John R. Searle

Rationality in Action. A Lecture (1)

Summary:

In this lecture, the author criticizes what he calls the classical model of rationality, which derives from Hume's thesis and continues up to recent scholars such as Davidson and Williams. This model is based on six points: (1) rational actions are caused by beliefs and desires; (2) rationality is a matter of following special rules; (3) there is a separate faculty or module for rationality; (4) weakness of will is impossible; (5) there must be a set of primary desires from which all reasoning derives; (6) if we are able to operate with rationality, then the set of primary desires must be consistent. Against this model the author opposes his idea of rationality and action, which is founded on certain basic points: (1) a consistent set of desires is impossible, because one desire cannot be satisfied without frustrating a whole lot of others; (2) there is a gap between rationality and action; (3) human beings, unlike animals, are capable of commitment. In particular, the author develops a thesis according to which free will, consciousness and rationality are mutually implicated.

SEARLE:
I'm going to talk about rationality, but I'm going to discuss it in the general situation in contemporary philosophy. The new millennium is a good excuse to announce a new beginning in philosophy. In fact we are coming out of the obsession with epistemology and skepticism that occupied philosophy from the time of Descartes into the 20th century. The problem in the 17th century was how to give a foundation for knowledge. The existence of knowledge was problematic and now, at the end of the 20th century, if there's a single fact of intellectual life which is dominant, it is that "knowledge grows" daily. It is absurd for philosophers to say: "knowledge is problematicÉ it's only really textÉ it's all perspectivalÉ it's all very much in doubtÉ you can't really know anything". Then the same philosopher gets on an airplane with his computer to give a lecture in Paris and works on his computer on the flight. There's no question that knowledge grows. And we come into a period in philosophy where we can assume the existence of knowledge as non -problematic and then we can do a kind of philosophy that for many 20th century philosophers seemed impossible: namely systematic, theoretical philosophy. Paradoxically Wittgenstein, by taking skepticism seriously and helping to overcome it, created the possibility of a kind of philosophy that Wittgenstein would have hated. Namely, the kind I do: *systematic theoretical philosophy*. And today I'm going to discuss the possibilities of a theory of rationality.

1.
Now, we already have, in our intellectual culture, what I will call "the classical model of rationality" and that's mostly what I'm going to discuss. This model of rationality begins with Aristotle's claim that deliberation is always about means and never about ends. It goes on in Kant's claim that "he who wills the end wills the means", and it finds its most famous expression in Hume's doctrine "reason is and ought to be the slave of the passions". Of course, all these philosophers don't agree about rationality, but there is a common thread that runs through our Western intellectual tradition. The most powerful statement is in David Hume and it receives its most powerful contemporary exposition in mathematical decision theory.

I knew there was a problem with decision theory even when I was an undergraduate and I first heard about it. It seemed to me that it was a consequence of this theory that, if I value, say, one dollar, and I value my life, there must be some odds at which I would be willing to bet my life against one dollar. Well, I wouldn't. I would pick up one dollar off the street, I'm glad to have it, but I won't risk my life for one dollar. And even if I would, I wouldn't risk my child's life for a small and trivial amount of money. And yet it is a consequence of mathematical decision theory that if I value any two things, there must be some odds where if I am rational, I'm willing to bet one against the other. I'm not willing to bet my life or my child's life against one dollar. I had a chance to argue this with several famous decision theorists. And after about half an hour of argument, they would always say to me: "well, you're just plain irrational". I'm not so sure. I think they have a problem with their theory of rationality.
I knew the point had practical applications when, during the Vietnam war, I went to visit a friend of mine who was a high official of the Defense Department in Washington and I tried to argue him out of the policy of bombing North Vietnam. He had a Ph.D. in Mathematical Economics and he just went to his black board and drew the traditional curves of marginal economic analysis and he said: "where these curves intersect, this point here, the marginal utility of resisting is exactly equal to the marginal dis -utility of being bombed, and at that point the enemy has to surrender. If they're rational, they have to surrender. All that we are assuming is that they are rational". At that point I knew we had a serious problem, not only with our theory of rationality but also with its application in practice.
So, I will lay out the basic principles of this classical model of rationality and then begin by criticizing them. And just so that we have a manageable basis for discussion, I have made a list of six principles that form the foundation of the theory of rationality.

2.
The first principle is that actions, where rational, are caused by beliefs and desires, and that you approach the decision -making situation with a set of beliefs and desires: you want to go to Milan, you have beliefs about the possibilities of getting to Milan and on the basis of those you form a desire to buy an airplane ticket to Milan. You reason from your ends and your beliefs to the action, and the action is then motivated by the combination of beliefs and desires.
I emphasize that the sense of causation in question is the ordinary sense of Aristotelian efficient causation. It is the sense of causation where the explosion caused the bridge to collapse, or the earthquake caused the building to fall down. It's just ordinary causation where something makes something else happen and it's absolutely essential to see that in the classical model, if the action is genuinely rational, it's caused in the right way by the beliefs and the desires, and consequently the beliefs and the desires set causively sufficient decisions. That is: given the set of beliefs and desires, the action follows by a kind of causal necessity. That is one of the essential features of the classical model. And it's a consequence of that feature that you have to have some basic desires before rationality can get going. I'll come back to that point.
Another feature of this classical model of rationality is that rationality is a matter of following special rules. There are rules to rationality. And our task as theorists is to try to be able to state these rules. But there are rules that people follow unconsciously. In the same way that, for example, the linguist tries to state the rule that people follow unconsciously when they speak correct English. So the agent needn't know the rules that he or she is following, but our task as theoreticians is to try and make explicit and conscious the rules that the agent follows implicitly and unconsciously.
A third feature of our theory of rationality is that there is a separate faculty or capacity. In the current terminology of cognitive science, we would say there is *a separate module for rationality*. And this view is so important that in many theorists it is regarded as a defining trait of human beings. Aristotle, after all, defined us as "rational animals". We are animals that differ from other animals in that we have a separate faculty, or module, of rationality. That module consists of a set of rules, and those enable us to have our actions caused in the right way by our beliefs and desires.

