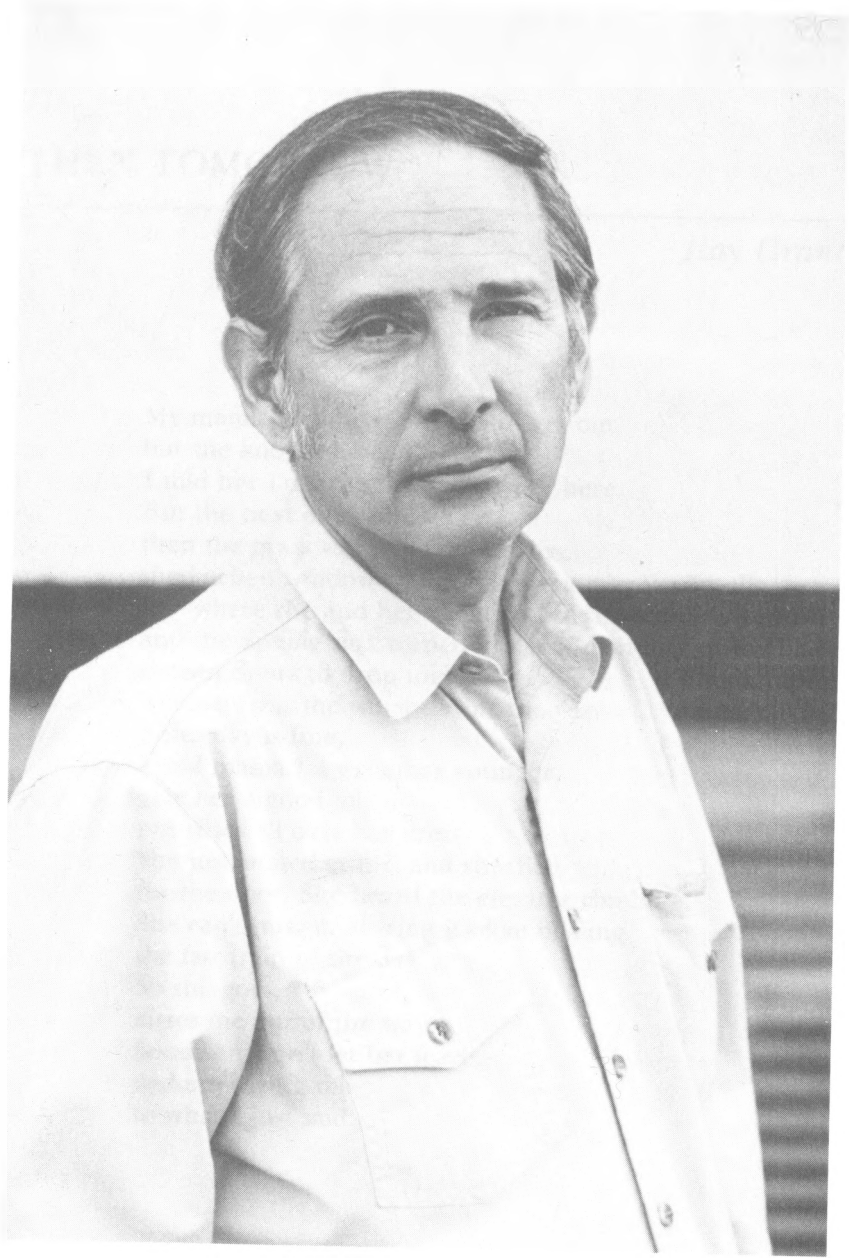

John Searle. Photograph by Anne Lattimore.

INTERVIEW WITH JOHN SEARLE

Bruce Krajewski

JOHN SEARLE IS a Professor of Philosophy at the University of California, Berkeley. He is the author of *Speech Acts* (1969), *The Campus War* (1971), *Expression and Meaning: Studies in the Theory of Speech Acts* (1979), *Intentionality: An Essay in the Philosophy of Mind* (1983), and *Minds, Brains and Science* (1984). Also, he edited an important collection of essays entitled *The Philosophy of Language* (1971).

