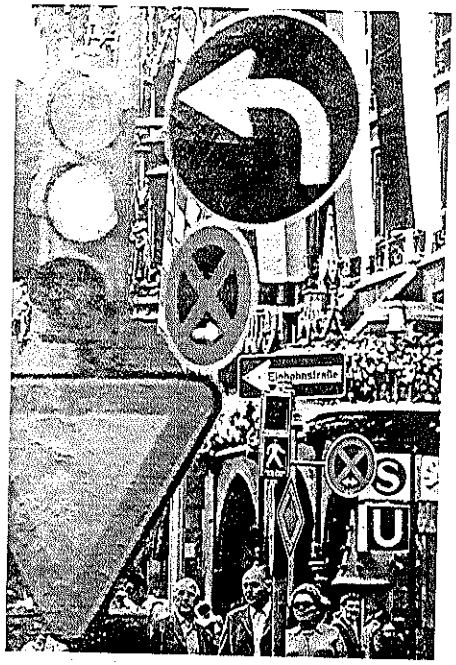
JOHN R. SEARLE The Construction of Social Reality



The Construction of Social Reality

JOHN R. SEARLE

things that exist. We will have to make some substantive presuppositions about *how the world is in fact* in order that we can even pose the questions we are trying to answer. We will be talking about how social reality fits into a larger ontology, but in order to do that, we will have to describe some of the features of that larger ontology.

The truth is, for us, most of our metaphysics is derived from physics (including the other natural sciences). Many features of the contemporary natural science conception of reality are still in dispute and still problematic. For example, one might think that the Big Bang Theory of the origin of the universe is by no means well substantiated. But two features of our conception of reality are not up for grabs. They are not, so to speak, optional for us as citizens of the late twentieth and early twenty-first century. It is a condition of your being an educated person in our era that you are apprised of these two theories: the atomic theory of matter and the evolutionary theory of biology.

The picture of reality derived from these two theories, to state it very crudely, is as follows: The world consists entirely of entities that we find it convenient, though not entirely accurate, to describe as particles. These particles exist in fields of force, and are organized into systems. The boundaries of systems are set by causal relations. Examples of systems are mountains, planets, H_2O molecules, rivers, crystals, and babies. Some of these systems are living systems; and on our little earth, the living systems contain a lot of carbon-based molecules, and make a very heavy use of hydrogen, nitrogen, and oxygen. Types of living systems evolve through natural selection, and some of them have evolved certain sorts of cellular structures, specifically, nervous systems capable of causing and sustaining consciousness. Consciousness is a biological, and therefore physical, though of course also mental, feature of certain higher-level nervous systems, such as human brains and a large number of different types of animal brains.

With consciousness comes intentionality, the capacity of the mind to represent objects and states of affairs in the world other

than itself.* Not all consciousness is intentional, and not all intentionality is conscious. There are, for example, forms of consciousness such as undirected anxiety that do not represent anything; and there are many forms of unconscious intentionality, such as my belief, even when I am not thinking about it, that Bill Clinton is president. However, though there is no necessary connection between being an intentional state at a given time and being conscious then and there, nonetheless, there is an important necessary connection between the two, in that every intentional state that is unconscious is at least accessible to consciousness. It is the sort of thing that could be conscious. An unconscious intentional state has to be in principle accessible to consciousness.

Here, then, are the bare bones of our ontology: We live in a world made up entirely of physical particles in fields of force. Some of these are organized into systems. Some of these systems are living systems and some of these living systems have evolved consciousness. With consciousness comes intentionality, the capacity of the organism to represent objects and states of affairs in the world to itself. Now the question is, how can we account for the existence of social facts within that ontology?

