with permanent credibility depends upon recognizing that credibility is no more to be equated with belief than being red is to be equated with looking red. We often believe what is not credible and disbelieve what is credible. Standards of credibility do not vary with individual opinion; they are constant in the world of worlds sketched in my book but may vary from one world of worlds to another.

And my argument that the arts must be taken no less seriously than the sciences is not that the arts "enrich" us or contribute something warmer and more human, but that the sciences as distinguished from technology, and the arts as distinguished from fun, have as their common function the advancement of understanding.

NELSON GOODMAN

Harvard University

UNDERSTANDING NATURAL LANGUAGE *

THE trouble with Artificial Intelligence is that computers don't give a damn—or so I will argue by considering the special case of understanding natural language. Linguistic facility is an appropriate trial for AI because input and output can be handled conveniently with a teletype, because understanding a text requires understanding its topic (which is unrestricted), and because there is the following test for success: does the text enable the candidate to answer those questions it would enable competent people to answer? The thesis will not be that (human-like) intelligence cannot be achieved artificially, but that there are identifiable conditions on achieving it. This point is as much about language and understanding as about Artificial Intelligence. I will express it by distinguishing four *different* phenomena that can be called "holism": that is, four ways in which brief segments of text cannot be understood "in isolation" or "on a one-by-one basis."

* To be presented at an APA symposium on Artificial Intelligence, December 28, 1979. C. Wade Savage and James Moor will comment; see this JOURNAL, this issue, 633/4, for an abstract of Moor's comment; Savage's paper is not available at this time.

I am grateful for suggestions from Nuel Belnap, Bob Brandom, Bert Dreyfus, Jay Garfield, and Zenon Pylyshyn. Kurt Baier helped with the German, and Genevieve Dreyfus with the French.

0022-362X/79/7611/0619$01.40

I. HOLISM OF INTENTIONAL INTERPRETATION

Consider how one might *empirically* defend the claim that a given (strange) object plays chess. Clearly, it is neither necessary nor sufficient that the object use any familiar chess notation (or pieces); for it might play brilliant chess in some alien notation, or it might produce "chess salad" in what appeared to be standard notation. Rather, what the defense must do is, roughly:

(i) give systematic criteria for (physically) identifying the object's inputs and outputs;
(ii) provide a systematic way of interpreting them as various moves (such as a manual for translating them into standard notation); and then
(iii) let some skeptics play chess with it.

The third condition bears all the empirical weight, for satisfying it amounts to public *observation* that the object really does play chess. More specifically, the skeptics see that, as interpreted, it makes a sensible (legal and plausible) move in each position it faces. And eventually, induction convinces them that it would do so in any position. Notice that, *de facto*, the object is also being construed as "remembering" (or "knowing") the current position, "trying" to make good moves, "realizing" that rooks outrank pawns, and even "wanting" to win. All these interpretations and construals constitute collectively an *intentional interpretation*.

Intentional interpretation is intrinsically holistic. It is supported empirically only by observing that its object makes generally "sensible" outputs, given the circumstances. But the relevant circumstances are fixed by the object's prior inputs and other outputs, *as interpreted*. Thus, each observation distributes its support over a whole range of specific interpretations, no one of which is supported apart from the others. For example, a chess move is legal and plausible only relative to the board position, which is itself just the result of the previous moves. So one output can be construed sensibly as a certain queen move, only if that other was a certain knight move, still another a certain bishop move, and so on.[1]

[1] A different argument for a similar conclusion depends on assuming that the inputs and outputs are semantically compound. Then, since each compound will in general share components with many others, their respective interpretations (in terms of their compositions) will be interdependent. Thus the (semantic) role of 'P' in 'P-K4' must be systematically related to its role in 'P-R3', and so on. The argument in the text, however, is more fundamental. There are fewer than two thousand possible chess moves. [Martin Gardner, in his June 1979 *Scientific American* column, gives the figure 1840; but he neglects castling and pawn promotion (see pp. 25/6)]. These could be represented unambiguously by arbitrary numbers, or even simple symbols; yet interpreting an object using such a system would still be holistic, for the earlier reasons.

This is the *holism of intentional interpretation*; and it is all too familiar to philosophers. Intentional interpretation is tantamount to Quine's "radical translation"—including, as Davidson emphasizes, the attribution of beliefs and desires. The condition that outputs be "sensible" (in the light of prior inputs and other outputs) is just whatever the ill-named "principle of charity" is supposed to capture. I have reviewed it here only to distinguish it from what follows.