Many students of literature know Searle through his "debate" about a/the philosophy of language with Jacques Derrida in the journal *Glyph*, a debate which carried over into the *New York Review of Books*.

Searle came to The University of Iowa in April, 1987 to give a talk based on the lectures in *Minds, Brains and Science*. In those lectures, Searle presents his refutation of arguments put forth by proponents of artificial intelligence (AI). The following interview was conducted during Searle's visit.

I want to ask you about some things you said in your talk last night about the refutation of the "strong" AI people, and about how you made a syllogistic-like argument regarding why it is a foolish idea to think that the mind is just a computer program. Even though this argument is rather silly, and easily refuted, as you say, these people are still receiving large grants, and they have, relatively speaking, some political power. Why doesn't the refutation of the argument prevent them from carrying on? This raises questions about the power of these people, and about why your story of the mind isn't as persuasive as theirs, or not persuasive to the "right" people.

I think, in the end, it will be persuasive. I think, in the end, rationality prevails. In the short run, there is too much invested in the contrary view. There are a lot of careers invested in strong AI, and a lot of reputations, and a lot of prestige, and a lot of money, and a lot of

institutional backing. If you spend thirty years betting on the thesis that by creating programs you are creating minds, and somebody comes along and refutes that thesis, you're not going to admit the refutation immediately. And what I think happens in all these cases, whenever you get a paradigm shift in the sciences, whenever a view has been refuted, it really takes time for a new generation of people to come in before the effect of the refutation takes hold. So I don't have much trouble with the younger generation of people in AI; I think most of them agree with me. They think the correct research program for AI is what I call "weak AI." It's the older generation that finds it harder to admit that. Basically, they will probably never admit it; they will simply retire. The way that views triumph in the long run is not by getting the establishment to accept the views, but by winning over a younger generation that just replaces the older establishment. That's what is likely to happen. Thirty years from now, I doubt that anyone will believe in strong AI, but that won't be because the old guard got converted. It will be because they retired or died. That's how it generally works. That's how Chomsky, for example, won in linguistics. It isn't that he converted the older generation of American structuralists; he converted their graduate students. That's what counts.

That's why I want to talk about stories, and how persuasive stories are, because I thought that was important to your talk last night. It's not so much whether the argument is sound or not, but whether you can be persuasive with your story.

That's partly right. It is important to have formulae that are accessible to an educated, but largely non-specialist public. The appealing thing about the presentation that I made on this issue is the parable of the Chinese room.[1] That story is something that anybody can understand. If I had just presented it in the form of an axiomatic derivation from three premises, and suppose I had formalized that in the predicate calculus—it would not have had anything like the impact that it did. Every freshman can understand the parable of the Chinese room. However, the contrary danger is that just as you can get an appealing parable that makes the point, so the opposition can equally get slogans and formulae that enable them to try to answer your objection or to evade the consequences of the argument. In the end, when you come to the nitty gritty, it's the logical power of the argument that prevails, but that takes a long time. In intellectual life rationality prevails, but it takes a while.

What got you interested in computers?

I'm not especially interested in computers. I love them; I love working with them, but the theory of computation doesn't interest me especially. I am more interested in AI than I am in computers because I am interested in how the mind works. Anything that is relevant to how the mind works, any new theory about how the mind works, I find very exciting. I got into this debate with strong AI because of my activities in cognitive science. I'm a member of the Cognitive Science Program in Berkeley. There are people who claim this is the way to do cognitive science, by designing computer programs. In fact, there is now a rather exciting development in programming called "parallel distributed processing," or sometimes called "connectionism," where you have a whole lot of interconnected networks. That's supposed to give us an account of cognition inspired by a certain conception of how the brain works. I believe connectionism is much more plausible as a model of the brain than the old-fashioned step-by-step linear AI was as a model of how the mind works.

Now on this particular issue of the Chinese room, that occurred to me quite by accident. I had to give some lectures at Yale, and I didn't know anything about what they were doing, so I bought a book and read it on the airplane going to New Haven, and I thought of an obvious objection to strong AI. I thought of the Chinese room example on United Airlines at 30,000 feet between cocktails and dinner.

