* has problems, then all theories that rest upon it have a common flaw, regardless of their individual differences.

II

What sorts of critical activity are liable to involve the hypothetical internalist in difficulties? The hypothetical formalist does not intend to end up with a fragmented array of linguistic insights. He fully intends to obtain a synoptic view of the work as a work of fiction. As a bare minimum, then, he ought to be able to tell us what the fictional world is like, what laws govern it, why the characters do what they do.[6] This kind of critical activity rests on an analysis of plot situations, posing hypotheses about the characters' probable psychology, delineating temporal sequences of events, locating and discriminating between possible and certain causal connections between events, inferring information that seems to be left out, etc. For want of a more glamorous term, I shall call this sort of criticism "explanatory criticism." The critic is in an obvious sense explaining what is going on. The critic who fails at this job fails pretty dismally.

Explanatory criticism is the most basic kind of interpretive criticism. Criticism of other kinds is dependent upon it. For example, in the cases of stream- and center-of-consciousness works or first-person narration, we need to know what the fictional world is like in order to evaluate the narrator's competence. Among the most important kinds of criticism dependent upon explanatory criticism is criticism in which the critic claims to show what bearing a work might or should have on our knowledge of our own actual world. Obviously, we have to know what the fictional world is like, whether it is similar to our own or not, what goes on in it, in order to support claims of this kind. Since explanatory criticism seems to be so central, let us first turn our attention to it and see if our "hypothetical formalist" can give a satisfactory account of conclusions about the way the fictional world is, character analyses, and explanations of situations and events. All these critical activities fall under the heading of "explanatory criticism."

In the description of the actual world, such explanations presuppose other, methodologically more basic principles having to do with the law structure or constitution of the universe in question. (What *sort* of principles are presupposed may vary with the kind of conclusions we are interested in.) We are in general more or less aware that many of the higher-level descriptive claims we make about the actual world are "theory-laden" or "theory-dependent" in this way. Quite simply, the terms we use have meaning, and the claims we make are intelligible and correct, given the tenability of certain basic theories about the way the world is.

What I wish to emphasize is that our situation with respect to a fictional world is similar to our real-world situation in an important way. To achieve in our descriptions of fictional worlds a level of sophistication comparable to our ordinary survival level in the actual world, we must rely on principles bearing on the lawlike constitution of a fictional world, which tell us, for

example, whether it is a "magical" or a "naturalistic" universe. The chosen principles and the explanations they generate constitute the explanatory interpretation of a work. Different sets of principles or laws will give rise to different explanations and hence to different interpretations. In interpreting a work of fiction, just as in dealing with the actual world, the set of principles used to ground and generate explanations must be justifiable, should the interpreter be challenged. It is quite true that most explanatory conclusions are, as assumed in a recent article, based on assignments of probability.[7] But the probabilities assigned must be relevant to some set of laws which are themselves demonstrably right for the work.

Given that he must adopt some such set of principles in dealing with a work of fiction, can our hypothetical internalist give a satisfactory justification of his choice of principles? His position is that, in describing and explaining the work, he restricts himself to its explicit elements and to inferences based on the meaning of the expressions used in the work. We have already noted the appeal of his position. I shall now argue that it is nonetheless fatally inadequate, that the formalist position, formulated as I propose, cannot handle the production and defense of explanatory criticism. From the meager resources at his disposal, the internalist simply cannot defend the adoption of a specific set of principles as *the correct set* for a given work.

The crux of the matter is that there is a marked difference between fictional worlds and the actual world. In the actual world, we suppose ourselves to be at least theoretically in a position to justify our basic lawlike principles in some objective way, by an appeal to our own metaphysical predilections, by an appeal to the appropriate science, or by reflection on our past experience of the actual world. In the case of a fictional world, these particular avenues are never open to us. Once the work has been read carefully, the fictional facts are all in. We cannot very well appeal directly to science about the *actual* world or to our experience of or views about the actual world and other fictional worlds to justify our decisions about the lawlike constitution of a given fictional world. To be sure, we must adopt *some* principles even to be able to read intelligently. We must assume either that certain nonlogical laws or connections between predicates that hold in the actual world also hold in the fictional universe or that they are replaced by laws that do not hold in the actual universe. The principles chosen may change in the course of reading. The question is how, in the end, the critic defends the chosen set of principles as the correct set of principles.

