

AN ANALOGICAL FRAMEWORK FOR LITERARY INTERPRETATION *

KEITH J. HOLYOAK

The ability to interpret literary works, which is essential to the writing of commentary, is treated as a product of analogical thinking. The framework proposed here addresses issues beyond the scope of current psychological models of text processing. This analogical framework is an attempted integration of theoretical ideas derived from models of analogical problem solving with ideas from the field of literary criticism. The use of analogy in problem solving is compared and contrasted with its use in understanding literary metaphors and allegories. A taxonomy of types of literary analogy is proposed as part of a task analysis of the interpretive process. A general model of the process of developing analogical interpretations is sketched. A major direction suggested for empirical research is to examine the types of textual elements that readers use to notice analogies and develop potential interpretations.

Introduction

Why should a cognitive psychologist be concerned with literature? A basic justification can be found in a claim forcefully expressed by the critic Frye: "...Verbal structures with meaning are verbal imitations of that elusive psychological and physiological process known as thought, a process stumbling through emotional entanglements, sudden irrational convictions, involuntary gleams of insight, rationalized prejudices, and blocks of panic and inertia, finally to reach a completely incommunicable intuition" (1957: 83). It is the business of cognitive psychology to strive toward a scientific reconstruction of the strange process Frye describes – thinking, complete with its unconscious elements and its links to affect and motivation. Literature is clearly a creation of thought, often an inspiration of it, and (if Frye is at all correct) perhaps its imitation as well. As such the cognitive psychologist would do well to consider what insights may be offered by analyses of literature and its interpretation.

My goal in the present paper is to examine the process of literary interpretation from the perspective of cognitive psychology. My most direct concern is

* Preparation of this paper was supported by an NIMH Research Scientist Development Award, 1-K02-MH00342-02. Eric Rabkin provided valuable directions for my foray into what would otherwise have seemed the trackless forest of literary criticism.

Reprint requests may be sent to Keith Holyoak, University of Michigan, Human Performance Center, 330 Packard Road, Ann Arbor, MI 48104, USA.

therefore not literature but literary commentary. The writing of commentary inevitably begins with reading; and as textbooks on the writing of commentary typically stress (e.g., Roberts 1977), the interpretive skills required to write commentary are continuous with those required to read literature with a degree of sophistication. The literary critic, in addition to exercising a specialized vocabulary and training as an essayist, will be an expert in a domain in which the typical educated reader is a novice. Nonetheless, the essence of the interpretive process may remain the same. My discussion of literary interpretation will therefore bear most directly on interpretive reading broadly construed.

Beyond literal test-processing

In the past few years the topic of text processing has become a major concern within both cognitive psychology and artificial intelligence (e.g., Black and Bower 1980; Kintsch and van Dijk 1978; Rumelhart 1977; Schank and Abelson 1977; Thorndyke 1977). This work has addressed a broad range of issues relating to the reading, writing, recall, and summarization of brief stories. But while a great deal has been learned about important aspects of text processing, a serious gap is apparent from my present perspective: none of the research to date bears directly on the process of literary interpretation. To illustrate this gap I will consider a study by Rumelhart (1977), justifiably considered one of the most important in the field. Rumelhart's goal was to predict the content of summaries and recall protocols produced for simple folk tales. He analyzed several stories in terms of a hierarchical "story schema", and predicted that propositions located relatively high in the resulting hierarchical structure would tend to be included in subjects' protocols. His model provided a very accurate account of the data he collected. For example, here is a summary of a story called "The Countryman and the Serpent" that was given by one of Rumelhart's subjects:

A farmer cut off a serpent's tail after the serpent had killed the farmer's son. The serpent killed many of the farmer's cattle in revenge, and refused peace offerings from the farmer (1977: 290).

According to Rumelhart's model, this is an entirely satisfactory high-level summary of the story; and indeed, it succinctly captures the gist of the events described. But such a summary is certainly not an interpretive commentary, nor even a particularly useful step toward one. (Of course, this observation is not a criticism of either Rumelhart or his quoted subject, neither of whom were concerned with the production of commentary.) What would a commentary look like? Let me try to construct one for "The Countryman and the Serpent". The first step is to return to the text of the story (Rumelhart 1977: 275). A few salient signs (most notably, the attribution of human-like intentions and the

power of speech to an animal) make it clear that the story is a fable. This classification is extremely important, because it immediately creates an expectation that the tale is meant to convey some general point about human conduct.

So, what point might that be? The story begins, "A countryman's son, by accident, trod upon a serpent's tail". Notice the casual phrase "by accident", buried in the middle of the sentence (superficially so insignificant as to be totally excluded from Rumelhart's propositional analysis of the story). That this small transgression, not even intentional, initiates the mutually destructive vengeance that ensues, provides the seed for ironic tragedy. The next sentence is, "The serpent turned and bit him, so that he died". While the story does not explicitly say so, we know that it is the nature of serpents to bite when they are stepped on, and that they exercise no subsequent control over the effects of their venom. The son thus suffers a punishment out of all proportion to his unintended offense; and yet the serpent's action (or reaction) was instinctive, not malicious. The story continues, "The father, in revenge, got his axe, pursued the serpent, and cut off part of his tail". This is the first intentional act of revenge; but from a human perspective it is easy to understand a father's desire to avenge the loss of a child. Then it is the serpent's turn: "So the serpent, in revenge, began stinging several of the farmer's cattle". Again from a human perspective, the serpent's reaction to his mutilation seems understandable. By now a pattern is apparent: the characters, none of whom are villains, have become locked in the grip of mutually destructive and escalating violence.

At this point the farmer gains insight into the situation, and makes a bid to rationally avoid a tragic outcome: "... The farmer thought it best to make it up with the serpent. So he brought food and honey to the mouth of its lair and said to it, 'Let's forget and forgive; perhaps you were right to punish my son and take vengeance on my cattle, but surely I was right in trying to revenge him; now that we are both satisfied, why should we not be friends again?'" This proposal sounds reasonable enough, but the fable ends with a darker vision: "'No, no', said the serpent, 'take away your gifts; you can never, never forget the death of your son, nor I the loss of my tail'".