Objectivity and Our Contemporary World View

Much of our world view depends on our concept of objectivity and the contrast between the objective and the subjective. Famously, the distinction is a matter of degree, but it is less often re-

*I use "intentionality" as a technical term meaning that feature of representations by which they are *about* something or *directed at* something. Beliefs and desires are intentional in this sense because to have a belief or desire we have to believe that such and such is the case or desire that such and such be the case. Intentionality, so defined, has no special connection with intending. Intending, for example, to go to the movies is just one kind of intentionality among others. For a fuller account of intentionality, see J. R. Searle, *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983).

marked that both "objective" and "subjective" have several different senses. For our present discussion two senses are crucial, an *epistemic* sense of the objective-subjective distinction and an *ontological* sense. Epistemically speaking, "objective" and "subjective" are primarily predicates of judgments. We often speak of judgments as being "subjective" when we mean that their truth or falsity cannot be settled "objectively," because the truth or falsity is not a simple matter of fact but depends on certain attitudes, feelings, and points of view of the makers and the hearers of the judgment. An example of such a judgment might be, "Rembrandt is a better artist than Rubens." In this sense of "subjective," we contrast such subjective judgments with objective judgments, such as the judgment "Rembrandt lived in Amsterdam during the year 1632." For such objective judgments, the facts in the world that make them true or false are independent of anybody's attitudes or feelings about them. In this epistemic sense we can speak not only of *objective judgments* but of *objective facts*. Corresponding to objectively true judgments there are objective facts. It should be obvious from these examples that the contrast between epistemic objectivity and epistemic subjectivity is a matter of degree.

In addition to the *epistemic* sense of the objective-subjective distinction, there is also a related *ontological* sense. In the ontological sense, "objective" and "subjective" are predicates of entities and types of entities, and they ascribe modes of existence. In the ontological sense, pains are subjective entities, because their mode of existence depends on being felt by subjects. But mountains, for example, in contrast to pains, are ontologically objective because their mode of existence is independent of any perceiver or any mental state.

We can see the distinction between the distinctions clearly if we reflect on the fact that we can make epistemically subjective statements about entities that are ontologically objective, and similarly, we can make epistemically objective statements about entities that are ontologically subjective. For example, the statement "Mt. Everest is more beautiful than Mt. Whitney" is about ontologically objective entities, but makes a subjective judgment about them. On the other hand, the statement "I now have a pain in my lower back" reports an epistemically objective fact in the sense that it is made true by the existence of an actual fact that is not dependent on any stance, attitudes, or opinions of observers. However, the phenomenon itself, the actual pain, has a subjective mode of existence.

The Distinction Between Intrinsic and Observer-Relative Features of the World

Historically in our intellectual tradition we make great distinctions between mind and body and between nature and culture. In the section on Fundamental Ontology, I tacitly abandoned the traditional dualistic conception of the relation of mind and body in favor of the view that the mind is just a set of higher-level features of the brain, a set of features that are at once "mental" and "physical." We will use the "mental," so construed, to show how "culture" is constructed out of "nature." The first step is to introduce a more fundamental distinction than those mentioned above. This is the distinction between those features of the world that exist independently of us and those that are dependent on us for their existence.

The features of the world I described in characterizing our fundamental ontology, e.g., mountains and molecules, exist independently of our representations of them. However, when we begin to specify further features of the world we discover that there is a distinction between those features that we might call *intrinsic* to nature and those features that exist *relative to the intentionality of observers, users, etc.* It is, for example, an intrinsic feature of the object in front of me that it has a certain mass and a certain chemical composition. It is made partly of wood, the cells of which are composed of cellulose fibers, and also partly of metal, which is itself composed of metal alloy molecules. All these features are intrinsic. But it is also true to say of the very same object that it is a

screwdriver. When I describe it as a screwdriver, I am specifying a feature of the object that is observer or user relative. It is a screwdriver only because people use it as (or made it for the purpose of, or regard it as) a screwdriver. The existence of observer-relative features of the world does not add any new material objects to reality, but it can add epistemically objective *features* to reality where the features in question exist relative to observers and users. It is, for example, an epistemically objective feature of this thing that it is a screwdriver, but that feature exists only relative to observers and users, and so the feature is ontologically subjective. By "observers and users" I mean to include makers, designers, owners, buyers, sellers, and anyone else whose intentionality toward the object is such that he or she regards it as a screwdriver.

Since the issues are important and the example is simple, I want to belabor these points a bit further.

1. The sheer existence of the physical object in front of me does not depend on any attitudes we may take toward it.

2. It has many features that are intrinsic in the sense that they do not depend on any attitudes of observers or users. For example, it has a certain mass and a certain chemical composition.

3. It has other features that exist only relative to the intentionality of agents. For example, it is a screwdriver. To have a general term, I will call such features "observer relative." Observer-relative features are ontologically subjective.