II. COMMON-SENSE HOLISM

Years ago, Yehoshua Bar-Hillel pointed out that disambiguating "The box was in the pen" requires common-sense knowledge about boxes and pens. He had in mind knowledge of typical sizes, which would ordinarily decide between the alternatives 'playpen' and 'fountain pen'.[2] In a similar vein, it takes common sense to determine the antecedent of the pronoun in: "I left my raincoat in the bathtub, because it was still wet." More subtly, common sense informs our appreciation of the final verb of: "Though her blouse draped stylishly, her pants seemed painted on."

Straightforward questioning immediately exposes any misunderstanding: Was the bathtub wet? Was there paint on her pants? And the issue isn't just academic; a system designed to translate natural languages must be able to answer such questions. For instance, the correct and incorrect readings of our three examples have different translations in both French and German—so the system has to choose. What's so daunting about this, from the designer's point of view, is that one never knows which little fact is going to be relevant next—which common-sense tidbit will make the next disambiguation "obvious." In effect, the whole of common sense is potentially relevant at any point. This feature of natural-language understanding I call *common-sense holism*; its scope and importance was first fully demonstrated in Artificial Intelligence work.

The difference between common-sense holism and the holism of intentional interpretation is easily obscured by vague formulas like: the meaning of an utterance is determinate only relative to *all* the utterer's beliefs, desires, and speech dispositions. This covers both holisms, but only at the price of covering up a crucial distinction. The holism of intentional interpretation is *prior* holism, in the sense that it's already accommodated *before* the interpretation of ongoing discourse. An interpreter *first* finds an over-all

[2] "The Present Status of Automatic Translation of Languages," in F. L. Alt, ed., *Advances in Computers* (New York: Academic Press, 1964), vol. 1, pp. 158/9. Quoted in H. L. Dreyfus, *What Computers Can't Do*, 2nd ed. (New York: Harper & Row, 1979), p. 215.

scheme that "works" and *then* can interpret each new utterance separately as it comes. For example, once a holistic chess-player interpretation has been worked out, its holism can be ignored—moves can perfectly well be translated "in isolation." [3] By contrast, common-sense holism is *real-time* holism—it is freshly relevant to each new sentence, and it can never be ignored. Even if a perfect dictionary and grammar were available, sentences like our three examples would still have to be disambiguated "in real time," by some appeal to common sense.

The point can be put another way. Prior holism is compatible with the (Fregean) ideal of semantic atomism: the meaning of a sentence is determined by the meanings of its meaningful components, plus their mode of composition. This ideal is (nearly) achieved by chess notations, formal logics, and most programming languages; but it is only grossly approximated by English—assuming that "meaning" is what one "grasps" in understanding a sentence, and that words and idioms are the meaningful components.[4] Real-time holism is precisely *in*compatible with semantic atomism: understanding a sentence requires *more* than a grammar and a dictionary—namely, common sense.[5]

The nature of common-sense holism is brought into sharper relief by current efforts to deal with it—those in Artificial Intelligence being the most concentrated and sophisticated. The hard problem, it turns out, is not simply the enormous volume of common knowledge, but rather storing it so that it can be efficiently accessed and used. Obviously, it is quite impractical to check every available fact for possible relevance, every time some question comes up. So the task is to design a system that will quickly home in on genuinely relevant considerations, while ignoring nearly everything else. This is the "memory organization" or "knowledge representation" problèm; what makes it hard is the quixotic way that odd little "facts" turn up as germane.

[3] Cryptography is comparable: code cracking is holistic, but once it succeeds, deciphering goes along on a message-by-message basis.

[4] Hilary Putnam argues that there is more to meaning than what competent speakers understand, but his point is orthogonal to ours ["The Meaning of Meaning," in *Mind, Language and Reality* (New York: Cambridge, 1975)].

[5] It is difficult to say what significance this has (if any) for formal semantics. The most common tactic is to relegate matters of real-time holism to "pragmatics," and apply the semantic theory itself only to idealized "deep structures" [in which ambiguities of sense, pronoun binding, case, mood, scope, etc. are not allowed—thus saving atomism (perhaps)]. A protective quarantine for semantics may or may not work out, but earlier experience with syntax hardly bodes well.