The message of the fable should now be clear. The dispute between the countryman and the serpent is an example of a kind of tragic pattern that can arise in human affairs. A point may be reached in a conflict after which forgiveness is no longer possible, because the mutual injuries are too severe and irreparable. At that point the origins of the conflict, the intentions of those involved in it, and conventional notions of fairness and adequate redress, become irrelevant. The emotions that have been unleashed are beyond rational control, and the conflict must continue until one or both sides have been destroyed.

A more subtle irony is worth noting. The father and serpent were both avengers and both victims, so the ending could have been reversed: the serpent

might have sued for peace, and been rejected by the countryman. But the actual ending is more effective. The human tries to be rational; is is the serpent, the subhuman one, who has a clearer grasp of the power of human emotions. In rejecting superficial rationality, and accepting the inevitable consequences of emotional forces, the serpent expresses a deeper if more pessimistic wisdom.

As this simple example illustrates, interpretive analysis necessarily goes beyond literal text-processing. Here I am using the term "literal" in a broader sense than is typical among cognitive psychologists. Many psychologists have stressed the importance of inferential processes that fill in elliptical gaps in the overt text, and inferred propositions are sometimes called "non-literal". But as I will use the term in the present paper, processing remains literal as long as it does not introduce meaning that goes beyond the overt topic domain. In "The Countryman and the Serpent" the overt topic domain is a conflict between a man and a serpent. The text does not say, for example, that the man hated the serpent; this would be a plausible inference, but in the present sense a literal one. Text processing only becomes non-literal when it seeks meaning that lies outside the overt topic domain. For example, in my commentary I derived an interpretation of the story in terms of a type of human conflict, thus introducing an ostensible covert topic.

The present distinction between literal and non-literal text-processing is tied to the traditional view that a literary work can be interpreted at multiple levels – an overt or "surface" level, and one or more "deeper" levels. The process of literary analysis is primarily a search for the deeper levels. As an answer to the question, "What does the fable mean?", the summary given by Rumelhart's subject would scarcely do as a reply, even though it certainly gives the story's meaning in some sense. My own commentary, whatever its shortcomings (for which I take refuge behind my amateur status as a critic), is clearly a more appropriate *type* of response. I do not mean to imply, however, that literal and non-literal text-processing are entirely separate processes. On the contrary, they will most often be intertwined. For example, in deriving my interpretation I relied on various literal inferences, such as that the serpent's biting of the son was an instinctive response. Nonetheless, the interpretive task required something more than extensive inference-making about the overt topic domain.

What is that "something more"? Several points are noteworthy. I began by identifying the genre, or type of literary form, that the story represented. My classification of the story as a fable directed my subsequent attempts to interpret it. (See Olson et al. 1981, for evidence that genre distinctions influence literal text-processing.) In particular, I knew that fables are meant to be *exemplary*, in the sense that they typically present symbolic examples of types of human interactions. I also knew that in fables animals usually represent people. I therefore attempted to interpret the overt interactions among characters in more general human terms, making a number of literal

inferences in the process. In deriving an interpretation I moved from an overt *narrative*, describing a sequence of events, to a covert *statement*, expounding a set of ideas about tragic human conflicts. It is this move across levels of interpretation that lies at the very heart of the interpretive process.

The allegorical nature of interpretation

One might well question whether the interpretation of a brief fable, as discussed above, is at all representative of the process of literary analysis. I will argue that it is indeed. The fable, like its religious cousin the parable, belongs to a family of literary forms called *allegory*. In an allegorical work symbolic figures and actions are used to represent generalizations about human conduct or experience. The purest form of allegory is based on personified abstractions, as in an example from Quintilian: "Pale death with equal foot knocks at the poor man's door" (quoted by Fletcher 1964: 86). But the concept of allegory is far broader than the venerable examples so far mentioned might suggest. Fletcher (1964), in an extensive discussion of allegory, characterizes it as a mode of expression that embraces satires, romances, westerns, detective stories, and a variety of other forms. He eschews formal definition, claiming simply that "allegory says one thing and means another". (I would quickly amend this description to say that it *also* means another.) This broad usage certainly has precedents. An especially far-reaching generalization was proposed by Chesterton: "Every great literature has always been allegorical – allegorical of some view of the whole universe" (1902: 47). And the position I am espousing here is expressed in Frye's remark that "all commentary is allegorical interpretation..." (1957: 89). The primary goal of literary interpretation is to elucidate the allegorical meaning of a work, and its relationship to the overt surface level.

The task of a cognitive psychologist, then, is to analyze the mental processes involved in moving from one level of text interpretation to another. The most central of these mental processes – the basis for understanding allegory and its smaller-scale relative, metaphor – is analogical thinking. At the most general level, analogical thinking involves finding a *mapping*, or set of correspondences, between the elements of two (or more) organized bodies of information. In the case most central to our present concern, the mapping will be between an overt and an allegorical level of interpretation. But as we will see, analogical thinking plays a yet broader role in the process of deriving meaning from texts.

My own research on analogical thinking (Gick and Holyoak 1980, in press), as well as other related work (Gentner in press; Winston 1980), has been primarily concerned with the role of analogy in problem solving. There are a variety of links between research on text processing and on analogical problem solving. On the one side, various models of story structure emphasize that

simple stories can be viewed as descriptions of problems and their resolution (e.g., Black and Bower 1980; Rumelhart 1977). On the other, my own research has used simple stories as analogs to subsequent problems. Theories in both domains have been based on hierarchical schematic representations. In the present paper I will try to draw a variety of theoretical links between the framework for analogy that has emerged in the domain of problem solving and the role of analogy in interpretive text-processing. I will first outline a general conception of analogical problem solving (based on Holyoak in press) and briefly review some relevant empirical findings. This will serve as a framework for a more detailed discussion of the role of analogy in text-processing.

Be warned, however – the extension of a model of analogical problem solving to text processing is by no means straightforward. As Gentner (in press) points out, problem-oriented and literary analogies serve quite different functions. Since the function is easier to specify in the case of problem solving, that domain will serve as a useful reference point with which literary text-processing can be contrasted. My goal is to use a discussion of analogy in problem solving to highlight certain major dimensions of analogical thinking, which will in turn serve as a guide in analyzing analogical aspects of text processing.

Analogical thinking in problem solving

Analogical mapping

The goal of analogical problem solving is to use a known problem situation (the *base*) to construct a parallel solution to a novel problem (the *target*). This process involves four major steps (not necessarily executed in serial order). (1) A mental representation of the target problem must be constructed. (2) The potential analogy must be noticed; i.e., some aspect of the target must serve as a retrieval cue that reminds the person of the base. (3) An initial *partial mapping* must be found between the elements (objects and their attributes and relations) of the two situations. (4) A solution to the target problem must be constructed by *extending* the mapping.