I thought that because many philosophy departments teach symbolic logic that there might have been some easy correspondence between philosophical logic and computer programming.

There is, but it's not my field. I don't do it. There are actually traditional programming languages that basically use the principles of formal logic, especially the predicate calculus.

I'd also like to find out how you appropriated the philosophical tradition. You seem to be influenced by Austin. You mention him in several places. Stanley Cavell is also someone who has been influenced by Austin, and yet you seem to have taken completely different paths from Cavell. Cavell is someone who writes about films and literature, and he is more likely to be read in the English department than you.

Is that right? I suppose so.

Cavell has an essay about the difference between modern analytic philosophy and what he calls continental philosophy. He sees that there's a definite split. There seems to be some sort of animosity or distrust between people who practice analytic philosophy and, say, existentialism or phenomenology. Do you see this same split, and if so, why does it exist?

I don't really see it in that way. I think it is a kind of vulgarization of intellectual life in general, and philosophy in particular, to think that the important questions are things like "Whose team is he on? Who's he playing for? What philosophical party does he belong to?" It is an outsider's or a journalist's view to suppose that there are these different philosophical teams, and you play on the analytic team or you play for the phenomenology team. But basically I do think that's vulgar and journalistic. In my own experience, I find I work on problems, and you might be surprised to discover where I get help with philosophical problems. Last year, I was invited by Jürgen Habermas, who is a professor of philosophy at Frankfurt, to come and teach with him in Frankfurt. I went; and Habermas and I and Habermas's colleague, Karl-Otto Apel, taught a seminar, the three of us together. The "seminar," by the way, was not a small group; it had about a hundred people. If somebody had told me in the middle of this, well, these guys are continental philosophers and you're an analytic philosopher, I would have said, "So what?" I mean, we have philosophical problems, and we are working on them.

Now, having said that, I have to add that there are, of course, differences in style. A French professor who I liked a lot, who used to come to Berkeley, and who I used to talk to a lot, was Michel Foucault. I don't know whether Foucault was a "philosopher" or not, but he certainly was fun to talk to. I think he did suffer from the fact that he didn't have the kind of conceptual apparatus as part of his taken-for-granted intellectual equipment that he would have had, had we gotten hold of him at an earlier age. He used to feel a little—I don't know quite what the right word is—envious almost, that American philosophers could take for granted certain standards of clarity and rigor, which, in the intellectual community that he was operating in, he couldn't. I once asked him, "Michel, how come you write such strange prose? You don't have to write like that. You don't talk like that." And he said that that was really required by the French intellectual environment. He once said to me, "If I wrote as clearly as you do, people in France wouldn't take me seriously, because they think that if they can understand everything, it must be superficial." And he wasn't joking. This was a very serious conversation we were having about different standards of clarity. I believe that Foucault's work got

much clearer as he got older and became more confident in his own intellectual distinction.

In response to your question, I would say, of course, there are different traditions, styles, ways of doing philosophy. What you will find, however, is that the really deep issues in philosophy cut across those distinctions. So I am much closer in outlook to Jürgen Habermas than I am to Quine, even though Quine and I were brought up in the same philosophical tradition. I am closer to Foucault, in many ways, than I am to Davidson, even though Davidson is a close colleague of mine, and I've known him for thirty years. So there are traditions and styles, but the deep issues in philosophy cut right across those traditions. The deep issues in philosophy have to do with such things as the role of truth in representation, the role of truth in the analysis of semantics, and that issue is neutral between analytic and continental philosophy.

Do you see it as also being some form of political argument to make this distinction between the continental and the Anglo-American analytic tradition? For instance, if you associate yourself with Habermas, some people would automatically assume that you must have Marxist political views. Is the philosophy separate from the politics?