4. Some of these ontologically subjective features are epistemically objective. For example, it isn't just my opinion or evaluation that it is a screwdriver. It is a matter of objectively ascertainable fact that it is a screwdriver.

5. Although the feature of being a screwdriver is observer relative, the feature of thinking that something is a screwdriver (treating it as a screwdriver, using it as a screwdriver, etc.) is intrinsic to the thinkers (treaters, users, etc.). Being a screwdriver is observer relative, but the features of the observers that enable them to create such observer-relative features of the world are intrinsic features of the observers. I will shortly explain this point further.

It is not always immediately obvious whether a feature is intrinsic or observer relative. Colors are a good example. Prior to the development of physics in the seventeenth century, people thought of colors as intrinsic features of the world. Since then many people have come to think of them as properties that exist only relative to observers. It is intrinsic that light differentially scatters when reflected from surfaces, and intrinsic to people that they have subjective color experiences caused by the impact of light on their visual systems. But the further attribution of color properties to objects in the world is observer relative, because it can be made only relative to the experiences of observers, as caused by the impact of light. I am not here trying to settle the issue about colors, but calling attention to the fact that whether a feature is intrinsic or observer relative is not always obvious.

A good rough-and-ready way of getting at this distinction is to ask youself, Could the feature exist if there had never been any human beings or other sorts of sentient beings? Observer-relative features exist only relative to the attitudes of observers. Intrinsic features don't give a damn about observers and exist independently of observers. One qualification has to be added immediately to this test, and it is stated in point 5 above, namely, that acts of observing and using are themselves intrinsic. So, to put it very crudely, something is a screwdriver only relative to the fact that conscious agents regard it as a screwdriver; but the fact that conscious agents have that attitude is itself an intrinsic feature of the conscious agents. Because mental states, both conscious and unconscious, are themselves intrinsic features of the world, it is not strictly speaking correct to say that the way to discover the intrinsic features of the world is to subtract all the mental states from it. We need to reformulate our explanation of the distinction to ac-

count for this exception as follows: Intrinsic features of reality are those that exist independently of all mental states, except for mental states themselves, which are also intrinsic features of reality.

From a God's-eye view, from outside the world, all the features of the world would be intrinsic, including intrinsic relational features such as the feature that people in our culture regard such and such objects as screwdrivers. God could not see screwdrivers, cars, bathtubs, etc., because intrinsically speaking there are no such things. Rather, God would see *us treating* certain objects as screwdrivers, cars, bathtubs, etc. But from our standpoint, the standpoint of beings who are not gods but are inside the world that includes us as active agents, we need to distinguish those true statements we make that attribute features to the world that exist quite independently of any attitude or stance we take, and those statements that attribute features that exist only relative to our interests, attitudes, stances, purposes, etc.

In each of the following pairs, the first states an *intrinsic* fact about an object, and the second states an *observer-relative* fact about the very same object.

1a. intrinsic: That object is a stone.

1b. observer relative: That object is a paperweight.

2a. intrinsic: The moon causes the tides.

2b. observer relative: The moon is beautiful tonight.

3a. intrinsic: Earthquakes often occur where tectonic plates meet.

3b. observer relative: Earthquakes are bad for real estate values.

I want this distinction to seem quite obvious, because it is going to turn out that social reality in general can be understood only in light of the distinction. Observer-relative features are always created by the intrinsic mental phenomena of the users, observers, etc., of the objects in question. Those mental phenomena are, like all mental phenomena, ontologically subjective; and the observer-relative features inherit that ontological subjectivity. But this ontological subjectivity does not prevent claims about observer-relative features from being epistemically objective. Notice that in 1b and 3b the observer-relative statement is epistemically objective; in 2b it is subjective. These points illustrate the ways in which all three distinctions cut across each other: the distinction between the intrinsic and the observer relative, the distinction between ontological objectivity and subjectivity, and the distinction between epistemic objectivity and subjectivity.

It is a logical consequence of the account of the distinction as I have so far given it that for any observer-relative feature F, *seeming to be F* is logically prior to *being F*, because—appropriately understood—seeming to be F is a necessary condition of being F. If we understand this point, we are well on the road to understanding the ontology of socially created reality.