Most contemporary systems employ some variant of the following idea: facts pertaining to the same subject are stored together ("linked") in structured clusters, which are themselves linked in larger structures, according as their subjects are related.[6]

We can think of these clusters as "concepts," so long as we remember that they are much more elaborate and rich than traditional definitions—even "contextual" definitions. For example, the concept for 'monkey' would include not only that they are primates of a certain sort, but also a lot of "incidental" information like where they come from, what they eat, how organ grinders used them, and what the big one at the zoo throws at spectators. It's more like an encyclopedia than a dictionary entry.

Three points will clarify how this is supposed to work. First, much of the specification of each concept lies in its explicit links or "cross references" to other concepts, in an over-all conceptual superstructure. For instance, part of the monkey concept would be an "is-a" link to the primate concept, which has in turn an "is-a" link to the mammal concept, and so on. So, the monkey, rat, and cow concepts can effectively "share" generic information about mammals. Second, entries in a concept can have modalities, like "necessarily," "typically," "occasionally," or even "only when . . . ." The "typically" mode is particularly useful, because it supplies many common-sense "assumptions" or "default assignments." Thus, if monkeys typically like bananas, the system can "assume" that any given monkey will like bananas (pending information to the contrary). Third, concepts often have "spaces" or "open slots" waiting (or demanding) to be "filled up" in stipulated ways. For example, the concept of eating would have spaces for the eater and the eaten, it being stipulated that the eater be animate, and the eaten (typically) be food.

A system based on such concepts copes with common-sense holism as follows. First, a dictionary routine calls the various concepts associated with the words in a given sentence, subject to constraints provided by a syntactical analyzer. Hence, only the information coded in (or closely linked to) these concepts is actually accessed—

[6] See, for example: Marvin Minsky, "A Framework for Representing Knowledge," in Patrick Winston, ed., *The Psychology of Computer Vision* (New York: McGraw-Hill, 1975); Yorick Wilks, "Natural Language Understanding Systems within the AI Paradigm," Stanford AI Memo-237, 1974; Roger Schank and Robert Abelson, "Scripts, Plans, and Knowledge," *International Joint Conference on Artificial Intelligence*, IV (1975); Daniel Bobrow and Terry Winograd, "An Overview of KRL, a Knowledge Representation Language," *Cognitive Science*, I, 1 (1977).

passing over the presumably irrelevant bulk. Then the system applies this information to any ambiguities by looking for a combination of concepts (from the supplied pool) which fit each other's open spaces in all the stipulated ways. So, for Bar-Hillel's example, the system might call four concepts: one each for 'box' and 'is in', and two for 'pen'. The "is in" concept would have two spaces, with the stipulation that what fills the first be smaller than what fills the second. Alerted by this requirement, the system promptly checks the "typical size" information under the other concepts, and correctly eliminates 'fountain pen'. An essentially similar procedure will disambiguate the pronouns in sentences like: "The monkeys ate the bananas because they were hungry" or ". . . because they were ripe" (cf. Wilks, *op. cit.* p. 19).

The other two examples, however, are tougher. Both raincoats and bathtubs typically get wet, so *that* won't decide which was wet when I left my coat in the tub. People opt for the coat, because being wet is an understandable (if eccentric) reason for leaving a coat in a tub, whereas the tub's being wet would be no (sane) reason to leave a coat in it. But where is *this* information to be coded? It hardly seems that concepts for 'raincoat', 'bathtub', or 'is wet', no matter how "encylopedic," would indicate when it's sensible to put a raincoat in a bathtub. This suggests that common sense can be organized only partially according to subject matter. Much of what we recognize as "making sense" is not "about" some topic for which we have a word or idiom, but rather about some (possibly unique) circumstance or episode, which a longer fragment leads us to "visualize." Introspectively, it seems that we imagine ourselves into the case, and then decide from within it what's plausible. Of course, *how* this is done is just the problem.

The ambiguity of 'painted-on pants' is both similar and different. Again, we "imagine" the sort of attire being described; but the correct reading is obviously a metaphor—for 'skin tight', which is both coordinated and appropriately contrasted with the stylishly draped blouse. Most approaches to metaphor, however, assume that metaphorical readings aren't attempted unless there is something "anomalous" about the "literal" reading (as in "He is the cream on my peaches," or ". . . faster than greased lightning"). But, in this case there is nothing anomalous about pants with paint on them—they would even clash with "stylish," explaining the conjunction "Though. . . ." On that reading, however, the sentence would be silly, whereas the metaphor is so apt that most people don't even notice the alternative.