I've never been able to take the "Marxism" in the Frankfurt school very seriously. By "the Frankfurt School," I mean the contemporary Frankfurt school, primarily Habermas and Apel. You see, in Berkeley, and in my upbringing, I knew real Marxists. By the standards of really mean Marxists, Habermas and Apel aren't Marxists. A Marxist, first of all, is somebody who wants to kill a very large number of people; and basically, Habermas and Apel don't want to hurt anybody. The sense in which they are Marxists is not at all the sense in which, by "Marxism," we mean the revolutionary tradition that goes through Marx, Lenin, Stalin, and the various forms of revolutionary movements that we know today. What they mean by "Marxism" is, roughly speaking, the idea that there are certain categories of economic analysis that they find useful in doing philosophy. To put it in one sentence: They are just not violent revolutionaries. They're not even close to being any kind of violent revolutionaries.

All of this relates to the issue of clarity and how philosophy is to be written or talked about in, say, simple language. You often use phrases in the "Minds, Brains and Science" lecture like, "Things are simply this," or "This point is clear and decisive," or "This demonstrates an obvious point." You like clarity, and you think things should be clear. Do you think that this sort of streamlined

language can do justice to the complex issues that you're dealing with? It seems that this is an argument that continental philosophers would make: Because things are much more complex, they cannot be put into simple language.

That particular book consists entirely of broadcast lectures for the BBC. I had to make the material completely clear to a lay audience. But in general there is no reason why complex issues can't be stated clearly. In general, there is a rough law I have discovered, and it is that if the author can't state his view clearly, he doesn't understand it himself. There are a lot of points in philosophy and elsewhere that are extremely difficult and complex, and if they're going to be stated precisely, they have to be stated in a way that doesn't disguise their difficulty and complexity. But complexity shouldn't be confused with obscurantism—obscurity for the sake of obscurity, or, worse yet, simple self-indulgence. Also, many obscure and complex arguments have rather simple consequences. Let me give you an example. Gödel's theorem is a very complex proof, but the implications of the proof can be stated quite simply. It means there are sentences in systems of the *Principia Mathematica* type which are true but not provable within the system. That's an enormous consequence for the philosophy of mathematics. That's a rather simple consequence, but the actual proof that states the derivation is very complex. Complexity is one thing and obscurantism another. My fight is not with complexity, but with obscurantism. Now, the other point that I was trying to make is that there are a lot of complex issues that have a simple overall structure, even though the details are complex. You can often state in a way that is accessible to any intelligent person what the overall issue is, even though you leave out a lot of the complexity of the detail. That's what I tried to do in *Minds, Brains and Science*. In addition I wanted to try to overcome the fact that many people are intimidated by what seems like inaccessible professional expertise. They think they can't argue with AI or cognitive science because they are not experts.

This is a change of topic here about the concept of mind, and how the mind is tied to your philosophy of language. It seemed to change for me from reading Speech Acts, *where you seem to sound a lot like Wittgenstein, and then in the book on minds, brains, and science. There are things you say about intention that, it seems to me, Wittgenstein would argue about. I just picked out an example from Wittgenstein's writings to get your reaction to it and to find out whether you would agree with this or not. Wittgenstein is talking about*

intentions, or internal states, or things that go on in a person's mind, and he says:

> "I'd like to know what he's thinking of." But now ask yourself this—apparently irrelevant—question. "Why does what is going on in him, in his mind, interest me at all, supposing that something is going on?" (The devil take what's going on inside him!) We judge the motives of an act by what its author tells us, by the report of eyewitnesses, and by the preceding history.

All of this has to be externalized for it to be real. So how do you see your view of the mind either agreeing or disagreeing with this?