The Assignment of Function

My main objective in this chapter is to assemble the apparatus necessary to account for social reality within our overall scientific ontology. This requires exactly three elements. The assignment of function, collective intentionality, and constitutive rules. (Later, in Chapter 6, to explain the causal functioning of institutional structures, we will introduce a fourth element, the Background of capacities that humans have for coping with their environment.) In explaining these notions I am perforce in a kind of hermeneutic circle. I have to use institutional facts to explain institutional facts; I have to use rules to explain rules, and language to explain language. But the problem is expository and not logical. In the exposition of the theory I rely on the reader's understanding of the phenomena to be explained. But in the actual explanation given, there is no circularity.

The first piece of theoretical apparatus I need I will call the "assignment (or imposition) of function." To explain this, I begin by noting the remarkable capacity that humans and some other animals

have to impose functions on objects, both naturally occurring objects and those created especially to perform the assigned functions.

As far as our normal experiences of the inanimate parts of the world are concerned, we do not experience things *as* material objects, much less as collections of molecules. Rather, we experience a world of chairs and tables, houses and cars, lecture halls, pictures, streets, gardens, houses, and so forth. Now all the terms I have just used involve criteria of assessment that are internal to the phenomena in question under these descriptions, but not internal to the entities under the description "material object." Even natural phenomena, such as rivers and trees, can be assigned functions, and thus assessed as good or bad, depending on what functions we choose to assign to them and how well they serve those functions. This is the feature of intentionality I am calling "the assignment—or imposition—of function." In the case of some artifacts, we build the object to serve a function. Chairs, bathtubs, and computers are obvious examples. In the case of many naturally occurring objects, such as rivers and trees, we assign a function—aesthetic, practical, and so on—to a preexisting object. We say, "That river is good to swim in," or "That type of tree can be used for lumber."

The important thing to see at this point is that functions are never intrinsic to the physics of any phenomenon but are assigned from outside by conscious observers and users. *Functions, in short, are never intrinsic but are always observer relative.*

We are blinded to this fact by the practice, especially in biology, of talking of functions as if they were intrinsic to nature. But except for those parts of nature that are conscious, nature knows nothing of functions. It is, for example, intrinsic to nature that the heart pumps blood, and causes it to course through the body. It is also an intrinsic fact of nature that the movement of the blood is related to a whole lot of other causal processes having to do with the survival of the organism. But when, in addition to saying "The heart pumps blood" we say, "The *function* of the heart is to pump blood," we are doing something more than recording these intrinsic facts.

We are situating these facts relative to a system of values that we hold. It is intrinsic to us that we hold these values, but the attribution of these values to nature independent of us is observer relative. Even when we *discover* a function in nature, as when we discovered the function of the heart, the discovery consists in the discovery of the causal processes together with the assignment of a teleology to those causal processes. This is shown by the fact that a whole vocabulary of success and failure is now appropriate that is not appropriate to simple brute facts of nature. Thus we can speak of "malfunction," "heart disease," and better and worse hearts. We do not speak of better and worse stones, unless of course we have assigned a function to the stone. If we use the stone as a weapon or a paperweight or an *objet d'art trouvé,* for example, we can asses its adequacy under these functional descriptions.

This point has to be understood precisely. We do indeed "discover" functions in nature. But the *discovery* of a natural function can take place only within a set of prior *assignments* of value (including purposes, teleology, and other functions). Thus given that we already accept that for organisms there is a value in survival and reproduction, and that for a species there is a value in continued existence, we can *discover* that the function of the heart is to pump blood, the function of the vestibular ocular reflex is to stabilize the retinal image, and so on. When we discover such a natural function, there are no natural facts discovered beyond the causal facts. Part of what the vocabulary of "functions" adds to the vocabulary of "causes" is a set of values (including purposes and teleology generally). It is because we take it for granted in biology that life and survival are values that we can discover that the function of the heart is to pump blood. If we thought the most important value in the world was to glorify God by making thumping noises, then the function of the heart would be to make a thumping noise, and the noisier heart would be the better heart. If we valued death and extinction above all, then we would say that a function of cancer is to speed death. The function of aging would be to hasten death, and the function of natural selection would be

extinction. In all these functional assignments, no new intrinsic facts are involved. As far as nature is concerned intrinsically, there are no functional facts beyond causal facts. The further assignment of function is observer relative.