However, there is a problem for this theory of rationality: some kinds of people have a set of rules and desires, they go through a set of deliberations, they make up their mind to do something and then, when the time comes, they just don't do it. They call it *akrasia* and we call it "weakness of will". And now you can

see why weakness of will must be a terrible problem for the rational theory, because if you have a set of beliefs and desires and you made the decision in the right way, then the causes are all set up and it's like lighting a fuse on a stick of dynamite, the action is supposed to follow by causal necessity. And consequently, if it doesn't follow, then it looks like there must be something wrong with the causes.
Most of the standard authors have an account of rational action that makes weakness of will strictly speaking impossible. So, according to Richard Hare, for example -the famous moral theorist in Oxford -if you really morally want to do something, if you really hold a moral principle that something is the best thing to do, and then if you didn't do that thing, that proves that you really didn't believe that was the best thing to do, you really didn't hold the moral principle. So, Hare eliminates weakness of will by definition: in a case where you thought you had a moral principle that required you to do something, and then you didn't do it, it really shows you didn't have a moral principle. That achieves security at the price of becoming trivial, and so it becomes logically impossible that there should be weakness of will.
Donald Davidson's account of weakness of will is more subtle but it commits exactly the same mistake. Davidson says: "if you decide to do something, on the basis of the conviction that it is the best thing to do all things considered, then if you don't do that thing, that only proves that your decision was not an unconditional decision. It was only a conditional, or *prima facie* condition." Once again it becomes impossible for there to be real cases of weakness of will: cases where you make an unconditioned decision to do something, you decide it the best thing to do and then you don't do it. There's something problematic about that, because weakness of will is not at all unusual. It very often happens to my students: they decide that they're going to write their term paper on Wednesday evening, they really believe that's the best thing to do, they permanently, unconditionally make up their mind to do it and at midnight they find themselves watching television having finished their six pack of beer, and no paper is written. Such a thing very often happens to me. I say: "I will drink no more wine this evening" and then I find I am reaching for another glass of Brunello di Montalcino. And this is not a case where you suddenly lose control of yourself and you grab the wine bottle in a frenzy. On the contrary, it can be done quite gracefully, by saying "could I have another glass of red wine, please? Thank you very much". It's simply the case that weakness of the will is a problem for supporters of this classical theory *and I don't think they have worked out this problem*.

3.
The heart of the classical model is that you have to start with a set of primary desires, where desires are construed to include moral evaluations, goals that you have adapted, a set of commitments that you recognize as binding on you and so on. This is what Hume is usually taken to admit when he says "reason is and ought to be the slave of passions". Unless you have these basic passions, then you would not be able to get rationality going. The classical model is almost a kind of hydraulic model, where you have this pressure of your desires and the question is how to satisfy the desires. Given your primary desires, you form secondary desires, so you have a primary desire to go to Milan, then you form a secondary desire to buy a plane ticket. When you go to the travel agent and you say "I want a plane ticket", they understand that's only a secondary desire, they don't say "well, what are you? A ticket fetishist? What is this passion for tickets?". They understand that's only a means to an end. So, the classical model is: all rationality begins with a set of primary desires and then, on the basis of the primary desires, you reason to your secondary desires and you make decisions as to what you're going to do to satisfy the primary desires.
The most sophisticated version of this is in Bernard Williams, who says that "all reasons must be internal reasons", that is to say there can't be any facts in the world that are not reflected in the agent's motivational set, which could constitute a reason for the agent to act. An agent can only act on the basis of elements in the agent's motivational set. The argument that Williams gives for that is that a reason for an action must be capable of motivating, and therefore of explaining, the action. But there's no way or reason for an action to explain the action, no way to motivate the action, if it isn't already represented in the agent's motivational set. I want you to notice that that seems a long way from Hume but, in fact, Hume's claim that you can't derive an *ought* from an *is*, and Williams's claim that there are no external reasons are at bottom very similar claims. If *ought statements* re -express reasons for actions, then the claim that "you can't derive an *ought* from an *is*" is the claim that you cannot get a reason for an action just out of neutral facts in the world. And Williams's claim that there are no external reasons -that all reasons are internal reasons -amounts to the

claim that there couldn't be any facts in the world that could succeed in motivating you, unless they were already recognized or reflected or represented in your motivational set; unless there is some sound deliberative route from reasoning about the recognition of the facts, together with the elements of your motivational set, that leads to the explanation of the action. So Hume's claim "you can't derive an *ought* from an *is*" and Williams's claim "there are no external reasons", though at first sight they look different, are closely related to each other. Both of them are expressed in Hume's other claim that reason is and ought to be the slave of passions: that rationality is a matter of reasoning how to satisfy your primary desires.
The final feature of the classical model -you find this in many authors -is the claim that if rationality is to have any grip, if we are able to operate with rationality, then the set of primary desires must be consistent. If you had inconsistent preferences or inconsistent desires, then rationality would be impossible. And it is often put in the form that you must have a well -ordered preference schedule in order that you can begin to reason, or that you can begin to make rational decisions; you must know that you prefer a -b to a -c and the preference schedule must be well ordered and transmitted. We could go on, but these six principles will give us enough to talk about.
Now, what I'm going to argue is: *they are all false, or at best they describe only special cases*. And I'm now going to attack all six principles.