There is one very general objection to the problem of justifying an interpretation as I have set it up. It is just as well to clear it out of the way at the start. It might be objected that the meanings of predicates are a result of the theories within which they occur and hence that even reading a work of fiction, as we obviously can, presupposes knowledge of the explanatory theories that apply to it. Such an objection might be held to show that in a

subtle sense the formalist is correct, that is, that in knowing the meanings of the expressions, we do implicitly know the law structure of fictional worlds. Thus, it might be argued, there can be no question of measuring the sets of principles against the "data" presented by the work, since there *are* no hard data, independent of meaning-defining principles. This objection is serious enough to merit an answer.

It is quite true that a complete interpretation of the predicates and their interrelationships presupposes a set of explanatory hypotheses adopted as principles that govern the fictional world. But for a given work, there will be several (at least) alternative choices of meanings and corresponding principles. Being able to read the work in the first place means that one has *a reading* of the work; it does not preclude the existence of *competing readings*. The critic's job is to choose between these. What is at issue is not so much "how to read a book" as "how to show that one has read it correctly."

Furthermore, there *are* "hard data" in the following sense. There are a certain number of uses of expressions that have to be interpreted consistently and cogently. We may, for example, discover that we can cogently explain all the tense expressions in their contexts in a work only if we assume a nonstandard tense logic; that is, we may discover that we get a consistent world discription only on the assumption of a nonstandard tense logic. No doubt in adopting these principles, making them the basis of our explanation, we diverge from the normal meaning of some tense expressions.[8] Nonetheless, whether the chosen principles allow us to interpret the expressions concerned so as to yield a consistent world description remains an objective question.

Finally, novels present some problems in explanatory criticism about which it is not even plausible to claim that they affect the meanings of the predicates. Suppose we have already decided that the universe in question is causally normal and are now faced with the problem of whether to construe the narrator's behavior in accordance with the laws of Freudian psychology. This decision may be quite important interpretively. On a Freudian interpretation, the possibility that the narrator is systematically misrepresenting the members of his family is substantially increased. Yet it does not seem that the choice affects the meaning of the predicates involved. When the narrator says that his father is angry with him, we may well suspect that his vision is warped; but the meaning of *angry* remains the same.

Thus, the attempt to vindicate formalism on the grounds of the theory dependence of predicates fails, and the formalist is left with the question how to show that a given interpretation of a work is the correct one.

There is no point in making the naïve claim that we simply begin reading, then learn from the facts of the fictional world that certain laws hold in it. We cannot presume that we are able to learn from a fictional world in just the way we do from a description of the actual world. For many of our common

procedures for dealing cognitively with the actual world—for example, simple induction—depend on prior assumptions about the constitution of the world to which we are not entitled in the case of fiction. We cannot merely assume causal regularity, for example. A unique choice of interpretation cannot be defended on the grounds that it allows us to explain all the "data." An ingenious critic can usually think up a dozen or so on the spot; and there is good reason to suppose that he could preceed to infinity, if time and patience were unlimited. If explanatory adequacy where the sole consideration, then any of these explanations should give us that same comfortable feeling of "fittingness." Ordinarily, most of them do not. Explanatory adequacy imposes on interpretations a minimal condition, one that even the internalist can rely on; but it seems not to be the only consideration.

The internalist may attempt to argue that we have underestimated the extent of the internal evidence. We know that certain laws hold in the fictional universe, he may say; and we are automatically entitled to claim that others hold also. Thus, our assumptions about the law structure of the work are therefore based on its internal features and what follows from those features. But this move is not legitimate. For the relationships between laws are in general nondeductive, and they do not seem to be analytic either. Connections between laws depend on the nature of the world to which they apply. Hence, we would have to argue that the fictional world was enough like the actual world that normal connections between laws still obtained. And it is the nature of the assumptions we are allowed to make that we are arguing about.

But the internalist has presumably not yet finished having his say. He may attempt to claim that he is entitled to the principles he chooses because such principles are true by virtue of the meanings of the expressions that constitute them. But is this so—even given the extensive concessions we have made about meanings? Among the inferences the internalist will want to carry through, presumably, will be some like that from:

(1) *S* knows that *p*

to

(2) *S* does not believe that *not-p*;

or, for example, from what Roger Chillingworth says to Hester in the prison scene to:

(3) Roger Chillingworth wanted revenge on Hester's partner in adultery.