I think that, at bottom, I'm probably very much in disagreement with the behaviorist strand that you find in Wittgenstein. Now it would be too crude to describe Wittgenstein simply as a "behaviorist." That's too simple but Wittgenstein is constantly anxious to emphasize the public character of our mental concepts, what he calls the way in which "an 'inner process' stands in need of outward criteria." Remember, that's a quotation from the *Investigations*. I find all of that suggestive, and certainly he's right in most of his observations, but the thrust of many of his observations about the mind seems to me profoundly misleading. It suggests something that I think he felt, and that a lot of people have taken him to suggest, namely, that what's important about the mind is its public, behavioral manifestations, and I think that's not right. Mental phenomena have a first-person ontology. They only exist from the point of view of the person whose mental phenomena they are, and all mental phenomena, if not conscious, are at least potentially conscious. So consciousness, and hence subjectivity, are primary to mental phenomena. Wittgenstein, I think, neglects that. Now he was reacting against a tradition that tried to make the mind something private and mysterious and Cartesian and ineffable. Naturally, he tries to state the case against that as strongly as he can. But we have now had fifty years or more of the anti-Cartesian, behaviorist tradition, and it's time to put a halt to it. That is, it's time to remind ourselves that what counts about mental phenomena is that they have mental features. They have such features as consciousness, subjectivity, intentionality, and rationality. And they function causally. Our own mental states are accessible to us in a way that they are not accessible to other people, and all of these facts have to be integrated into our overall scientific worldview. That's what I was trying to do in my book *Minds, Brains and Science*. I think, though it's always difficult to generalize about Wittgenstein, that the direction of Wittgenstein's philosophy of mind is, in the end, not the same as the direction of mine.

Now, you mentioned *Speech Acts*. I think, in fact, the direction of *Speech Acts* is also counter to the way Wittgenstein wanted to do the

philosophy of language. I want a theory, and Wittgenstein was anti-theoretical. I think Wittgenstein would have been opposed to having a theory of speech acts. He would have said it's impossible to get a general theory of speech acts of the sort that I have. For example, I think that you can classify different types of speech acts; and Wittgenstein would have said that you can't do that, because there are countless different kinds. For Wittgenstein there is no way to classify or taxonomize types of speech acts. So, I think, at bottom, I'm opposed to many themes in Wittgenstein, though he has been enormously influential in my work.

Wittgenstein is more popular with people who study literature. For instance, Gadamer says he has been influenced by Wittgenstein, and many people in literary theory study Bakhtin, who says that you really don't have a private language. The language that you learn is given to you by society, and that any sort of speech act you would make would reflect, say, your social class, for instance, or your background, who brought you up—those sorts of things. So that you can never be an individual entity.

If you don't overstate that point, I think I would agree with most of it. That is, I certainly think that language is public, and that we learn it in particular social situations, and words have publicly accessible meanings, and we speak to each other in a publicly understandable vocabulary. But I wouldn't go the Marxist route and say that then we can never transcend our class background. I think that's rather silly.

One more question about this. One of the people I studied with here is Gerald Bruns, and he was upset by people who talked about the mind. He believes that the mind is an Enlightenment invention. That it really doesn't exist. People invented the mind in order to do analytic philosophy, for instance.

That's an odd thing to say, because the term "mind" has been around long before anybody ever thought of analytic philosophy. Did he really think that it was invented by analytic philosophers?

He attributed it to Descartes, and I want to present this quotation by Bruns to you, to see what you think of this view, because the notion of the mind causes all sorts of problems which we can get around by turning things to the social rather than to the mind. Bruns writes:

> The way you characterize a given phenomenon determines the way in which you will go on to study it. If, for example, you characterize interpretation as a mental act, you will be constrained to practice a form of epistemology whether by constructing a theory of interpretation on the model of a theory of knowledge or by practicing epistemological skepticism. If, however, you characterize interpreta-

> tion as something that goes on in the world, that is, something that human beings regularly do under a variety of complicated circumstances, and to accomplish certain things that need to get done, then you will be constrained to study the history of such goings on.

Bruns is trying to characterize interpretation, or how we understand things, as a social act or custom rather than a mental process.