These examples are meant only to illustrate the subtlety of common sense. They show that no obvious or crude representation will capture it, and suggest that a sophisticated, cross-referenced "encyclopedia" may not suffice either. On the other hand, they don't reveal much about what's "left out," nor (by the same token) whether that will be programmable when we know what it is. The real nature of common sense is still a wide-open question.

III. SITUATION HOLISM

Correct understanding of a sentence depends not only on general common sense, but also on understanding the specific situation(s) to which it pertains. I don't have in mind the familiar point about descriptions and indexicals, that only the "context" determines *which* table is "the table . . ." or "this table . . . ," and so on. Much more interesting is the situation-dependence of examples like Bar-Hillel's; Dreyfus (*op. cit.*) points out that

> . . . in spite of our *general* knowledge about the relative sizes of pens and boxes, we might interpret "The box is in the pen," when whispered in a James Bond movie, as meaning just the opposite of what it means at home or on the farm (216).

This is not just a problem about "exotic" contexts, where normal expectations might fail; both of the following are "normal":

> When Daddy came home, the boys stopped their cowboy game. They put away their guns and ran out back to the car.

> When the police drove up, the boys called off their robbery attempt. They put away their guns and ran out back to the car.

The second sentence is not exactly ambiguous, but it means different things in the two situations. Did they, for instance, put their guns "away" in a toy chest or in their pockets? (It makes a difference in German: *einräumen* or *einstecken*.) Could 'ran' be paraphrased by 'fled'?

So far, the role of "situation sense" seems comparable to that of common sense, though more local and specific. A fundamental difference appears, however, as soon as the stories get interesting enough to involve an interplay of several situations. A Middle-Eastern folk tale gives a brief example:

> One evening, Khoja looked down into a well, and was startled to find the moon shining up at him. It won't help anyone down there, he thought, and he quickly fetched a hook on a rope. But when he threw it in, the hook snagged on a hidden rock. Khoja pulled and pulled and pulled. Then suddenly it broke loose, and he went right

> on his back with a thump. From where he lay, however, he could see the moon, finally back where it belonged—and he was proud of the good job he had done.

The heart of this story is a trade-off between two situations: the real one and the one in Khoja's imagination. The narrative jumps back and forth between them; and it is up to the reader to keep them straight, and also to keep track of their interaction and development.

In the first sentence, for example, the embedded clauses "Khoja found the moon" and "it shined up at him," are clearly about the epistemic situation, despite their grammar. One must understand this at the outset, to appreciate Khoja's progressive misperceptions, and thus his eventual pride. A trickier shift occurs in the clause "It won't help anyone down there . . . ," which must mean "*if it stays* down there" (not: "anyone *who is* down there"). In other words, it's an implicit hypothetical which refers us to yet another situation: a counterfactual one in which people are left in darkness while the moon is still in the well. This too is essential to understanding the pride.[7]

The important point is how little of this is explicit in the text: the clauses as written exhibit what can be called "situational ambiguity." It's as if situations were "modalizers" for the expressed clauses, generating "mini-possible-worlds" and implicit propositional operators. I'm not seriously proposing a model theory (though, of course, this has been done for counterfactuals, deontic modalities, and epistemic states) but only suggesting what may be a helpful analogy. Thus the clause "Khoja found the moon" would have not only the modality "Khoja thought that . . ." but also the modality "while looking into the well. . . ." The latter is a crucial modalization, for it (along with common sense) is what forces the former.

Given this way of putting it, two things stand out. First, rather than a fixed, lexically specified set of possible modalities, there are indefinitely many of them, more or less like sentences (or indeed, whole passages). Second, many of these have to be supplied (or inferred) by the reader—often, as in the last example, on the basis of others already supplied. That is, to understand the text, the reader must provide for each clause a number of these generalized

[7] There are also a number of "background counterfactuals" involved in understanding what happens. Thus, a reader should be able to say what would have happened if the hook hadn't caught on the rock, or if it hadn't broken loose. Anyone who couldn't answer, wouldn't really "have" it.

or "situational" modalities, and must do so largely on the basis of some over-all situational or modal coherence. This demand for over-all coherence—that all the various "situations" (with respect to which clauses are understood) should fit together in an intelligible way—is what I call *situation holism*. It is a general feature of natural-language text, and coping with it is prerequisite to reading.