Even the inference from (1) to (2) is open to debate. Very few people, I think, would be inclined to grant it on the basis of the meanings of the expressions in question. The inference to (3) certainly requires the mediation of nonanalytic psychological laws. And at any rate, our critic will more likely want more colorful inferences. Perhaps he will, for example, want to go from a description of Roger Chillingworth's behavior to the conclusion that Roger Chillingworth had heretofore lacked a goal structure and therefore jumped at the chance to get one, even one based on aversion, rather than

on positive desire. Can such inferences be legitimated on the basis of the given data, syntax, and analytic connections? I think not. In comparable claims about the actual world, we would base our inferences of this kind on deeply embedded generalizations about the actual world, which were originally learned from experience but which subsequently themselves play an important role in our further acquisition of knowledge and thereby form a foundation of knowledge. In this case, we would doubtless rely on psychological generalizations. In drawing conclusions, we would use some rules of nondeductive inference. But all of these generalizations and procedures depend in turn on more basic principles bearing on the way the world is—for example, that it is causally regular, that intentions and personality traits are evinced by what people say and do. To ground comparable claims about the characters of *The Scarlet Letter*, the formalist must adopt parallel principles and use methods of reasoning analogous to those assumed legitimate with respect to the fictional world. Unfortunately for the "hypothetical internalist," the normal choice of principles is not always warranted, as is readily evident from science fiction and fantasy works.

A concrete example, I think, will make the exact nature of the hypothetical internalist's predicament clearer. Let us suppose that a Balzac novel assigns two different dates to the same event. If something of this sort occurred in a history book, we would know immediately that at least one of the dates is incorrect. On the other hand, although we should be very surprised if a Balzac fictional universe were not like this one in its space-time structure, the possibility is not automatically excluded. We *could* claim in such a case that Balzac was, contrary to popular opinion, writing science fiction and employing a nonstandard tense logic. The example is farfetched, but it is not clear how the internalist could argue against this hypothesis if the interpretation is explanatorily adequate. Of course, it is *simpler* to explain that Balzac must have lost track of his earlier report or overlooked a misprint and that the universe is normal after all. But it is by no means clear that the simpler theory is automatically to be preferred. Both theories do account for the facts, as, no doubt, do numerous other theories. In this case, the "simpler" solution involves reducing some "data" to non-data status, always a drastic move. And there are surely many cases in criticism in which we would *not* opt for the simpler solution. In a science fiction work with a first-person narrator, for example, it is usually simpler to assume that the narrator has taken an overdose and dreamed the whole thing than to allow that we are being presented with a non-normal universe. It is also patently incorrect. Similar considerations affect other standard criteria of theory preference.

The strict internalist will probably at this point retreat to the position that in a problem case like the one hypothesized, he need not claim that there is some particular set of principles that lead to a *unique* correct interpretation. If, as in the present case, we are faced with two explanations that both

explain the data, then we simply have to admit that the "problem work" is ambiguous. The critic's job in such a case is simply to delineate the explanatory alternatives that would account for the facts. But this move merely dramatizes a fatal weakness in his position. It turns out that every work is in principle a "problem work."

Consider *The Scarlet Letter*. Most of us would accept it as a work that poses none of the sort of interpretive problems we are worried about. Yet in the opening scene, we are told that Roger Chillingworth sees Hester on the scaffold *and* that an expression of horror crosses his face. The normal reader assumes—though he is not told—that Chillingworth's face contorts in horror *for some reason*. People generally do not react in this way unprovoked. The reader further assumes that it is what the character sees that causes his reaction. People do sometimes recoil in horror from things that they see, and we have been given no other relevant information. But in connecting the events in this way, we have in fact assumed that the causal and psychological laws of *The Scarlet Letter* are very much like those which govern this world. This choice of principles seems natural. But even in this case, no doubt a sufficiently ingenious critic could construct alternative sets of principles that would force us to account for the explicit facts in quite different ways. Thus, even in a work that would normally be called nonproblematic, the problem of the bizarre—but adequate—interpretation occurs. In effect, every work becomes a problem work, and *ambiguous* becomes worthless as an aesthetic predicate. What is surprising is that there is as much consensus as there obviously is about the interpretation of works of literature.