Each analog can thus be conceptually divided into two parts: that which provides the basis for an initial partial mapping, and that which constitutes the extension of the mapping (i.e., the derived solution). To use the terms introduced by Hesse (1966), analogy involves two distinct types of relationships: the “horizontal” mapping between elements of the two analogs, and the “vertical” relationship between the two parts of a single analog. In the case of analogies between problems, vertical relationships correspond to causal relations within the person’s mental model of each situation (Winston 1980). For example,

certain aspects of the base problem will be viewed as “enabling conditions” for the attained solution. The critical point to note is that the vertical and horizontal relationships within an analogy are intimately related. For while it will seldom be possible to map *all* elements of the base and target, those base elements causally related to its solution must be mappable. If causal elements cannot be mapped, the putative analogy can be rejected as inappropriate.

To provide a concrete example of analogy it will be useful to introduce the main target problem used in the experiments of Gick and Holyoak (1980, in press), as well as a story used as a base analog. The target was Duncker’s (1945) radiation problem:

Suppose you are a doctor faced with a patient who has a malignant tumor in his stomach. It is impossible to operate on the patient, but unless the tumor is destroyed the patient will die. There is a kind of ray that can be used to destroy the tumor. If the rays reach the tumor all at once at a sufficiently high intensity, the tumor will be destroyed. Unfortunately, at this intensity the healthy tissue that the rays pass through on the way to the tumor will also be destroyed. At lower intensities the rays are harmless to healthy tissue, but they will not affect the tumor either. What type of procedure might be used to destroy the tumor with the rays, and at the same time avoid destroying the healthy tissue?

Prior to their attempt to solve the radiation problem, our subjects often read a story about an analogous military problem and its solution, in which a general wishes to capture a fortress located in the center of a country. There are many roads radiating outward from the fortress. All have been mined so that while small groups of men can pass over the roads safely, any large force will detonate the mines. A full-scale direct attack is therefore impossible. The general’s solution is to divide his army into small groups, send each group to the head of a different road, and have the groups converge simultaneously on the fortress. Note that there is an analogous “convergence” solution to the radiation problem. The doctor could direct multiple low-intensity rays toward the tumor simultaneously from different directions, so that the healthy tissue will be left unharmed, but the effects of the low-intensity rays will summate and destroy the tumor.

This pair of problem analogs can be used to elucidate several general points about analogy that will bear on literary interpretation. First, analogs vary in the *similarity* of their respective domains. The military and medical domains are quite dissimilar, as evidenced by the fact that the corresponding objects are drawn from very different semantic categories (e.g., army and rays, fortress and tumor). Only the corresponding *relations* (e.g., capturing and destroying) seem similar; accordingly, the analogy can be classed as *metaphorical*. It is easy to imagine a much more similar analog to the radiation problem (e.g., a story about a doctor treating a brain tumor with multiple lasers), in which the corresponding objects are instances of the same superordinate categories (e.g.,

Table 1

Correspondences among two convergence problems and their schema. (Source: M.L. Gick and K.J. Holyoak, Schema induction and analogical transfer. *Cognitive Psychology*, in press. Reprinted with permission.)

	Initial state	Solution plan	Outcome
Military problem	<i>Goal:</i> Use army to capture forces <i>Resources:</i> Sufficiently large army <i>Constraint:</i> Unable to send entire army along one road	Send small groups along multiple roads simultaneously	Fortress captured by army
Radiation problem	<i>Goal:</i> Use rays to destroy tumor <i>Resources:</i> Sufficiently powerful rays <i>Constraint:</i> Unable to administer high-intensity rays from one direction	Administer low-intensity rays from multiple directions simultaneously	Tumor destroyed by rays
Convergence schema	<i>Goal:</i> Use force to overcome a central target <i>Resources:</i> Sufficiently great force <i>Constraint:</i> Unable to apply full force along one path	Apply weak forces along multiple paths simultaneously	Central target overcome by force

lasers and rays, brain tumor and stomach tumor). Such an analogy could be classed as non-metaphorical [1].

Second, analogical correspondences can be defined at multiple levels of abstraction. Various models of text processing have incorporated hierarchical representation systems. Perhaps the clearest example is Kintsch and van Dijk's (1978; van Dijk 1980) concept of "macrostructure" levels. Table 1 presents an informal representation of the hierarchical structure of the military and medical problems. Each analog is represented as an instance of a very general "problem schema", composed of an initial state (goals, available resources, and constraints), a solution plan, and an actual or anticipated outcome of realizing

[1]I will refer to an analogy between relatively similar analogs as "non-metaphorical" rather than "literal" in order to avoid introducing ambiguity into my usage of the latter term. In accord with the earlier discussion, I will reserve the term "literal" to refer to inferential processes that operate on a single topic domain (or interpretive level). Analogical processing is never literal in this sense, since by definition it involves a comparison between multiple topic domains.

the plan. The problem schema reflects the vertical causal organization of the analogs. The components of the initial state are all causally related to the solution plan: the goal is a *reason* for it; the resources *enable* it; and the constraints *prevent* alternative plans. The outcome is then a *result* of executing the solution plan.

The hierarchical nature of analogical mapping has several important consequences. In general, finding correspondences at an abstract level (e.g., matching the initial states of the two problems) will constrain additional correspondences at lower levels. It is typically the case that at relatively concrete levels many mapping failures occur (e.g., the general disperses his troops by a verbal command, which certainly does not suggest a promising strategy for controlling the strength of rays). In addition, since not all possible inferences about the base and target are likely to be drawn, some correspondences will remain indeterminate (Holyoak in press). Problem analogies are thus typically incomplete (as is also the case, we will find, for allegorical interpretations).

Schema induction

Table 1 also includes a statement of a “convergence schema” – a representation of the type of problem for which convergence solutions are possible. For example, the goal in such problems is to use a force to overcome a central target. The schema preserves the commonalities between the two analogs while eliminating the differences between them; accordingly, it can be viewed as an abstract category of which the analogs are instances. There is a sense in which the schema is in fact implicit in each of the analogs. But as we will see, there are psychological advantages in storing a schema in memory as an independent concept. Note also that each analog can be viewed as a transformation of the abstract schema into a set of concepts appropriate to a specific domain. An analogy between problems thus consists of three key elements – the base, the target, and the transformation relating each to the schema (and hence to each other). We will see that variations in the content of the same three basic elements help to define a taxonomy of analogical mappings relevant to literary interpretation.