4.
Let's start with the first claim, that "rational actions are caused by beliefs and desires". I have to tell what is paradoxical about this. Those actions that are genuinely caused by beliefs and desires, where the beliefs and desires are really causally sufficient for the action, seem to me precisely cases of irrational actions. In fact, far from them being a model of rationality, those are typically cases that lie outside the scope of rationality altogether. That's the case, for example, where a man is addicted to a certain drug: he has a tremendous desire for heroin, he believes that the stuff in front of him is heroin, so he takes the heroin. In that case the belief and the desire genuinely cause the action by a causally sufficient condition. But that's hardly a model of rationality. In the normal case of rationality, you have to consider various options and then decide what you want to do on the basis of the options. In the normal case of voluntary rational decision -making you have a sense of alternative courses of action open to you. You have a set of reasons for action, and on the basis of the set of reasons you have to decide what you are going to do.
For example, within the next few months I will have to decide which of various unattractive candidates I will vote for in an election. But I have to decide what I'm going to do on the basis of a set of reasons that do not set causally sufficient conditions. That is, I do not experience my decision -making as psychologically determined by antecedent causes. Rather there is a gap between my awareness of the reasons for acting and my decision. And there is indeed a series of gaps. There's a gap between the psychological antecedence of the decision and the actual making of the decision. Then there's a second gap between the making of the decision and actually carrying it out. That is, once I've made up my mind what to do, I still have to do it, I can't just sit back and let the action occur. Furthermore, for any more complex series of acts, such as if I'm trying to learn Italian, it seems to me there's a third gap: there's a gap between the initiation of the action and its continuation to completion. Just by starting the action, I do not set causally sufficient conditions for its continuation to completion.
This gap has a traditional name in philosophy: it's called *freedom of the will*. I don't much like the old terminology, but it was driving at the sense of alternative possibilities open to us; and the fact that alternative possibilities are open to us is a reflection of the fact that the antecedent psychological states prior to the action do not set causally sufficient conditions. It's true that I made up my mind to vote for Clinton in the last election. But nothing forced me to, I could have voted for the other candidate or I could have decided not to vote at all. What's the proof of that? One proof of it is that, given a number of reasons for acting, I have to select which one to act on: I can't just sit back and let the reasons determine the actionÉ And notice: I know without observation which reason I acted on. I had four reasons for voting for Clinton: I thought he'd be better for the economy, I thought he had a better foreign policy, I thought he had a much better education than the other candidate, and besides he too went to my old college in Oxford. Of those four reasons, I might act on only one, and I know which one I acted on. I don't just have a sense of alternative possibilities, but when I close off the alternatives by selecting an action, I know without observation which reason I made

effective by acting on that reason. Notice furthermore that it is impossible to get rid of the gap in the actual decision -making situation.
Kant was aware of this. Kant said: “you cannot think away your own freedom, you cannot abandon the presupposition of your own freedom, you’re forced to act under the presupposition of freedom.” For example, you’re in a restaurant and they give you the menu where you have the choice of the *spaghetti alle vongole* and the *vitello alla scaloppina*. And you can’t say: “Well, I am a determinist, *che sarà sarà*, I’ll just wait and see what happens”. Because, and this is the key point, even if you do that, that is only intelligible to you as an exercise of freedom. That is to say, you cannot intelligibly deny your own freedom, because the attempt to deny your own freedom is only intelligible to you as an exercise of the freedom you are trying to deny. So there’s no thinking away of the gap: we are confronted with the gap. Now, I’m going to come back to the gap later, when we talk about some of these aberrations.