One of Darwin's greatest achievements was to drive teleology out of the account of the origin of species. On the Darwinian account, evolution occurs by way of blind, brute, natural forces. There is no intrinsic purpose whatever to the origin and survival of biological species. We can, arbitrarily, define the "functions" of biological processes relative to the survival of organisms, but the idea that any such assignment of function is a matter of the discovery of an intrinsic teleology in nature, and that functions are therefore intrinsic, is always subject to a variant of Moore's open-question argument: What is so functional about functions, so defined? Either "function" is defined in terms of causes, in which case there is nothing intrinsically functional about functions, they are just causes like any others. Or functions are defined in terms of the furtherance of a set of values that we hold—life, survival, reproduction, health—in which case they are observer relative.

I realize that many biologists and philosophers of biology will disagree. Over the past few decades there has developed a large literature on functions and functional explanations. Much of it is influenced by Larry Wright's article[3] in which he defines function as follows:

The function of X is Z *means*

1. X is there because it does Z.

2. Z is a consequence (or result) of X's being there.

If such an analysis were correct, it would eliminate the observer relativity of function. Intuitively the idea is to define "function" in terms of causation: X performs the function F just in case X causes F, and at least part of the explanation for X's existence is that it causes F. Thus, for example, the heart has the function of pumping blood because it does pump blood and the explanation for the existence of hearts in evolutionary history is that they do in fact pump blood. This seems to give a naturalistic definition of "function" whereby functions would be intrinsic. Ruth Millikan has a similar, but more complex, idea in her notion of "proper function," though she insists that she is not trying to analyze the ordinary use of the notion of function but to introduce a new technical expression defined in terms of "reproduction" and causation.* So construed no one could object. You can introduce any new technical terms you like. However, it is important to emphasize that such definitions fail to capture certain essential features of the ordinary notion of function, for at least three reasons. First, in Millikan's case it makes the definition of function dependent on a particular causal historical theory about "reproduction." In fact I believe my heart functions to pump blood and I also believe in a Darwinian account of how "reproduction" gives a causal historical account of the evolution of hearts. But even if no such account of reproduction, Darwinian or otherwise, turned out to be true, my heart would still function to pump blood. On her definition the

*R. G. Millikan, *Language, Thought, and Other Biological Categories: New Foundations for Realism* (Cambridge, Mass.: MIT Press, 1984). In R. G. Millikan, "In Defense of Proper Functions," in *The Philosophy of Science* 56 (1989), 288–302. She writes:

> The definition of a "proper function" is recursive. Putting things very roughly, for an item A to have a function F as a "proper function," it is necessary (and close to sufficient) that one of these two conditions should hold. (1) A originated as a "reproduction" (to give one example, as a copy, or a copy of a copy) of some prior item or items that, *due* in part to possession of the properties reproduced, have actually performed F in the past, and A exists because (causally historically because) of this or these performances. (2) A originated as the product of some prior device that, given its circumstances, had performance of F as a proper function and that, under those circumstances normally causes F to be performed by *means* of producing an item like A. Items that fall under condition (2) have "derived proper functions," functions derived from the functions of the devices that produce them.(p. 288)

very meaning of the claim that the heart has the (proper) function of pumping blood can be explained only in terms of a causal historical account of how hearts are reproduced, and that cannot be right as far as our ordinary notion of function is concerned. Second, if we take such definitions as capturing the essential features of our ordinary notion, there are counterexamples to the analyses. On Wright's account and apparently on Millikan's as well, we would have to say that the function (proper or otherwise) of colds is to spread cold germs. They do in fact spread cold germs, and if they did not spread cold germs they would not exist. But on our ordinary notion colds do not have a function, or if they do it is certainly not to spread germs. Third, the normative component of functions is left unexplained. Though analyses such as Millikan's can account for the fact that some entities that have a function do not in fact carry out the function, the reduction of function to causal notions still leaves out the normative component. Why do we talk of malfunctioning hearts, of heart disease, of better and worse hearts? The usual dilemma shows up: either we are talking about brute, blind causal relations, in which case hearts pumping blood and colds spreading germs are in the same basket, or we think there is something really functional about functions, in which case this type of definition leaves out the observer-relative feature.