The concept of a schema is now current in cognitive modelling (for an overview see Rumelhart 1980), including the areas of both text processing and problem solving (Chi et al. 1981; Larkin et al. 1980). If a schema can be learned by induction from analogous examples, the abstract knowledge structure may effectively summarize the information conveyed by the examples. The most obvious way a schema could be acquired is by mapping two examples to identify their structural commonalities. Gick and Holyoak (in press) had subjects write descriptions of the similarities between two stories of the convergence type (e.g., the military story described earlier, and a story about a fire-fighter who used multiple hoses to extinguish a blaze). These descriptions

were later scored for presence and quality of the convergence schema. The subjects then attempted to solve the analogous radiation problem. A striking relationship was evident between the quality of the induced schemas and success in generating the convergence solution. Subjects whose descriptions clearly expressed the convergence schema (e.g., "Both stories used the same concept to solve a problem, which was to use many small forces applied together to add up to one large force necessary to destroy the object") were much more likely to produce the target solution than were those whose descriptions noted less relevant similarities (e.g., "In both stories a hero was rewarded for his efforts"). Interestingly, having subjects write summaries of a single story analog (as opposed to attempting to recall it verbatim) had no clear positive effect on subsequent analogical transfer. The kind of abstraction that serves to induce a problem schema appeared to depend on a comparison between multiple examples.

Mapping processes thus seem to serve several interrelated functions. Mapping is used to transfer information from one concrete situation to another, or from a more abstract schema to a concrete example. In addition, mapping plays a key role in the process by which schemas are learned from examples. We will see shortly that various types of schematic concepts are central elements in literary interpretation.

I alluded earlier to the psychological advantages of storing an independent schema in memory, and the Gick and Holyoak (in press) results, described above, confirmed that induction of a problem schema facilitates analogical transfer. One key reason the schema is useful is that the elements of the schema, which are common to both the base and target, afford potential retrieval cues by means of which the target problem may evoke a relevant base analog. If the problem solver fails to encode elements of the schema, in either the base or the target, the potential analogy may be missed. Noticing an analogy between semantically disparate domains is clearly a non-trivial problem. In many of our experiments (Gick and Holyoak 1980, in press) subjects first attempted to solve the radiation problem without any hint from the experimenter suggesting that the prior story analog might be helpful. Only about a third of the subjects who would potentially apply the analogy (once a hint was given) spontaneously noticed it. Manipulations that fostered schema induction (such as provision of multiple analogs) increased the frequency of spontaneous noticing. We will see that the corresponding problem in literary interpretation – noticing an allegorical level of meaning – is also a probable source of processing difficulty.

Analogical thinking in literary interpretation

Overview

In this section I will attempt to develop in more detail an analogical approach to the psychology of literary interpretation. My aim is to erect a tentative theoretical framework that may guide future research in this area. I hope this framework will point toward testable predictions; however, it will necessarily lack immediate empirical support, for the simple reason that no directly relevant research has been done. My discussion will largely focus on the question of what literary interpretation *is*, but with a view toward the broader question of how it is accomplished. In this emphasis on *task analysis* I am following precedents in the development of theories of problem solving, in which analyses of the nature of cognitive tasks have provided strong constraints on models of how they are performed (Newell and Simon 1972). Since a task analysis of literary interpretation can be informed by theoretical approaches to literary criticism, my discussion will attempt to integrate ideas from the field of criticism with ideas from cognitive psychology.

Form and function in literary metaphor

As argued earlier, literary analysis can be viewed as the process of moving from an overt topic domain to a covert level of interpretation. In relation to my earlier description of analogical problem solving, the overt topic domain – what the text is directly about – corresponds to the *base* analog, and the interpretive level corresponds to the target. This terminology happily converges with Kintsch and van Dijk's (1978) concept of a *text base*. In the following discussion I will take as given some mental representation of a text base elaborated by appropriate inferences. This is admittedly an idealization, since the inferences drawn about the overt topic domain will often be guided by attempts to find an analogical interpretation. However, as an initial approximation it is useful to view the reader as having encoded a reasonably complete representation of the surface-level text base, and to focus analysis on the processes involved in relating the text base to other interpretive levels.

An important point to notice is that the initial phase of literary interpretation is essentially the reverse of analogical problem solving. In the case of problem solving, the person faces an inadequately-understood target problem, and must notice and retrieve a known base analog in order to develop a solution to the target. In the literary case, the idealized reader fully understands the text base, but must notice a covert target topic and then use the text base to generate an analogical interpretation. This is but the first of several salient differences between the form and function of analogy in problem solving versus literary analysis.

Various types of analogy are relevant to literary interpretation (see below), but the most central of these can be termed *allegorical*. An allegory, in the sense intended here, is a discursive metaphorical analogy – metaphorical in that it relates disparate semantic domains, and discursive in that the metaphor makes an extended statement about the target topic. I thus assume continuity between the mental processes involved in understanding metaphor as it is embodied in single sentences (in classical terms, a kind of *trope*) and allegory involving large texts (a kind of *figure*; Fletcher 1964: 84) [2].

The above assumption is bolstered by the recent analysis of metaphor offered by Lakoff and Johnson (1980). Their linguistic evidence suggests that sentential metaphors are commonly based on broad conceptual analogies. For example, the concept of an “idea” seems to be understood in part by analogy to food. This analogy yields an indefinitely large set of sentential metaphors, such as “That’s food for thought”, “I can’t swallow that claim”, and “That idea has been fermenting for years” (Lakoff and Johnson 1980: 46–47). There thus seems to be a continuity between large-scale analogies and individual metaphors. Many originally metaphorical expressions are so ingrained that their analogical basis is no longer immediately apparent. However, the underlying conceptual analogy may still be revealed by creative extensions. Consider this example from Hemingway: “They say the seeds of what we will do are in all of us, but it always seemed to me that in those who make jokes in life the seeds are covered with better soil and with a higher grade of manure” (1964: 104). Here the conventionalized analogy between “initiating conditions” and seeds is revitalized by an original extension.