5.
Point number two: it seems to me that, in general, rationality is not a matter of following rules, because the constraints of rationality are already internal to the phenomena of language and mind. They are internal to language, to speech acts and to intentional states. You see this very clearly if you look at the attempt to see how rules work in what seems to me the most favorable case, the case of logical deduction. It’s tempting to think that, when we make rational deductions of a logical kind, what we are doing is applying rules of logic. We’re applying such rules as the rule of *modus ponens*. And it’s tempting to think that typical arguments of the form “*p*, and if *p* then *q*, therefore *q*” derive their validity from the fact that they follow the rule of *modus ponens*.
That is a disaster and it is a famous disaster. It leads to what is called the *Lewis Carroll paradox*, after a paradox that Lewis Carroll stated about a hundred years ago. Achilles is talking to the tortoise and Achilles presents a simple deductive argument: Achilles says -this isn’t Carroll’s example but it makes the point -“if it rains tonight the ground will be wet: it will rain tonight therefore the ground will be wet”. And the tortoise says “well, I understand all that part until the *therefore*; I don’t understand how you can get to Ô*therefore* the ground will be wet’”. And at that point Achilles makes a disaster, he says: “what you have to do is recognize that we have the law of *modus ponens* that says Ôwhere you have *p*, and if *p* then *q*, you can derive *q*’”. And the tortoise says, “just write all that down, will you?” So Achilles writes down the rule of *modus ponens: “p*, if *p* then *q*, then *q*” and now you know that the tortoise is going to say: “well, I understand, but what I don’t see is how that entitles you to infer Ôtherefore the ground will be wet’”. And Achilles then says: “but, you idiot, don’t you understand? Whenever you have premises that form *p*, and if *p* then *q*, and you have a rule of the form *p* and *p* then *q* then *q*, then you can infer *q*”. And you know what the tortoise is going to say: “write all that down, will you? Just write all that down”. Well, I’m not going to write it down, because you see it leads to an infinite regress.
What is the way to avoid the infinite regress? Most authors say that the mistake was to treat *modus ponens* as a separate premise. But that’s not the mistake. The mistake is to treat the rule of *modus ponens* as playing any role whatever. That argument I gave you as a piece of standard English is a perfectly valid argument as it stands. “If it rains tonight the ground will be wet, it will rain tonight therefore the ground will be wet” - that’s all you need to say. The rule of *modus ponens* is the representation of an infinite number of inferences that are independently valid. But the inferences don’t derive their validity from the rule of *modus ponens*, indeed it would be accurate to say the rule of *modus ponens* derives its validity from the fact that it expresses a pattern of an infinite number of inferences that are independently valid, but it plays no role in the validity of the inference.
We are blinded to this fact by our very sophistication, because, of course, if we treat those marks as syntactical, proof -theoretical, if we treat them just as uninterpreted symbols, then of course there is a substitute rule. You can program a computer: whenever the computer sees *p* and *pàq* , it writes down *q*. You can give it a program rule in the form of *modus ponens*. But notice: the syntactical proof -theoretical analogue of the substantive semantic content is not itself a semantic content: it’s just a set of rules for operating on formal symbols. But if we’re talking about actual semantic contents, then the original argument I gave you is valid as it stands.

6.
Let's move on to the third principle, which is the idea that there must be a separate faculty of rationality. Again we can see why that's wrong.
If I'm right in thinking that the constraints of rationality are internal to language and mind, that is you couldn't perform speech acts except under the constraints of rationality, because the rules of speech acts are constitutive, they constitute what counts as making an assertion or asking a question, or making a promise. If I am right in thinking that those rules are constitutive, then there couldn't be any separate faculty of rationality independent of the faculties of mind and language, because the constraints of rationality are already built into mind and language. The constraints of rationality are not some further module, some further capacity in the way that vision is a capacity in addition to language: you can have language without vision and vision without language. But you cannot have language and intentionality without constraints of rationality; because they're internal, there couldn't be a separate faculty of rationality.

7.
Fourth: The problem of weakness of will. Now we have the means for solving it. How is weakness of will possible? It's possible because of the gap. It's an obvious manifestation of a gap. In any decision -making situation I have a sense of alternative possibilities open for me. I make up my mind what I'm going to do, but when the actual time comes to act, there are still those other possibilities open. I've decided I'm going to work on my book this evening, but there are all sorts of other attractive possibilities. I can watch television, I can go for a walk, etc. And sometimes I find those appealing, even though I fully recognize that they are not the best thing for me to do. And so cases of weakness of will arise more often than we'd like to admit. And the fact that this seems so puzzling and bizarre, shows us that there's something wrong with the classical model. What exactly is wrong with it?
What's wrong is that it has neglected a gap. Once you see the gap, once you see that the antecedents of the action, however rational, do not set causally sufficient conditions for the action, then you see that weakness of will is a perfectly natural and indeed inevitable phenomenon. One way to show the irony of this is to say that what is often presented as the model of rationality, namely the computer, is not even within the scope of rationality, because it's not capable of irrationality. It has no gap, it has no consciousness. Rationality is about a certain form of human consciousness, and the way human consciousness operates in the gap. The computer does not give you a model of rationality because it has no consciousness and it has no possibility of irrationality, hence it's outside the scope of rationality and irrationality. Rationality is only possible where irrationality is possible. And that is only possible for an agent that operates under the precepts of a position of freedom. That is to say an agent that operates consciously in a gap.

8.
The fifth point -the central point of the classical model -says "there must be a set of primary desires from which all reasoning can start." Williams says: "there couldn't be anything for the agent to reason from, if he did not have a reason as an element of his motivational set".
There's a simple argument to show that that must be wrong: that this would have the consequence that for any decision that I make, unless there is some desire right then and there, at that moment when I have to make the decision, act on the decision; unless there is some desire right then and there that would be satisfied by that decision, then I have no reason at all to make the decision.
You could not apply that to real life. I imagine myself going into a bar and ordering a beer. They bring the beer and I drink it and then they bring the bill, $4, and I say: "look, I have consulted my motivational set and I can find no desire to pay for this beer". I have no primary desire to pay for the beer and there is no other desire I have that would be satisfied by paying for the beer. The point is not that I don't have a strong enough reason to pay for the beers, I have zero reason, no reason at all to pay for the beer.
When I present this example to my students, they say "maybe they'll call the police or maybe he's a big bartender". I imagine he's a very small bartender, I am bigger than he is and the phone lines are cut, there's no danger and prudential consideration at all. The point that I'm making is that it's still absurd for me to say "I have no reason whatsoever to pay for the beer if there's nothing in my motivational set that paying for the beer would satisfy".