The internalist thus finds himself faced with a difficult choice. He can claim that all explanatory criticism (and all criticism dependent upon it) is illegitimate. But once the presuppositions involved in our normal reading procedures are made explicit, it turns out that on this alternative we lose most of criticism. Or the internalist can allow explanatory criticism, thereby granting that all works are interpretively ambiguous. It becomes impossible to give any theoretically significant explanation of critical consensus or any defense of a particular interpretation against competing interpretations that are bizarre but adequate.

The real source of the problem is the internalist's initial restriction of data relevant to determining the interpretation of a work to "internal evidence." It is natural to assume that he means by this that any interpretation that adequately explains his chosen data is an acceptable interpretation. Recognizing the difficulty, he may attempt to add to the criterion of explanatory adequacy without sacrificing the internalist restriction of relevant data.

One popular special criterion for interpretive theories in aesthetics is the claim that the correct set of explanatory principles is the one that "makes the work come off best." One obvious problem with the suggestion is that, for any work whatsoever, it is possible to dig up an interpretation that is explanatorily adequate and that makes the work appear intriguing, complex,

etc. As a result, the number of negative evaluations we are entitled to make decreases radically. We might, for example, have to rate a given romantic work very high (although our initial impulse is to damn it for unbearable mawkishness) because it is exquisite when read as a parody. Another problem is that there is massive and energetic disagreement about what makes any work "come off best."

The internalist critic can, of course, appeal to critical intuition. All the data we are entitled to use are internal, he may claim; but mere explanatory adequacy is not the sole criterion for an acceptable interpretation. The correct interpretation, he may claim, just does emerge from the internal elements, gradually dawning and constituting itself in full clarity and embracing every minute detail. I surely do not wish to deny that the phenomenon that the internalist describes does indeed occur. But as a justification for an interpretation, this kind of appeal to "critical intuition" is a desperate move.

It is a good deal more to the point to note that much of our evidence about the correct interpretation of a work come from factors external to the work and that such evidence plays a legitimate and important role in interpretive criticism. The "intuition" of the competent critic has its feet rather firmly planted in historical good sense and extensive background knowledge. And the factors that influence the interpretive decision should be given their due in the account of justification. This means abandoning the internalist restriction of relevant critical evidence to "internal evidence."

III

The noninternalist critic is in a much better position to deal with explanatory criticism than is his internalist opponent, both with regard to critical decision making and with regard to explaining critical consensus. He can, for example, argue that nearly all of Balzac's novels are plausibly construed as causally normal relative to the actual world, as are nearly all novels written before the 20th century. As in other cases of human actions, we come to expect a given kind of work from a given author or in a given period.[9] In Balzac's case, the novel is one of a series, *The Human Comedy*, that exhibits an evolution of skill and sharpening of focus if construed as causally normal. In this case, as it so happens, we have the artist's stated intentions—Balzac's opinions in his prefaces and letters. Surely most of our preferred and unquestioned literary interpretations can be traced to a semiautomatic appeal to factors external to the work which delimit the interpretive options and cause some to be preferred to others.[10]

Although we cannot, as we have seen, use our knowledge of the actual world directly in arguments about the fictional worlds, we can use it indirectly in arguments about literary works as products of real-world actions. Explanation of actions and intentions is tricky business.[11] Agents

often misdescribe their actions, and the ingenious psychologist can nearly always present competing explanations of actions. Artistic creativity introduces "maverick factors" in addition to the problems normally encountered in explaining actions, intentions, and action products. The questions are nonetheless this-worldly. And in this-worldly affairs, we have certain advantages. We have preferred explanatory models and some sense of how to defend them. And we have a very strong vested interest in separating useful theory from idle speculation.

The externalist approach is not infallible. Arguments based on the author's stated intentions may mislead. An author may give an interpretation of his work that is not a good explanation of the facts presented in the work or that is incompatible with a set of basic principles and resultant explanations that did deal more adequately with the presented facts. In such a case, it would be quite appropriate to say that the author had misdescribed his own (quite complex) speech act. But in the present case, none of Balzac's individual novels gives evidence that he has misdescribed his action in telling us, as he does now and then, that he is constructing a normal universe. And there is no independent biographical evidence that shows that he cannot be trusted. We can be misled by broader genetic arguments also. We might, for example, conclude from the date of a work that it was a romantic work and proceed to interpret it as if it had the special symbolic vocabulary of romantic works, then find the work recalcitrant. Parallel mistakes occur in writing history. But in the absence of a counterargument from features internal to the work, an argument based on the author's stated intentions or upon the circumstances surrounding the creation of a work is admissible as an argument about the nature of the world of the work, hence about the correct interpretation of the work. We are at a definite advantage when we are arguing about our own world.