I guess I don't see the mutually exclusive character of the mental and the social. It seems to me both are real and interconnected. When you interpret something, you have to *think* about it, but, of course, interpretation goes on normally in *social* communities involving exchange of ideas between members of that community. I guess what I would want to do is reject the either/or implication, the implication that either interpretation is a private mental act, or it's just a lot of public noises. Of course, it's public and, of course, it has to be an expression of thinking. We have inherited from Descartes the illusion that there are these two realms, the mental and the physical. But I want to say once we recognize mental phenomena as part of the world we all live in, then they cease to be terrifying; then they cease to be mysterious. We can recognize that things like going on a picnic, or teaching a course, or buying a used car, are all both "mental" and "physical." What that tells you is that the terminology was the wrong terminology from the beginning. We are slaves to the terminology here, and I'm stuck with the terminology as much as anybody else. If you use the expression "the mind"—it sounds like you're naming a thing, or an entity, or an arena. But when we use the expression "mind" we are in fact just talking about human beings at a certain level of operation; and we are talking about certain features that human beings have, such as consciousness, that are absolutely crucial for their functioning as human beings.

The mind includes things that aren't easily explained, or that can't be explained through the physical?

I don't know why not. What can't be explained? I don't think the mind is mysterious or ineffable. There are a lot of things we don't know the answers to, but if somebody says you can't get a theory of the mind, they are wrong. I just wrote part of such a theory. I wrote a 250-page book (*Intentionality*) about how the mind works. Of course, I only scratched the surface. There's much much more to be said. The problem with the quotation that you gave is not that it's wrong, but that it tacitly accepts the dualism that it attempts to be militating against. It accepts the idea that we have to choose between a public and a private conception of interpretation. We don't. Interpretation

is, of course, a matter of thinking, which is a conscious mental process, and, of course, it is also a social phenomenon. The mistake is to think that somehow those are in conflict. They are not.

Bruns gets around to the conflict later. The conflict is about whether the mind leads you to epistemology, and then you have to have methodologies in order to prove how the mind works. Then you get, at least as far as interpretation goes, different camps of how you understand a text. You have a Marxist reading, a psychoanalytic reading, and a hermeneutic reading. People choose camps, and that's where the conflict seems to begin. I think that's what Bruns was trying to get to there. People do choose, and this is somehow wrong, because what they're doing is choosing a methodology rather than trying to understand the complexity of the problem; it exceeds the methodology.

I'm sure you are right about that. These methods always look incredibly superficial when you see them described. Reader-response theory, for example, looks very superficial when you see the description of it. Roughly speaking, the method I use in philosophy, or anything else, is: use any weapon you can lay your hands on. In the middle of the fight, do what you can. When I'm working on the philosophy of mind, I don't say myself, "I'm an analytic philosopher. What does an analytic philosopher do?" I go down and buy every textbook I can find about the brain, or I go and talk to psychiatrists, and find out what they think they're doing. Use any weapon that you can. Some of them will turn out to be useless, and others will be useful. But you can't say in advance what is going to be the right method. The proper method to follow in philosophy, or literary criticism, or anything else, is to use what works.

Let's turn to the mysterious and ineffable. Acts that aren't really acts are difficult to understand, like communicating by doing nothing, or by being silent. I ask you a question, and you just sit there. How am I going to interpret that?

There is a way to answer that. Ask yourself, "Is it intentional or not?" To ask yourself that question is to ask yourself whether there is a certain kind of intention. For example, if you ask me a question, and I decide I'm not going to answer it because I think it's an insulting or a dumb question, and I just sit there, deliberately silent, that silence is itself a kind of intentional speech act. Whereas if you ask me a question, and I don't hear you, then my external behavior might be the same—I might just sit there silently—but in that case it is not a speech act. So the same behavior can be an expression of two

completely different intentional phenomena. That's even true when the behavior is null, that is, when the physical behavior is simply zero. In one case, it might be an intentional non-performance of a speech act, which might itself be a type of speech act, and in the other case, it might simply be the non-performance of an intentional speech act.