Why is that absurd? Because when I ordered the beer and drank it, I did so under a certain set of presuppositions -I did that under the presupposition that that type of action was a commitment. I was committing myself to pay for the beer when the time came. The general point here is that the notion of a *commitment* is absolutely essential in human social life. We could not get on without the notion of commitment. But the notion of a commitment is precisely the notion of a desire -independent reason for action. The key element in the classical model and the key element in point five is that there cannot be any desire -independent reasons for action. The key way in which human rationality differs from animal rationality is that we have the capacity to create or act upon desire -independent reasons for action. I promise to come to Milan and give a talk. On the date in question I can't say: "well, do I really feel like going to Milan today? Does that fit into my motivational set? Would it be fun?" I am not allowed to ask myself that under rational considerations because I have created a desire -independent reason for action.

Nietzsche said that's the amazing thing about human beings: they can make promises. And when I was a student at Oxford, promising was regarded as so awesome that some philosophers never made any promises. Pritchard, it was said, would never make any promises. If he was invited to a party he'd say: "I fully intend to come to your party". But he would never say "I promise to come", because the sheer burden of the metaphysics of that were too enormous for him. This feature of promising, namely you create a desire -independent reason for action, is pervasive in language. You do that whenever you order a beer or you call somebody on the telephone or invite someone to a party: you see that language in general has this feature of promising; that you create desire -independent reasons for action. The obvious case is an assertion. If you ask me what the weather is like and I say "it's sunny out today", I commit myself to the truth of a certain proposition about the weather.

In our culture there's supposed to be a big distinction between practical reason and theoretical reason, and on the account that I'm suggesting, theoretical reason is just a special case. It is the special case where you reason what to believe, or what to accept, it's just a special case of practical reason. And in the case of theoretical reason, there's no question that you have desire -independent reasons for accepting truth. If you've asserted that *p*, and if you've asserted that "if *p* then *q*", then you are committed to the truth of *q* in the same sense that you are committed if you've made a promise to somebody to do something. You have a desire -independent reason for accepting truth, in the same way that you have a desire -independent reason for carrying out courses of action that you have committed yourself to doing.

So, far from it being the case of there being something unusual about desire independent reasons, that it only happens in the case of promising, I think it is the typical form of human social life. And, indeed, the creation of desire -independent reasons is built into the structure of almost every kind of speech act.

The oddity of the tradition is that, far from it being the case that you can't derive an *ought* from an *is*, we're committed to this derivation pretty much whenever we have a belief, then it ought to be a true belief, other things being equal. Whenever we make an assertion, it ought to be a true assertion. It ought to be one for which you have reasons. It ought to be one that is consistent with your other assertions. Whenever you make a promise, you have created a desire -independent reason for carrying out the action you promise. So, far from it being the case that there are no desire -independent reasons, which is really what point number five says, desire independent reasons are pervasive in human social life.

9.

Finally, what about the claim that the set of desires must be consistent? This is put by a number of theorists, and it's stated explicitly in a book by Elster, which says that rationality couldn't get going, we'd have no way to reason rationally, if we had inconsistent desires. From inconsistency anything follows, so you could not have rationality if you had inconsistent desires.

But I think you couldn't possibly have a consistent set of desires, because there's no way that you can satisfy one desire without frustrating a whole lot of others. Let's suppose that I want to go to Milan, and I've made up my mind that I'm going to go by airplane. All the same, I don't want to go to airports and stand in line in airports, I don't want to eat airplane food, and so on. There is just an enormous number of frustrated desires that come whenever you're trying to satisfy any desire. Now, the standard answer to that point in the literature is to say: "what you have there is a preference schedule, and you prefer this to this, to thisÉ and it's the preferences that enable the set of desires to be consistent, because, though it's true that you frustrated one

desire you have satisfied another, and you have an order of preferences". The difficulty with that is that the preferences are typically the result of rational deliberation, rather than a presupposition. That is, you don't know in advance of the deliberation of what you prefer -I'd like to be in Milan today but I would also like to be in Paris -which do I prefer? I have to think about it. It's not something that comes prior to the deliberating situation. Typically, in short, the set of preferences is the result of deliberation and not its presupposition.
We need to take seriously three absolutely fundamental features which are left out of the classical model. One is consciousness. You cannot talk about rationality without talking about conscious agents: only for conscious agents is rationality possible. So, our theory of rationality is part of our theory of consciousness.
Secondly, the gap. We cannot begin to discuss rationality without talking about *freedom*, because, just as consciousness is a presupposition throughout the whole discussion, so a certain form of consciousness, namely conscious -voluntary decision -making, provides the specific field in which rationality operates. And that is only under the presupposition of a gap. So, if consciousness and freedom are two essential presuppositions, our theory of rationality is a theory of that form of consciousness where freedom is manifest. I hate these grand terms like consciousness or freedom, but let's always bring it down to actual cases, where you're wondering how you're going to pay for a beer in a bar. And then you'll see you've got to have consciousness and you've got to have a gap.
The third feature that's essential to our whole discussion is the notion of commitment. It is characteristic of human beings as opposed to animals, that they are capable of commitment. That, by the way, is what was wrong with my friend in the Pentagon. He thought that the decision to fight a war was like the decision to buy a tube of toothpaste, strictly a matter of margin, and of getting a better marginal utility. But it isn't, it's a matter of weighing seriously what your deepest commitments are. And the whole notion of a commitment is a notion of a desire -independent reason for action.
So, the topic I have begun today and by no means finished, is that feature of consciousness that has to do with voluntary decision -making. That has to do with pre -conscious decision -making operating in the gap, and the capacity that human beings have to create desire -independent reasons by way of undertaking commitments.