Situation holism is especially characteristic of longer texts. We had a brief sample in our folk tale; but it really comes into its own in the forms of dialectic, characterization, and plot. Mystery novels, for example, are built around the challenge of situation holism when pivotal cues are deliberately scattered and ambiguous. Translators (who read the book first, naturally) must be very sensitive to such matters—to use 'ran' or 'flew' instead of 'fled', for instance—on pain of spoiling the suspense. But only the over-all plot determines just which words need to be handled carefully, not to mention how to handle them. Engrossed readers, of course, are alert to the same issues in a complementary way. This is situation holism, full-fledged.[8]

IV. DIGRESSION: HERMENEUTICS

Hermeneutics, in the classical (nineteenth-century) sense, is the "science" of textual interpretation—i.e., exegesis. It is often described as "holistic," on something like the following grounds: the meanings of particular passages, doctrines, and specialized ("technical") terms, are only apparent in the context of the whole; yet the whole (treatise, life's work, or genre) is composed entirely of particular passages, containing the various doctrines and special terms. So the interpreter must work back and forth among part, subpart, and whole, bootstrapping each insight on one level into new insights on the others, until a satisfactory over-all understanding is achieved.

Hermeneutics is like intentional interpretation, insofar as the point is to translate baffling expressions into others more familiar or more intelligible. And the constraint on adequacy is again that the text, as construed, make a maximum of sense. But in exegesis, "sensibleness" is not so easy to determine as it is, say, in translating chess notations. For each sentence will have various presuppositions

[8] In AI, work on this problem has only just begun. See, e.g. David Rumelhart, "Notes on a Schema for Stories," in Bobrow and Allan Collins, eds., *Representation and Understanding* (New York: Academic Press, 1975); Bob Wilensky, "Why John Married Mary: Understanding Stories Involving Recurring Goals," *Cognitive Science*, II (1978): 235–266; and Robert de Beaugrande and Benjamin Colby, "Narrative Models of Action and Interaction," *Cognitive Science*, III (1979): 43–66. Compare also David Lewis, "Scorekeeping in a Language Game," forthcoming in *Journal of Philosophical Logic*.

or "facts" taken for granted and will make sense only in the light of these. Part of the interpreter's task, in determining what the text means, is to ferret such assumptions out and make them explicit. So hermeneutic interpretation must deal explicitly with common-sense holism (though it may be "common" only to the initiated few). But the paramount concern in formal exegesis is exposing the over-all structure and purport of the original. A construal cannot stand unless it renders sensible the progression and development of arguments, examples, incidents, and the like. But this is just situation holism, made more articulate. Thus, I don't think the holism of classical hermeneutics is different from the three kinds so far discussed, but is instead a sophisticated combination of them all.[9]

V. EXISTENTIAL HOLISM

In the section on intentional interpretation, we noticed how naturally we construe chess-playing computers as "trying" to make good moves, and "wanting" to win. At the same time, however, I think we *also* all feel that the machines don't "really care" whether they win, or how they play—that somehow the game doesn't "matter" to them. What's behind these conflicting intuitions? It may seem at first that what machines lack is a "reason" to win: some larger goal that winning would subserve. But this only puts off the problem; for we then ask whether they "really care" about the larger goal. And until this question is answered, nothing has been; just as we now don't suppose pawns "matter" to computers, even though they subserve the larger goal of winning.

Apparently something else must be involved to make the whole hierarchy of goals worth while—something that itself doesn't need a reason, but, so to speak, "matters for its own sake." We get a hint of what this might be, by asking why chess games matter to people (when they do). There are many variations, of course, but here are some typical reasons:

(i) public recognition and esteem, which generates and supports self-esteem (compare the loser's embarrassment or loss of face);
(ii) pride and self-respect at some difficult achievement—like finally earning a "master" rating (compare the loser's frustration and self-disappointment); or

[9] It can be argued (though not here) that genuine radical translation is less like the interpretation of a chess player than like a hermeneutic investigation of a whole culture—including (so far as possible) an "interpretation" of its practices, institutions, and artifacts. For a good account of what hermeneutics has become in the twentieth century (very roughly, it adds my fourth holism), see Charles Taylor, "Interpretation and the Sciences of Man," *Review of Metaphysics*, XXV, 1 (September 1971): 3–51.