The continuity between sentential metaphors and large-scale allegories may be clarified by an example that falls between them – an extended but directly-stated metaphor. This example, again drawn from Hemingway, is a brief depiction of Scott Fitzgerald:

His talent was as natural as the pattern that was made by the dust on a butterfly’s wings. At one time he understood it no more than the butterfly did and he did not know when it was brushed or marred. Later he became conscious of his damaged wings and of their construction and he learned to think and could not fly any more because the love of flight was gone and he could only remember when it had been effortless (1964: p. 147).

This passage, which is overtly an analogy, can be used to illustrate several distinctions among metaphoric devices, literary interpretations, and problem analogies. Table 2 presents an informal description of the analogy between a writer’s talent and a butterfly’s flight. The former is the target and the latter is the base. Unlike cases of interpretation in which the target must be inferred,

[2]The present terms “target” and “base” can be respectively viewed as approximate generalizations of Black’s (1962) terms “principal subject” and “subsidiary subject”, and of Richard’s (1936) terms “tenor” and “vehicle” (at least for one usage of each of the latter terms).

Table 2
Hemingway's analogical depiction of Fitzgerald.

Writer's Talent	Butterfly's flight
(1) Writer's talent is natural	(1') Butterfly's pattern on wings is natural
(*2) Talent is beautiful/fragile	(*2) Wings are beautiful/fragile
(3) At t_1 , writer was unconscious of talent	(3') At t_1 , butterfly was unconscious of wings
(4) At t_2 , writer became conscious of talent and its shortcomings	(4') At t_2 , butterfly became conscious of wings and their damage
(5) Consciousness causes loss of love of exercising talent	(5') Consciousness causes loss of love of flight
(6) Loss of love of exercising talent causes loss of ability to write	(6') Loss of love of flight causes loss of ability to fly
(7) At t_3 , writer remembers lost effortlessness of writing	(7') At t_3 , butterfly remembers lost effortlessness of flight
(*8) This memory adds to the writer's sense of loss	(*8') This memory adds to the butterfly's sense of loss

Schema:

Some natural skills are both beautiful and fragile.

Such skills are best exercised unconsciously.

If the skill is made conscious, inadequacies will be made salient, and love of exercising the skill will be lost.

At that point the skill will be lost, and its possessor will be saddened by memory of it.

and cases of problem solving in which the base must be inferred, the metaphorical form in this passage directly indicates both domains. Indeed, Hemingway explicitly draws the major correspondences between target and base, which are indicated by numbered propositions in table 2. It is instructive to note the progression in the metaphoric devices used to convey the analogy. The first sentence and the first clause of the second use similes to match aspects of the two domains. These explicit comparatives ("as natural as", "no more than") point out key correspondences (e.g., writer and butterfly). In the second clause of the second sentence the passage shifts to a stronger metaphorical form; comparatives are no longer used, and the writer's talent is directly described as "brushed and marred" – terms appropriate to a butterfly. Finally, by the third sentence the two domains have been completely identified. The subject pronoun "he" indicates that the writer is the topic, but the rest of the sentence is a description of a personified butterfly.

Although the passage expresses an analogy, the above literary devices make it very different in surface form from a problem analogy, in which the two analogs are compared but not identified with each other. Such surface differences in part reflect differences in function. In problem solving, the goal is

to derive a solution to the target. On the other hand, a literary metaphor typically is not directed at a “problem”; it simply serves to make a statement about the target domain. Table 2 indicates various major propositions that match across the domains compared in Hemingway’s analogy. Although these can be roughly separated into the initial mapping and extensions of it, the division is less clearcut than in the case of problem analogs. Two pairs of propositions that I take to be salient inferred extensions are marked by asterisks in table 2. Neither of these is really a “conclusion” to the analogy, in the sense that a derived solution to the target is a conclusion in the case of problem solving. They simply reflect propositions about the base domain that can be used to infer parallel propositions about the target. To the extent that the metaphor has any real “conclusion”, it would be the entire set of propositions about the target. Because the metaphor’s primary function is to make a statement, rather than to derive a solution, causal relations within each domain are relatively less central to the mapping process than they are in the case of analogical problem solving.

Literary metaphors often yield less precise mappings than do problem analogs (Gentner in press). That is, the mapping between elements may be less consistent, and mapping failures tend to be less damaging. For example, in Hemingway’s passage the writer’s talent is first mapped with the pattern on a butterfly’s wings, and later with the wings themselves. If the analogy were decomposed in further detail than was attempted in table 2, it arguably becomes incoherent. A butterfly’s pattern is not causally related to its flight, so if talent is mapped with the pattern, then there is no reason why consciousness of the talent/pattern should interfere with the ability to exercise it. However, from the perspective of literary function one could argue that such a mapping failure simply indicates the analogy is being decomposed too far. A strong associative relation (technically, *metonymy*) links the concepts of pattern, wings, and flight; accordingly, they can be interchanged quite freely, so that talent may be mapped with any of them.

Indeed, the apparently dubious initial mapping between talent and pattern may serve a broader purpose. The phrase “pattern made by the dust on a butterfly’s wings” is strongly suggestive of painting, and hence provides a clue that the passage can be interpreted more broadly, as a statement about various modes of artistic skill. Just as in the case of problem analogs, a literary metaphor not only conveys information about the target, but also potentially allows the induction of a *schema* that captures the commonalities between the target and base [3]. At the bottom of table 2, I have sketched a more abstract schema derived from Hemingway’s passage – a statement not just about writing, but about a broader class of skills. The passage thus has multiple levels

[3] Schema induction seems to correspond closely to Black’s (1962) concept of an interaction between principal and subsidiary subjects.

of interpretation. It is (we see from its title) about Fitzgerald, and about writing, and about artistic skills, and about skills in general. Note that these possibilities are conjoined by “and”, not “or”. The passage can be interpreted as being about any or all of these topics, and perhaps others as well. And so we move along the continuum from overt metaphorical comparison to covert allegorical interpretation.