In the case just discussed, intentional and genetic considerations are used to show that a work does have normal law structure. But there are times when it is important to argue that we are *not* justified in relying on laws or connections of laws that hold in the actual world to map out the nomological geography of the fictional world. In Kafka's works, for example, or in science fiction, non-normal universes are quite common. In a few such cases, we have internal evidence that some of the generalizations we normally rely on do not hold. In some cases, assumption of normal causal laws gives an interpretation that is not a consistent world description. But external considerations are usually important. In a science fiction novel with a personal narrator, we could, of course, always argue that the narrator has taken an overdose and hallucinated the subsequent story. The contents of hallucinations need not be consistent. Such interpretations become extremely implausible, however, in the face of appeals to the author's stated intentions or to his other works or to works by authors writing in the same tradition or the same artistic circle, which are also plausibly explained by the

assumption of abnormal law structures. We can also have recourse to the fact that the author was considering some problem in his diaries or nonfictional works and proposed to write a fictional work along the same lines. If such considerations support an explanation that is already conceded to be explanatorily adequate, then that interpretation is probably the best.

It will no doubt be objected at this point that the theory I propose is excessively intentionalistic and headed toward all the traditional pitfalls. And certainly it *is* what I should prefer to call "externalistic." It is quite true that on the theory I propose, some genetic arguments and some intentional ones turn out to be good ones. One answer to the internalist objections is that the alternative to intentionalism is universal interpretive ambiguity, not with respect to the specifically aesthetic properties of works or our evaluations of them, but with regard to the "facts" of the work. If I am right, we cannot conclusively decide on the interpretation of a work on "internal grounds," although internal considerations may rule out some interpretations as adequate explanations. Furthermore, it is not clear that the theory I propose is vulnerable to the stock objections to intentionalistic theories. Very likely the author knows better than the rest of us what his intentions were and how he wants the work to be taken—although this is open to debate. But his word is surely not the only evidence we have for choosing a way of interpreting the work. The author does not, on my theory, become the final arbiter of the nature of his fictional speech act, any more than any user of language is the ultimate arbiter of the nature of the speech act he performs or the effect it has. The artist may ascribe to his work aesthetic qualities that it lacks, or he may advance an interpretation that is not adequate. Thus the simplest and most traditional arguments against reliance on intentions simply do not count against the theory I am advancing.

A parallel with graphic art will perhaps help here. Our claim that Monet painted water lilies is due, I suspect, largely to his claims that that is what his pictures were of. Many of our higher-level interpretive claims are based on the assumption that the paintings are of water lilies. If we had no information about Monet, we could base our interpretive arguments about his work upon other works similar to his in their internal features, works about which we had further information. We might at least narrow down the range of possible picture subjects in this way. If we had no historical knowledge at all about the impressionists but did have a large body of their works and, in addition, some photographs of the scenes they represented (paper-clipped to the backs of one or two of the pictures, let us say), then we could begin to theorize about the "facts" of the works *via* our conclusions about the representative conventions involved. If we had no information at all, however, we would not be able to decide whether a group of works was representational, let alone *what* was represented by the individual works, if anything.[12] This conclusion, far from being undesirable, is clearly what common sense dictates. The parallel for works of fiction, which are rep-

resentational in a slightly different sense, is obvious. Here, again, we are dependent upon external considerations, but I cannot see that we are relying on the author's intentions in any objectionable sense. The theory I am proposing is externalistic. And I think I have shown that externalism is a feature of any acceptable theory of interpretation.

Of course, the formalist is right in claiming that internal evidence is primary, that the data given in the work provide the initial testing ground for any explanatory theory about the work and for descriptions that rest on such explanatory theories. What I have tried to point out is that categorically restricting criticism to the primary or "internal" data cripples interpretive criticism in a rather unexpected and entirely unacceptable way.

1. No doubt some theorists would like to distinguish formalism altogether from internalism, which, they might claim, is a mere consequence of formalism (though perhaps a characteristic and unavoidable one). For the purposes of this paper, I will ignore this distinction to some extent. This procedure has the effect of forcing us to bypass some of the more interesting features of individual formalistic theories, and the formalist may well regard this as unjust. Nonetheless, the procedure is legitimate. For internalism is a central, but mistaken, feature of all formalist theories. Many of the other techniques and tenets of formalist theories seem unobjectionable, or at least open to debate. But if the internalist position is mistaken, they are left without theoretical justification.