(iii) proving one's prowess or (as it were) "masculinity" (compare the loser's self-doubt and fear of inadequacy).

What these have in common is that the player's self-image or sense of identity is at stake. This concern with "who one is" constitutes at least one issue that "matters for its own sake." Machines (at present) lack any personality and, hence, any possibility of personal involvement; so (on these grounds) nothing can really matter to them.[10]

The point is more consequential for language understanding than for formal activities like chess playing, which are largely separable from the rest of life. A friend of mine tells a story about the time she kept a white rat as a pet. It was usually tame enough to follow at her heels around campus; but one day, frightened by a dog, it ran so far up her pantleg that any movement might have crushed it. So, very sheepishly, she let down her jeans, pulled out her quivering rodent, and won a round of applause from delighted passersby. Now, most people find this anecdote amusing, and the relevant question is: Why? Much of it, surely, is that we identify with the young heroine and share in her embarrassment—being relieved, at the same time, that it didn't happen to us.

Embarrassment, however, (and relief) can be experienced only by a being that has some sense of itself—a sense that is important to it and can be awkwardly compromised on occasion. Hence, only such a being could, as we do, find this story amusing. It might be argued, however, that "emotional" reactions, like embarrassment and bemusement, should be sharply distinguished from purely "cognitive" understanding. Nobody, after all, expects a mechanical chess player to *like* the game or to be thrilled by it. But that distinction cannot be maintained for users of natural language. Translators, for instance, must choose words carefully to retain the character of an amusing original. To take just one example from the preceding story, German has several "equivalents" for 'sheepish', with connotations, respectively, of being simple, stupid, or bashful. Only by appreciating the embarrassing nature of the incident, could a translator make the right choice.

A different perspective is illustrated by the time Ralph asked his new friend, Lucifer: "Why, when you're so brilliant, beautiful, and

[10] There are many problems in this vicinity. For instance, people (but not machines) play chess for *fun*; and, within limits, winning is more fun. It's very hard, however, to say what fun is, or get any grip on what it would be for a machine actually to *have* fun. One might try to connect it with the foregoing, and say (in a tired European tone of voice) that fun is merely a temporary diversion from the ever-oppressive burden of self-understanding. But that isn't very persuasive.

everything, did you ever get kicked out of heaven?" Rather than answer right away, Lucifer suggested a little game: "I'll sit up here on this rock," he said, "and you just carry on with all that wonderful praise you were giving me." Well, Ralph went along, but as the hours passed, it began to get boring; so, finally, he said: "Look, why don't we add some variety to this game, say, by taking turns?" "Ahh," Lucifer sighed, "that's all I said, that's all I said."

Here, even more than Ralph's embarrassment, we enjoy the adroit way that Lucifer turns the crime of the ages into a little *faux pas*, blown out of proportion by God's infinite vanity. But why is that funny? Part of it has to be that we all know what guilt and shame are like, and how we try to escape them with impossible rationalizations—this being a grand case on both counts. It's not the *psychology* of guilt that we "know," but the tension of actually *facing* it and (sometimes) trying not to face it. And actually "feeling" guilty is certainly not just a cognitive state, like believing you did wrong, and disapproving; nor is it that, with some unpleasant sensation added on. It is at least to sense oneself as diminished by one's act—to be reduced in worth or exposed as less worthy than had seemed.

Crime and Punishment, too, is "about" guilt, but it isn't especially funny. The novel is powerful and didactic: the reader's experience of guilt is not simply drawn upon, but engaged and challenged. We enter into Raskolnikov's (and Dostoyevsky's) struggle with the very natures of guilt, personal responsibility, and freedom —and in so doing, we grow as persons. This response, too, is a kind of understanding, and asking questions is a fairly effective test for it. Moreover, at least some of those questions will have to be answered in the course of producing an adequate translation.

One final example will demonstrate the range of the phenomenon I'm pointing at, and also illustrate a different way in which the reader's personal involvement can be essential. It is a fable of Aesop's.