A QUESTION from the audience:
It seems to me that the definition of the classical position involves some more empirical questions than philosophical statements, for example, the statement that rationality is a separate cognitive faculty is something that cognitive science is likely to explain, so philosophy has little to say here.
Also, regarding rules, we can imagine that the question of whether we are using rules of reasoning is something empirical, to be subjected to normal empirical analysis.
I'd like to ask you about commitment and desire, which could simply be a question of vocabulary. I don't think there are two kinds of desires, one consisting of the more stable sort of desires, which are a part of personality: for example paying for the beer because of the sort of person I'd like to be, or a desire to conform to a certain society, which is a commitment that could be defined as a sort of desire.
About the gap, my first reaction is that it is the statement of the problem, not the solution. Could you explain further what its physiological basis is and thus how akrasia *exists?*
Also, we agree that consciousness is important for rationality. But could we consider that, if a desire is consciously entertained, it is not the cause of something, but is more like a judgement or a perception? And so couldn't this be a reason not to be so worried about the connection between desire and motivation?

SEARLE:
Of course there are empirical questions, but there is also a conceptual point. In a way that it is possible to conceive language without vision, or vision without language -because those are two separate faculties -it is not possible to conceive language without constraints of rationality. We can only speak if we understand ourselves as having certain rational commitments when we speak, and we can only understand others if we assume that they are operating under some constraints of rationality. So, of course there are empirical questions about how all this operates in the brain, and I hope cognitive science will investigate those

questions. But there is a conceptual point, namely that rationality is not a separate faculty from language and thought, but is a set of constitutive principles.

Secondly, *rules*. Certainly there are all kinds of rules that we try to follow. Here's one that I try to follow in the stock -market not very successfully: "buy cheap, sell high; buy low, sell high". That's a good rule. I don't always follow it, but those are strategic principles and not a set of separate constitutive principles of rationality.

My position is a little bit subtle. The point is this: there are constitutive rules of speech acts and constitutive constraints on thought and those, in a sense, are the rules of rationality. There isn't any separate extra -set of rules. So the idea that there is no separate rationality faculty is really the same as the idea that there is no separate set of independent rules. It isn't that you have the rule of *modus ponens*, but you might have had a lot of other rules, rather there's a structural constraint on rationality and commitment, that has to do with the kind of commitments that are involved in your truth claims.

Thirdly, *the gap*. Of course the gap doesn't solve all the questions, it just poses other questions. But we have to take our experiences seriously. It is just the fact of my experience that I have an awareness of my own freedom. If I raise my right hand or my left hand there's a sense of freedom and action that is not true of perception. If I hold this up, its not up to me whether or not I can see it, whether I perceive it. I can turn my head away, but if Im looking at it point blank, then perceptual consciousness is deterministic in a way that volitional consciousness is not deterministic. Thats just a fact that I experience volitional consciousness as free.

But now the question arises: couldnt it be the case that we have the illusion of freedom, but in fact its all determined at the level of the neurons and the synapses. And the answer is: it could turn out that way. But if we came to believe that, then this idea that our consciousness does not make any difference to the rest of our life would be the biggest revolution in our thinking in human history -much bigger than Newton and Galileo and Einstein all put together: that it doesn't really make any difference, that its just going on for the ride and that everything is actually determined by operations of the "plumbing" in the brain, operations at the level of the neurons and the synapses, and that it is completely deterministic.

However, this poses a serious question to us. How is it supposed to work in the brain? My whole approach is utterly naturalistic. Most of the philosophers confronted with this question cheat. Kant cheats, he says: "youve got to think of yourself as free in the noumenal world and determined in the phenomenal world." But I live in just one world. I cant just get on the train to the noumenal world and everything will be all right. That train is in a permanent "*sciopero*"!

Because I live in one world, Kants solution is not a solution. Kant at least faced the problem. Most philosophers dont face the problem and they say: "of course you're free -that just means we say youre free and of course youre determined, because everythings determined, now lets talk about something else!" But thats an evasion, you have to confront a fact. Either all this conscious apparatus and decision -making actually makes a difference to my life or it doesn't. If it doesnt make a difference to my life, then, as I said, that is the most massive illusion that weve had in human history. One of the reasons why Im suspicious of that conclusion is that evolution doesnt work that way. The evolutionary and the biological costs of having this enormous conscious decision -making apparatus is just too high. Its too high for the organism, it's too high for the blood -flow in the organism -all this blood going to the brain and this enormous amount of biological cost in sustaining this kind of consciousness, not to mention the ecological cost of raising children in a painful and exhausting way so they develop this kind of consciousness. If it turned out that that made no difference at all, then it would go against everything we know about evolution, because evolution doesnt involve such an elaborate phenotype if it has no function whatever. On the other hand, our current models of the brain make it look as if the brain must be as deterministic as a car engine. The problem is we really dont know how the brain works, so there are two possibilities. One possibility is that we have the illusion of the gap at the top of the whole, but in fact its all mechanically determined at the level of neurons and synapses. Another possibility is that the movement of rational decision -making, where you have reasons for an action, though the reasons do not set causally sufficient conditions, that movement from left to right across time goes all the way down. It goes right down to the synaptic cleft. So that what youve got is a sequence of conscious decision -making processes, which are conscious and rational, but where the causal antecedents at any step are not causally sufficient to determine the next step.