Taxonomy of literary analogy

At this point we need to grapple directly with a task analysis of literary interpretation. As a first step I will outline a tentative taxonomy of the types of analogical processes that seem most central to literary analysis. This taxonomy is based on four crucial dimensions of variation. (1) The first of these is the *content* of the target domain to which the text base is compared. (2) A second is the *level of abstraction* of the target domain, which may vary from concrete to highly schematic. (3) A third dimension, which interacts with the second, may vary whenever two domains at comparable levels of abstraction are compared. The reader may emphasize either the direct mapping from base to target, or induction of a more abstract schema that captures their commonalities. Schema induction introduces a new, more abstract interpretive level. (4) The fourth dimension concerns what I will term the *polarity* of the analogical relation between the base and target, which may be either *compatible* or *ironic*.

The second and third dimensions (level of target abstraction, and the relation between mapping and schema induction) have already been discussed at some length. I will therefore concentrate on providing a fuller exposition of variations in target content and relational polarity, illustrating variations in the other dimensions in passing.

Content of the target domain

Four broad classes of literary analogies can be defined by variations in the content of the domain treated as target.

Extended self

For the typical reader, the most immediate form of analogical processing is based on comparisons between characters in the story and what I will term the “extended self”. Such comparisons generally involve the identification of one’s self with the protagonists; however, the broader term is meant to encompass cases in which analogies are drawn with friends, lovers, or other people closely tied to the self. A good literary work can provoke rather remarkable mental transformations of the self: “Ordinarily I would prefer not to think of myself as a murderer, as a suicide, or as a middle-aged failure cuckolded by his wife. Yet in Dostoevsky’s *Crime and Punishment*, Faulkner’s *The Sound and the*

Fury, and Joyce's *Ulysses* I am forced to recognize and come to terms with my participation in the fate of Raskolnikov, of Quentin Compson, and of Leopold Bloom" (Cawelti 1976: 18).

Identification of the self with characters is clearly one of the primary sources of emotional involvement in the reading of literature. In "formulaic" or "escapist" works, identification with an heroic protagonist (such as James Bond) provides a satisfying confirmation of an idealized self-image (Cawelti 1976). Analogies involving the extended self are concrete and non-metaphorical (involving a "real" and a "hypothetical" person). However, a comparison of the self and a protagonist may, through schema induction, create a more abstract interpretation at the level of general human behavior.

Social world

The second major type of content domain that may serve as a target for interpretation is what I will term the "social world". This is the broadest domain of all, and might as well be called "life". When a literary work seems to make a statement about the human condition, the interpretation may arise from a perceived analogy between a situation described by the text base and the reader's knowledge of parallel situations that occur in "real life". To draw a small example from a giant work, consider the following situation from Part Six of Tolstoy's *Anna Karenina*. Over several chapters an expectation has been building that Koznyshov will propose marriage to Varenka. Finally they are walking together in the woods and the climactic moment is evidently at hand, as Koznyshov has made up his mind to propose.

He had to declare himself now or never.... He...repeated to himself the words in which he had intended to propose; but instead of these words, because of some unexpected thought that came to mind, he suddenly asked:

'What difference is there between the edible white mushrooms and the birch-tree kind?'

Varenka's lips quivered with emotion as she replied: 'There's hardly any difference in the caps, it's just in the stems.'

And the moment she said this, both he and she realized that it was all over... (Tolstoy 1960: 604).

At an abstract level of macrostructure, this text base describes a situation in which a person makes a momentous decision only to turn aside from it at the last moment for no clear reason. This event will heavily influence the future course of human lives; whether the impact of the reversal is good or bad will never be surely known. At this representational level it is fairly easy to think of actual or hypothetical analogs (such as a career decision, perhaps); noticing such "real life" analogs may suggest an interpretation of the passage as a statement about certain general ironic possibilities of life.

Intertextual analogies

Literary criticism has often been polarized between the view that works

should be viewed in relation to life, and the view that they should be viewed in relation to other works. A psychological model of the interpretive process must clearly embrace both. I will refer to analogies from a text base to representations of other texts or categories of texts as *intertextual*. Such analogies can be found at various levels of abstraction: between two particular works, such as *Macbeth* and Kurosawa's film *Throne of Blood*; between an individual work and a formula or genre, such as *The Godfather* and the generic concept of "gangster story"; and between two categories, such as the western and tales of knightly adventure (Cawelti 1976). Analyses of intertextual analogies often serve to define the unique qualities of a work, by identifying aspects that distinguish it from comparable texts.

One experiment in the cognitive literature on story memory in fact provides evidence that college students sometimes spontaneously notice intertextual analogies. Thorndyke (1977) had subjects study and then recall two successive stories. The stories were constructed to have either the same or different plot structures, as defined by a "story grammar". Recall of the second story was more accurate if the two stories shared the same structure. It is clear from inspection of Thorndyke's stories that those based on the same structure are semantically analogs of each other (e.g., one pair consists of a description of a political dispute in which the characters are either humans or fabulous animals). Indeed, Thorndyke reported that most of his subjects in the "same-structure" conditions noted during debriefing that the two stories had the same "idea" or "theme". Accordingly, Thorndyke's results provide evidence that recall can be facilitated by perception of an intertextual allegory.

Intratextual analogies

The fourth variant in the content of target domains, which I will term *intratextual*, involves analogies between elements of the text base. Such analogies have a rather different status than the types previously mentioned, in that the base and target are *both* part of the text base. Overt metaphors, such as the example by Hemingway discussed earlier, are intratextual analogies. But many potential intratextual analogs are not so explicitly marked for comparison, nor are such analogies necessarily metaphorical. For example, two couples are the central protagonists in *Anna Karenina* (Anna and Vronsky, Levin and Kitty). All four interact as the couples are first paired; then the plot tends to alternate between their relatively separate lives. The reader is therefore likely to compare the development of the two relationships (thus constructing a non-metaphorical analogy). Kitty initially rejects Levin, but they eventually achieve a happy, relatively conventional marriage. Anna and Vronsky fall in love immediately, but their passionate relationship eventually ends in Anna's suicide (which occurs at nearly the same time as the birth of Kitty's first child). An analysis of the similarities and differences between these two relationships is likely to contribute to an interpretation of the work. A variety of stylistic devices for

juxtaposing story elements may often provide clues to intratextual analogies (Rabkin 1977). The most striking of these involve extensive "doubling" of plots and characters, as in "Jekyll and Hyde" stories (Fletcher 1964).