Another question you brought up last night was about how you decide or interpret something that seems to be very strange. You were talking about the Chinese room, and how your actions could be interpreted as your understanding Chinese. If you keep giving out the right symbols, someone could say this person obviously understands Chinese. At least the Greeks called this the pseudos, *something that looks real or true, but isn't. This interests people in literary theory, especially with things like the movie* The Return of Martin Guerre, *in which a man assumes someone else's position, actually takes the place of a husband who has disappeared. It is a question of identity, a problem of disguise. So where does the identity come from? Is it from within—that you, as a person giving out the symbols, know that you don't understand Chinese? But what if you kept this up long enough, and people assumed on the outside of the room that you knew Chinese, and then these people wrote books about you after you died, saying that you were the greatest understander of Chinese ever. Wouldn't you then become a real speaker of Chinese?*

Not the way you describe it. Let's go through the steps of the story. Suppose that I'm very well programmed, so that I can continue to give the right Chinese symbols as output in response to the right Chinese input—only I'm always kept locked in this room, so I can't ever learn the meanings of any of these symbols. Let's suppose that after my death people get excited by my answers to the questions, and they discover that they have wonderful stylistic features, and that they show a deep grasp of Chinese. They attribute all of that to me. But, if so, they have made a mistake. The fact that they all agree on something doesn't make it true. They are still mistaken. The brilliance of undertanding Chinese should have been ascribed to the programmers who programmed me, and not to me. It's perfectly possible that these mistakes might proliferate, and become generally accepted as true, even though they're just plain false. The way you describe it, it would be just plain false to say of me that I understood Chinese. There is a kind of bullshit theory that goes around today that says that if you get enough people to agree on something that's all the truth you need. That's obviously wrong, because a lot of things that people agree on are just plain false.

But if you're dead, you can't come back and correct us.

There are a lot of cases in which the truth may never be known. What were Lee Harvey Oswald's motives? We may never know.

The last thing I want to ask you about is, now that it's ten years after your debate with Derrida, what do you feel the debate accomplished? And do you feel the debate is settled?

I never had a debate with Derrida. What happened was that some people presented me with an article in French by a French professor of philosophy about speech acts and asked me what I thought of it. The professor was Derrida, whose work I had never read before, and I told them that I didn't think the article was very good. They asked me if I would write down what I thought of it, because they wanted to include it in a journal they were starting. I agreed to do that, and over a weekend I wrote up my notes. But that's all I did. It was not a debate; I wrote about ten pages in response to his twenty-five page article, and then to my total amazement, he produced nearly a hundred pages in response to me. If there were going to be a debate, I would want equal time. I found his response to be more of a hysterical outburst, a rather low-level temper tantrum, than a serious piece of philosophical analysis. However, I did later publish a more general piece in the *New York Review of Books* where I tried to give an overall assessment of the intellectual level of so-called deconstruction. (See "The Word Turned Upside Down," *New York Review of Books*, 30 [October 27, 1983].) So, we might think of the *New York Review* piece as part of a "debate," but there never was anything approaching a debate in the original sandwich whereby Derrida had a long article, I was allowed a short reply, and then without warning or consulting me, the editors published a very long reply by him. I was never, by the way, given any opportunity or invitation to reply to his piece.

Perhaps the most interesting thing about the whole incident is the very low intellectual level of the work of Derrida, in particular, and deconstruction, in general. If you think about the things that we have been discussing in the course of the interview—artificial intelligence, analytic philosophy, Habermas, Foucault, Wittgenstein, or Austin—Derrida's work simply cannot be discussed at that level. The only interest that the discussion would have would be to explore the status of a certain sort of pathology—the popularity of "deconstruction"—in contemporary American intellectual life.

NOTE

1 The parable goes like this: Imagine that a bunch of computer programmers have written a program that will enable a computer to simulate the understanding of

Chinese. So, for example, if the computer is given a question in Chinese, it will match the question against its memory, or data base, and produce appropriate answers to the questions in Chinese. Suppose, for the sake of argument, that the computer's answers are as good as those of a native Chinese speaker. Now then, does the computer, on the basis of this, understand Chinese? Does it literally understand Chinese in the way that Chinese speakers understand Chinese? Searle's answer is no, and what he does in the parable is substitute a human being in a closed room for the computer that seems to understand Chinese. If you are the computer and you give the right Chinese answers, that is not enough to *guarantee* that you will understand Chinese.