So, were not talking about the noumenal and the phenomenal, we're talking about the thalamo -cortical system, about how the brain works. Then it seems there are at least two options open, and the truth is we dont know how the brain works.
In any mature science there is a guiding principle. The guiding principle of physics is the atom, of genetics it is DNA, of geology the tectonic plate. Here is the scandal in Neurobiology: we dont have a guiding principle; we dont know. Most textbooks assert that the neuron is the right level of analysis. And they try to give an analysis of the functioning of the neuron as if it were a straight deterministic system. But we dont know if thats the right level. It may well be that you've got to have all clouds of neurons. That the functional unit for things like consciousness has to do with the function of whole clouds of neurons and that the form of mathematics is chaos theory. That were talking about a chaotic dynamics, which would at least make the gap an empirical possibility.
Other people say: "no, a mass of neurons is much too high a level, youve got to go down to the sub -neuron level, down to the quantum mechanical level, down to the level of the micro -tubules". At this point we dont know.
This is the most serious question we raised: *how does free will work in the brain*? And the short answer is: *we dont know*. There are two possibilities. One is that its a massive illusion. Nature is playing its biggest joke on us in all history. We don't have free will, all our consciousness is just an illusion and everything is controlled by systems as mechanical as tropisms. I cant believe that, it runs against all we know about nature. But the other view sounds just as crazy: it is that consciousness moves forward from left to right and the indeterministic element in consciousness, namely the gap, where you have rational decision -making - which is causable -but not on the base of causally sufficient conditions: that goes all the way down to the synaptic cleft, and, presumably, all the way down to the quarks and the muons, down to the quantum mechanical level.
Those are the two options. Neither is very attractive. I do not solve the problem of the freedom of will, but that is the problem, it is a neuro -biological problem.
Now the last question. It was about whether or not we should think of consciousness as part of the world of reasons rather than as part of the world of causes. I dont make that distinction. One of the things that drives me in philosophy is the feeling that many of our problems come from the fact that we accept a set of obsolete categories: mind and body, subjectivity and objectivity, in this particular case causes and reasons. What Im trying to do is give an account of us as biological and perfectly natural beasts in the world. But at the same time we have these special features: consciousness, intentionality and rationality.

QUESTION from the audience:
I have a question about the gap and modus ponens. You pointed out that the conclusion is not true in virtue of the rule, but I dont think you can point that out to the tortoise to convince it that it should believe the conclusion of modus ponens. You say in your hand -out that the tortoise is free not to believe the conclusion of modus pones, but I dont understand why the tortoise is free not to believe the conclusion. Do you mean that the tortoise can assert "I believe P and I believe if P then Q but I don't believe Q" and it is sincere and it understands what it is saying?

SEARLE:
The point that Im making is that the gap is pervasive and it pervades even logic. Of course its irrational and self -contradictory of the tortoise to assert the premises of the argument and deny the conclusions. But nonetheless its a possibility. The point that Im making here is that the reasons that you have for accepting the conclusion, though they are rationally compelling, are not causally deterministic. Theyre not causally compelling, because it's always open to people to announce self -contradictory views. They often do this. When its a political or religious issue, people often hold views which are inconsistent.

QUESTION:
Do you mean that for instance one could say: "it is raining but I do not believe it is"?

SEARLE:
Ive written a book trying to explain why one cannot consistently say that. The point, however, is not a causal point. Are the causes operating on me causally sufficient, so that they make it causally impossible that I should say “its raining but I dont believe it” in the same way that the causes operating on an object are such that they make it causally sufficient that it will fall when I drop it? No, the causes are not like that. Theres a distinction between the rational grounds that I have for making the decision -and those do function causally - and causally sufficient conditions. I’m glad about this question, because it points out that there is a gap even in logic, even though here the gap is less visible because its part of the definition of rationality that you are rational and you accept the valid conclusions of your argument.

QUESTION:
I have a question about desire -independent reasons. As Bernard Williams says, when one has a duty, it does motivate and at least indirectly it is a part of your motivational set, even if it is not a real desire.
Id also like to know if in the book you are writing there is something about ethics connected with your ideas of commitment, which seem to lead to some sort of idea of virtue or ethics.

SEARLE:
Bernard Williams has been very helpful to me because I gave him a draft of this and then he wrote some detailed comments and we never had a chance to discuss it. But theres a serious difference between us. Williams is prepared to recognize that people often act out of a sense of duty. But that he thinks it is because they have a desire to do their duty and thats just a desire like any other. Here is the exact point of disagreement. I think there are a lot of things where you recognize the reason as the ground for the desire rather than the desire as the ground of the reason. The difference is this: I think that often you recognize that you have a duty to do something and you want to do it because you have a duty to do it. But its not the case that you have a duty to do it that functions as a reason: just because you want to do your duty.
This goes to the point that came out earlier, that I didn’t respond to. How about the guy in the bar who just wants to be the sort of person who pays his bar bills? But why would anyone have that desire? Is it just like a desire for chocolate as opposed to vanilla? I think its different and the reason is that in the case of desire - independent reasons like duty, commitment and obligation, those form the ground of desires. The reason why I want to do it is that I recognize I have an obligation to do it. I accept the principle that in some sense all actions are expressions of the desire to perform that action. The point where I differ from the classical model is that I think you can have a desire which is motivated by something not a desire, but a desire - independent reason.
The second question is “whats this got to do with ethics?” I think much of the weaknesses of ethical theory derive from the fact that people are too eager to talk about ethics and not talk about rationality and reasons. We should first get clear about rationality and reasons. I think that the study of practical reason is a much more fundamental study. Ethics is just a branch of practical reason. The most interesting subject is practical rationality. And it will have consequences for ethics. You cannot do ethics without a theory of action and rationality. So, in this book Im doing an introduction to the basic principles which will make possible a theory of ethics. If you look at the literature, it really is very disappointing. People start off talking about free will and immediately they start talking about ethical responsibility. Or they start talking about commitment and then immediately theyre off about moral obligation. Lets put that off a bit. Let’s get clear about what a commitment is, and then it turns out that a commitment arises when you have this peculiar phenomenon whereby you impose conditions of satisfaction on conditions of satisfaction. Thats a complicated notion that I explain in my book about intentionality. But the idea is that we can create reasons for ourselves that my dog cant create. He cant make a promise to do something next Wednesday, because he cant in that way create a desire -independent reason, by imposing conditions of satisfaction upon conditions of satisfaction. So, thats a more complex idea.