Polarity of the analogical relationship

I have now outlined variations in the content of the target domain, and touched upon the dimensions of level of target abstraction and direct mapping versus schema induction. The fourth dimension of literary analogy concerns the polarity of the relationship between the two analogs. This relationship may be one of either *compatibility* or *irony*. An analogical relationship is compatible if the two interpretive levels are mutually consistent or supportive. For example, when I treated the text base of "The Countryman and the Serpent" (a description of a conflict between humans and an animal) as an example of a class of human conflicts, the relation between the two levels was compatible.

In contrast, an analogical relationship is ironic whenever the two domains are somehow in opposition to each other. Irony typically arises when there is an incongruity between what might be expected, or what appears to be, and what actually is (Kernan 1965). In "The Countryman and the Serpent" the relationship between the text base and my interpretation was compatible, but that between the interpretation and the "expected" social world was ironic (in that irrational passion unexpectedly seemed more wise than rationality). Frye (1957) has argued that literature in modern times has become increasingly ironic, often conveying the message that things are somehow less than they seem. Irony is commonly associated with a negative or pessimistic attitude, ranging from gentle humor at life's imperfections to bleak despair. It is not surprising that irony remains the dominant pole of twentieth century literature: to go about everyday life under the shadow of holocaust is an inherently ironic task. Frye is misleading, however, when he refers to irony as "anti-allegorical" (1957: 91-92). Irony is inevitably a product of analogical thinking, in that it depends on the perception of opposition between domains. An ironic interpretation of a text can clearly be allegorical, albeit with a negative polarity (see Fletcher 1964: 340-341).

It is an admitted oversimplification to treat polarity of relation as a unidimensional concept, since the relation of opposition can take a variety of forms. As Kernan (1965: 82) points out, irony blends into such similar relations as ambiguity and paradox. But as Kernan also argues, irony appears in a particularly clear form in *satire*. A major form of satire, exemplified by Swift's *A Modest Proposal*, hinges on an ironic analogy between the text base and some aspect of the social world. Satire, by virtue of its inherent duality, is the natural weapon of choice for attacks on all forms of hypocrisy and pretentiousness. The text base often superficially praises its target, yet contains ample clues that the intended interpretation is one of condemnation. A favorite

device is to seriously laud aspects of the target domain that have been exaggerated to transparently grotesque proportions. A satire may also be based on an ironic intertextual analogy, in which case it is a *parody*. A parody typically satirizes another work either by exaggerating less prepossessing aspects of its style, or by using its style to describe manifestly inappropriate content (Hight 1962). Parody is more than a literary category; it also appears in other art forms, such as painting and music, and in rituals, such as the satanic Black Mass (which consists of ironic reversals of the Catholic rite). Such illustrations highlight an important general point: allegorical literary interpretations are but one of many manifestations of analogical thought.

Toward a process model of literary interpretation

In the preceding sections I attempted to survey the forms of literary analogy with a degree of systematicity. With this taxonomy in mind, let me outline a general task analysis of the process of literary interpretation.

At a global level the interpretive process can be divided into several interactive steps, analogous to those included in my earlier description of analogical problem solving. (1) The reader must develop a mental representation of the text base, elaborated by appropriate literal inferential processes. (2) One or more clues in the text base may lead the reader to notice a potential covert topic. (3) The reader may then develop an initial partial mapping between salient elements of the text base and elements of the hypothesized target domain. (4) The reader may extend the mapping between the base and target domains to construct an analogous statement about the latter. (5) As in the case of problem solving, the reader may induce a more abstract schema that captures the commonalities between the base and target. The resulting schema will then constitute an additional "higher-order" level of interpretation.

A process model of literary interpretation can be expected to share a variety of features with a model of analogical problem solving (Holyoak in press). The steps outlined above are likely to be highly interactive and under the strategic control of the reader. Thus an initial hypothesis regarding a possible interpretation for a work may trigger a conscious effort to work out an analogical mapping; if the outcome seems unsatisfactory, the reader may seek an interpretation in an alternative target domain. Often the mapping process is likely to proceed in a "top-down" fashion, beginning with the detection of correspondences at an abstract level of macrostructure. For example, the behavior of a protagonist may be classified as an example of a broad class of human behavior or human interactions, such as "pride" or "jealousy" (cf. Abelson 1973). The fate of the protagonist can then be interpreted as a statement about this general characteristic; e.g., Kipling's *The Man Who Would Be King* can be readily interpreted as a reiteration of the perennial truth, "Pride goeth before a fall".

The top-down nature of the interpretive process helps to clarify a controversial issue: to what extent are literary works “really” allegorical? People often seem to read works simply “for the story”, with little apparent concern for any sort of secondary “interpretation”. Sometimes the reader may simply fail to notice an allegorical interpretation (see below). However, the present analogical framework clearly allows works to vary along a continuum of inherent interpretability. While some works yield a detailed allegorical interpretation (e.g., Orwell’s *Animal Farm*), in others the allegorical meaning may be confined to an abstract level of macrostructure (such as the above interpretation of *The Man Who Would Be King*). The detail of a derived interpretation will be limited both by the structure of the text base and by the perceptiveness of the reader. In any case, the impact of a work will be determined by the psychological coherence and interest of the text base itself, in addition to whatever covert interpretation it may suggest. Indeed, some of the more strictly allegorical works can be criticized on artistic grounds because the surface plot seems unnaturally forced by a heavy-handed interpretation imposed by the author (Fletcher 1964). Conversely, a work that affords a detailed allegorical interpretation (e.g., Golding’s *Lord of the Flies*) may well be read simply as a “good story”.

One of the most central issues encountered in studies of analogical problem solving has a clear counterpart in literary interpretation. How can a reader notice a potential interpretation? Under what conditions will the reader develop a secondary interpretation at all? Such questions, which have long faced teachers of literature courses, have so far escaped the attention of cognitive psychologists. Doubtless there is a tremendous range in the difficulty of extracting analogical meaning from texts, and some interpretations may elude all but specialized critics. However, many literary analogies are very likely noticed by almost all readers of any degree of sophistication. In particular, given the universality of egocentric interests, virtually everyone is likely to attempt to map their extended self with protagonists in the text.

Other levels of interpretation are fostered by a wide range of literary devices. One so obvious as to be easily overlooked is simply the author’s *selection* of what is described in the text base. For example, a typical “hard-boiled” detective story by Mickey Spillane contains numerous scenes of sordid violence; it is only a small step to infer the message that life is, by and large, sordid and violent (Cawelti 1976). A more subtle form of selection is the repetitive mention of some detail, which may suggest it has symbolic importance. And as mentioned earlier, the juxtaposition of plot elements may foster the detection of intratextual analogies.