2. A distinction can be drawn between considerations directly relevant to what the artist intended and more general considerations having to do with the circumstances surrounding the creation of the work—e.g., the age in which the author lived, his circle of friends, or his aesthetic theories. These latter factors we might call "genetic," rather than "intentionalistic." This is a distinction that has traditionally been ignored in practice; see M. Beardsley and W. K. Wimsatt, "The Intentionalist Fallacy," in *The Verbal Icon* (Kentucky, 1954), pp. 3–18. There are theorists who still defend the Beardsley-Wimsatt position unconditionally. A. J. Ellis, "Intention and Interpretation in Literature," *British Journal of Aesthetics* 14 (1974), does so; and Stein Haughom Olsen, "Authorial Intention," *British Journal of Aesthetics* 14 (1974) seemed of the same mind. His more recent position, spelled out in "Interpretation and Intention," *British Journal of Aesthetics* 17 (1977), is somewhat more moderate. Though he still seems to think that the work itself is the only evidence commonly accepted for the intent of a work, he does speak of a practice-defined matrix of intentions similar to those important for assessing moves in chess; his position is thus similar to that of Mark Roskill, "On the Intention and Meaning of Works of Art," *British Journal of Aesthetics* 17 (1977), who speaks of a notion of "the intent of a work," which is not to be identified with the intentions of the author as independently determinable. Berel Lang, "The Intentional Fallacy Revisited," *British Journal of Aesthetics* 10 (1967), and George Yoos, "The Work of Art as a Standard in Itself," *Journal of Aesthetics and Art Criticism* 26 (1967–68) propose "compatibilist" solutions that covertly reintroduce anti-intentionalism.

3. Throughout this paper, I shall consider a critical interpretation satisfactory only if there are good arguments to support the claim that this is the correct or best interpretation (or at least that no other interpretation is better). An interpretation is satisfactory, then, only if it is

defensible. Thus, internalism as I construe it is both a theory about how to choose an interpretation of a work of fiction and a theory about what kind of evidence is admissible in arguments for a given interpretation.

4. This account is somewhat oversimplified. Obviously, in a basic description, if p is a sentence of dialogue, we precede the sentence corresponding to the quoted material with an indication of who said it; and if p falls within a center- or screen-of-consciousness passage, we precede the sentence corresponding to the sentence of the work with an indication of who thought it. In addition, we make the appropriate indexical adjustments.

5. We have to assume, for example, that our formalist reader knows the difference between *the*-constructions in "*The gladiator* kicked *the bucket*" and "*The whale* is becoming extinct" and that he can recognize meaning-preserving transformations.

6. Throughout this paper I will be talking as if the work of fiction were a world description (more properly speaking, a world presentation) and thus subject to many of the same restrictions as a complete state description of the actual world. This way of talking is intelligible and relatively clear and does not prejudice the issue at hand.

7. Philip Devine, "The Logic of Fiction," *Philosophical Studies* 26 (1974): 390–91.

8. In the case of Heinlein's *The Door into Summer*, for example, the expression *before* acquires a "new meaning." Some sentences normally entailed by sentences in which *before* occurs are no longer legitimately derivable. Usually we are not overly precise about such "new meanings." We simply come to accept without undue worry a claim to the effect that the time traveler genuinely experiences the distant future *before* the immediate future, a claim that would be downright bizarre in a description of the actual world.

9. It seems relevant, for example, that no one at the time of Balzac had even considered nonstandard conceptions of objective time. This sort of argument is allowed by even such a strict anti-intentionalist as Wimsatt in "History and Criticism," in *The Verbal Icon* (Kentucky, 1954), pp. 253–65. Probably Wimsatt should not allow such arguments, given his internalism.

10. This is somewhat similar to justifying the claim that someone is carving a figurehead (although it might to all appearances just as well be a free-standing statue) because he is known to be working in his basement on what appears to be and what he claims is a viking ship.

11. Externalist arguments in criticism are usually complex and difficult, as any argument about human actions and their products.

12. It will be fairly clear that my view of the conventional nature of representation is very similar to the views of Nelson Goodman and Gombrich.