QUESTION:
In the contemporary literature in ethics, is Elizabeth Beck's work very similar to this?

SEARLE:
Elizabeth Beck has had a couple of ideas that are terrific in this regard. One is that she has challenged the idea that primary desires are beyond the scope of rational assessment. What I dont find in her work is a well -worked -out theory. But I certainly admire enormously the contribution that she has made.

QUESTION:
I have a question about the relationship between your theory of rationality and your speech acts analysis. [incomprehensible words]. Commitment plays a special role in your theory of rationality. How does it affect speech act analysis? Could we say that commitment or commissive function is a general trait of every kind of speech act, and not a special feature as in commissive ends in speech acts?

SEARLE:
In my taxonomy of basic types of speech acts, I make a distinction between assertives, directives, commissives, expressives and declaratives; and commissives are a special class where you are committed to performing a future course of action -thats the primary point of the speech act. The questioner says that what Im saying today suggests that every speech act is a commissive. So, whats going on here? The answer is that every speech act involves some level of commitment. But you still need a special class of commissives. Why? Because the only point of the commissive is to commit the speaker to the performance of some future course of action. And of course thats not the point of a directive or an assertive. If I say "It's raining", that does commit me. But thats not the point, maybe theres some future course of action, but what Im committed to there is the existence of a state of affairs. So, though there is an element of desire -independent reasons in all speech acts, it doesnt follow that you dont need a separate class. And there is a well -defined class of commissives.
I see all of my work as hanging together, the way each book follows from the previous book. I've now published ten books, and in a way they are all chapters of the whole thing. Sometimes I realize Ive made a mistake and I correct it in subsequent books. But the basic idea is that all these issues in philosophy -mind, language, society, commitment, rationality, intentionality -they all hang together. And in order to get any part right, you have to get the rest right, absolutely.

QUESTION:
I have a question about the roles of language. You said that rationality is not separate from language and thought. Apes and monkeys have no rationality because they have no language, they cant learn language. I wonder if this is because only with language you can create desires and beliefs and thus initiate rational behavior, but perhaps another reason is that language is fundamental for consciousness, which is important for the gap between reasons and action. Which of these two is more important for language?

SEARLE:
The question seems to be in two parts: one is about apes and the other about the special relation between consciousness and language.
Its not my thesis that animals cant have any rationality at all. Indeed, one of the examples to cite is the example of Khlers chimpanzees on the Island of Tenerife. He proved that they were capable of simple rational means and decision -making. They figured out how to reach up with a stick and get the bananas. That is a type of rationality. The point is that animals are capable of a rationality of a very limited kind. But elaborate cases of rationality require language. So the apes can figure out how to get the bananas with a stick, but the ape can't figure out how its going to get bananas next year or next week, or what its going to do on Christmas day. Because it does not have a linguistic capacity for representing it.
So, language is essential not to the existence of rational decision -making as such, but to any kind of complexity. Language gives us much more power.
The second point is about the relation between language and consciousness, and its not my thesis that consciousness requires language, because animals do have consciousness but do not typically have language. It is not clear what the significance of the efforts to teach the apes is. The results I have seen are very

ambiguous, they do not show any evidence for the performance of speech acts of any interesting kind. Most researchers would agree with me. The point is that language is not essential to consciousness, but an elaborate kind of human consciousness requires a language.
So, I can have a conscious experience of red and my dog can have a conscious experience of red. And I can hear sounds and my dog can hear the same sounds. But the ability for me is to see the red as a left -wing political symbol: you have to have language for that. La Rochefoucauld said: “very few people would ever fall in love if theyd never read about it”. If theyd never seen it on TV, the movies and so on. For a lot of human emotions, language is partly constitutional. If you didn’t have the vocabulary, you couldnt fall in love. You can have sexual attraction without vocabulary. But given a vocabulary, you can do all sorts of things with sexual attraction which you cant do without it. And that goes for a lot of other feelings. My dog has a lot of feelings, but he doesnt, for example, suffer the angst of post -industrial man, or dog, under late capitalism. I dont either, but I know people who claim they suffer the angst of post -industrial man under late capitalism.
So, language is not essential to consciousness, but human consciousness -falling in love, wondering what political party youre going to vote for, wondering if you should get married, and whether or not you should spend the summer in Calabria -requires language.

Notes:

1.1) Lecture held in Milan, May 30, 2000, at the State University of Milan.
1) *Strike* in Italian [*Editor’s Note*].