Other clues to interpretation depend on various forms of shared cultural knowledge. For example, experience with a category of literature may predispose the reader to seek certain types of interpretation for a work identified with that category (e.g., westerns often convey messages about the relationship

between violence and civilized society; Cawelti 1976). In other cases certain conceptual analogies have become conventionalized, such as the identification of contagion with sin in Christian literature (Fletcher 1964). Some conceptual analogies that figure prominently in literature span such a broad range of history and culture that they may be justifiably termed *archetypes* (Frye 1957, 1963; for a cautionary critique of Frye's views, see Wimsatt 1966). The mythic journey, from the *Odyssey* to modern science fiction, has symbolized the quest for understanding of the human condition. The progression of the seasons suggests the transition from birth through maturity to death, as well as the hope of resurrection (as in Mishima's *Spring Snow*). Islands, which are powerfully associated with physical isolation, have often figured in allegorical statements about social or spiritual isolation (as in Defoe's *Robinson Crusoe*). Such archetypal symbols have deep roots in near-universal aspects of human experience and cognition.

One of the many potential directions for psychological research on the interpretive process is to investigate the roles played by different types of text-base elements in noticing and developing potential interpretations. A closely related direction suggests empirical studies of the influence of various forms of prior knowledge, such as experience with a formulaic category (e.g., the western). In addition, a descriptive theory of analogical reasoning may eventually help to address the normative question of how putative allegorical interpretations can be justified (a central issue in literary criticism).

The analogical framework outlined here suggests that a cognitive psychologist has good reason to be concerned with literature and its interpretation. The ability to notice and understand analogies, one of the most striking hallmarks of human intelligence, is nowhere manifested more clearly. Analogical thinking may be profitably examined through the mirror of its verbal imitations.

References

- Abelson, R.P. 1973. 'The structure of belief systems'. In: R.C. Schank and K.M. Colby (eds.), *Computer models of thought and language*. San Francisco: W.H. Freeman.
- Black, M. 1962. 'Metaphor'. In: M. Black, *Models and metaphors*. Ithaca, NY: Cornell University Press.
- Black, J.B. and G.H. Bower. 1980. Story understanding as problem-solving. *Poetics* 9: 223-250.
- Cawelti, J.G. 1976. *Adventure, mystery, and romance: formula stories as art and popular culture*. Chicago: University of Chicago Press.
- Chesterton, G.K. 1902. 'A defence of nonsense'. In: G.K. Chesterton, *The defendant*. New York: Dodd, Mead & Co.
- Chi, M.T.H., P.J. Feltovich and R. Glaser. 1981. Categorization and representation of physics problems by experts and novices. *Cognitive Science* 5: 121-152.
- Duncker, K. 1945. On problem solving. *Psychological Monographs* 58 (whole no. 270).
- Fletcher, A. 1964. *Allegory: the theory of a symbolic mode*. Ithaca, NY: Cornell University Press.
- Frye, N. 1957. *Anatomy of criticism*. Princeton, NJ: Princeton University Press.

- Frye, N. 1963. *Fables of identity: studies in poetic mythology*. New York: Harcourt, Brace & World.
- Gentner, D. In press. 'Are scientific analogies metaphors?' In: D.S. Miall (ed), *Metaphor: problems and perspectives*. Brighton, Sussex: Harvester Press.
- Gick, M.L. and K.J. Holyoak. 1980. Analogical problem solving. *Cognitive Psychology* 12: 306-355.
- Gick, M.L. and K.J. Holyoak. In press. Schema induction and analogical transfer. *Cognitive Psychology*.
- Hemingway, E. 1964. *A moveable feast*. New York: Charles Scribner's Sons.
- Hesse, M.B. 1966. *Models and analogies in science*. Notre Dame: University of Notre Dame Press.
- Highet, G. 1962. *The anatomy of satire*. Princeton, NJ: Princeton University Press.
- Holyoak, K.J. In press. 'Analogical thinking and human intelligence'. In: R.J. Sternberg (ed.), *Advances in the psychology of human intelligence*, Vol. 2. Hillsdale NJ: Erlbaum.
- Kernan, A.B. 1965. *The plot of satire*. New Haven: Yale University Press.
- Kintsch, W. and T.A. van Dijk. 1978. Toward a model of text comprehension and production. *Psychological Review* 85: 363-394.
- Lakoff, G. and M. Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.
- Larkin, J.H., J. McDermott, D. Simon and H.A. Simon. 1980. Expert and novice performance in solving physics problems. *Science* 208: 1335-1342.
- Newell, A. and H.A. Simon. 1972. *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Olson, G.M., R.L. Mack and S.A. Duffy. 1981. Cognitive aspects of genre. *Poetics* 10: 283-315.
- Rabkin, E.S. 1977. Spatial form and plot. *Critical Inquiry* 4: 253-270.
- Richards, I.A. 1936. *The philosophy of rhetoric*. London: Oxford University Press.
- Roberts, E.V. 1977. *Writing themes about literature*. (4th ed.) Englewood Cliffs, NJ: Prentice-Hall.
- Rumelhart, D.E. 1977. 'Understanding and summarizing brief stories'. In: D. Laberge and S.J. Samuels (eds.), *Basic processes in reading: perception and comprehension*. Hillsdale, NJ: Erlbaum.
- Rumelhart, D.E. 1980. 'Schemata: the building blocks of cognition'. In: R. Spiro, B. Bruce and W. Brewer (eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Erlbaum.
- Schank, R.C. and R.P. Abelson. 1977. *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.
- Thorndyke, P.W. 1977. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology* 9: 77-110.
- Tolstoy, L. 1960. *Anna Karenina*. Translated by J. Carmichael. New York: Bantam Books.
- van Dijk, T.A. 1980. *Macrostructures*. Hillsdale, NJ: Erlbaum.
- Wimsatt, W.K. 1966. 'Northrop Frye: criticism as myth'. In: M. Krieger (ed.), *Northrop Frye in modern criticism*. New York: Columbia University Press.
- Winston, P.H. 1980. Learning and reasoning by analogy. *Communications of the ACM* 23: 689-703.

Keith J. Holyoak is an Associate Professor of Psychology at the University of Michigan. A Stanford Ph.D., his main interests are in the areas of reasoning, memory, and language.