> One day, a farmer's son accidentally stepped on a snake, and was fatally bitten. Enraged, the father chased the snake with an axe, and managed to cut off its tail. Whereupon, the snake nearly ruined the farm by biting all the animals. Well, the farmer thought it over, and finally took the snake some sweetmeats, and said: "I can understand your anger, and surely you can understand mine. But now that we are even, let's forget and be friends again." "No, no," said the snake, "take away your gifts. You can never forget your dead son, nor I my missing tail."

Obviously, this story has a "moral," which a reader must "get" in order to understand it.

The problem is not simply to make the moral explicit, for then it would be more direct and effective to substitute a non-allegorical paraphrase:

> A child is like a part of oneself, such as a limb. The similarities include:
> (i) losing one is very bad;
> (ii) if you lose one, you can never get it back;
> (iii) they have no adequate substitutes; and thus
> (iv) they are literally priceless.
> Therefore, to regard trading losses of them as a "fair exchange," or "getting even," is to be a fool.

But this is just a list of platitudes. It's not that it misrepresents the moral, but that it lacks it altogether—it is utterly flat and lifeless. By comparison, Aesop's version "lives," because we as readers identify with the farmer. Hence, we too are brought up short by the serpent's rebuke, and that makes us look at ourselves.

The terrifying thing about losing, say, one's legs is not the event itself, or the pain, but rather the thought of *being* a legless cripple for all the rest of one's life. It's the same with losing a son, right? Wrong! Many a parent indeed would joyously give both legs to have back a little girl or boy who is gone. Children can well mean more to who one is than even one's own limbs. So who are you, and what is your life? The folly—what the fable is really "about" —is not knowing.[11]

A single event cannot be embarrassing, shameful, irresponsible, or foolish in isolation, but only as an act in the biography of a whole, historical individual—a person whose personality it reflects and whose self-image it threatens. Only a being that cares about who it is, as some sort of enduring whole, can care about guilt or folly, self-respect or achievement, life or death. And only such a being can read. This holism, now not even apparently in the text, but manifestly in the reader, I call (with all due trepidation) *existential holism*. It is essential, I submit, to understanding the meaning of any text that (in a familiar sense) *has* any meaning. If situation holism is the foundation of plot, existential holism is the foundation of literature.

In the context of Artificial Intelligence, however, there remains

[11] Rumelhart (*op. cit.*) analyzes a different version of this story in terms of an interesting "story grammar," loosely analogous to sentential grammar. Significantly, however, he addresses only the continuity of the story and never touches on its moral or meaning.

an important question of whether this sets the standard too high—whether it falls into what Papert somewhere calls "the human/superhuman fallacy," or Dennett "the Einstein-Shakespeare gambit." Wouldn't it be impressive enough, the reasoning goes, if a machine could understand everyday English, even if it couldn't appreciate literature? Sure, it would be impressive; but beyond that there are three replies. First, if we could articulate some ceiling of "ordinariness" beyond which machines can't pass or can't pass unless they meet some further special condition, that would be very interesting and valuable indeed. Second, millions of people can read—really read—and for most of the others it's presumably a socio-historical tragedy that they can't. Existential holism is not a condition just on creative genius. Finally, and *most important*, there is no reason whatsoever to believe there is a difference in kind between understanding "everyday English" and appreciating literature. Apart from a few highly restricted domains, like playing chess, analyzing mass spectra, or making airline reservations, the most ordinary conversations are fraught with life and all its meanings.

VI

Considering the progress and prospects of Artificial Intelligence can be a peculiarly concrete and powerful way of thinking about our own spiritual nature. As such, it is a comrade of the philosophy of mind (some authors see AI as allied to epistemology, which strikes me as perverse). Here, we have distinguished four phenomena, each with a claim to the title 'holism'—not to trade on or enhance any mystery in the term, but rather, I would hope, the opposite. The aim has not been to show that Artificial Intelligence is impossible (though it is, you know) but to clarify some of what its achievement would involve, in the specific area of language understanding. This area is not so limited as it seems, since—as each of the four holisms testifies—understanding a text involves understanding what the text is "about." The holisms, as presented, increase in difficulty relative to current AI techniques; and my own inclination (it's hardly more than that) is to regard the last, existential holism, as the most fundamental of the four. Hence my opening remark: the trouble with Artificial Intelligence is that computers don't give a damn.

JOHN HAUGELAND

University of Pittsburgh