Another, and perhaps decisive, clue that functions, unlike causes, are observer relative is that functional attributions, unlike causal attributions, are intensional-with-an-s.* Substitution of coreferential terms in function contexts fails to guarantee preservation of truth value. Thus "The function of A is to X" together with "X-ing is identical with Y-ing" do not imply "The function of A is to Y." For example, it is trivially true that the function of oars is to row with, and rowing consists in exerting pressure on water relative to a fixed fulcrum; but it is not the case that the function of oars is to exert pressure on water relative to a fixed fulcrum.

*Intensionality-with-an-s should not be confused with intentionality-with-a-t. Intentionality is that property of the mind by which it is directed at objects and states of affairs in the world. Intensionality is that property of sentences and other representations by which they fail certain test for extensionality. One of the most famous of these is Leibniz's Law: If two expressions refer to the same object they can be substituted for each other in a sentence without changing the truth value of the sentence. Sentences that fail this test are said to be *intensional* with respect to substitutability. Another expression used to name this sort of intensionality is "referential opacity." Typically sentences that are about intentional-with-a-t states are intensional-with-an-s sentences, because in such sentences the way in which an object is referred to affects the truth value of the sentence. For extensive discussion of these matters see Searle, *Intentionality, An Essay in the Philosophy of Mind.*

To summarize, the first feature we need to note in our discussion of the capacity of conscious agents to create social facts is the assignment of functions to objects and to other phenomena. Functions are never intrinsic; they are assigned relative to the interests of users and observers.

I have not attempted to analyze the sentence form "The function of X is to Y" into logically necessary and sufficient conditions. But I am calling attention to certain central conditions.

1. Whenever the function of X is to Y, X and Y are parts of a *system* where the system is in part defined by *purposes, goals, and values generally.* This is why there are functions of policemen and professors but no function of humans as such—unless we think of humans as part of some larger system where their function is, e.g., to serve God.

2. Whenever the function of X is to Y, then X is *supposed to* cause or otherwise result in Y. This normative component in functions cannot be reduced to causation alone, to what in fact happens as a result of X, because X can have the function of Y-ing even in cases where X fails to bring about Y all or even most of the time. Thus the function of safety valves is to prevent explosions, and this is true even for valves that are so badly made that they in fact fail to prevent explosions, i.e., they *malfunction.*

The examples we have considered so far suggest a further distinction between *agentive* and *nonagentive* functions. Sometimes the assignment of function has to do with our immediate purposes, whether practical, gastronomic, esthetic, educational, or whatever. When we say, "This stone is a paperweight," "This object is a screwdriver," or "This is a chair," these three functional notions mark *uses* to which we put objects, functions that we do not discover, and that do not occur naturally, but that are assigned relative to the practical interests of conscious agents. Not all these interests are "practical" in any ordinary sense, because such functions are also assigned when we say "That is an ugly painting." Because all these are instance of uses to which agents intentionally put objects, I will call them "agentive functions." Some of the objects to which we assign agentive functions are naturally occurring, such as a stone that we use as a paperweight; some are artifacts made specifically to perform these functions, such as chairs, screwdrivers, and oil paintings. An object manufactured to perform one agentive function can be used to perform another, as reported, e.g., by "This hammer is my paperweight." As in the case of the heart, the function is not intrinsic to the object in addition to its causal relations, but in contrast to the ascription of function to the heart, in these cases the ascription of the function ascribes the *use to which we intentionally put* these objects.

Some functions are not imposed on objects to serve practical purposes but are assigned to naturally occurring objects and processes as part of a theoretical account of the phenomena in question. Thus we say "The heart functions to pump blood" when we are giving an account of how organisms live and survive. Relative to a teleology that values survival and reproduction, we can discover such functions occurring in nature independently of the practical intentions and activities of human agents; so let us call these functions "nonagentive functions."[4]

There is no sharp dividing line between the two, and sometimes an agentive function can replace a nonagentive function, as when, for example, we make an "artificial heart." It is generally,

though by no means always, the case that agentive functions require continuous intentionality on the part of users for their maintenance, whereas nonagentive functions continue to chug functionally along without any effort on our part. Thus bathtubs, coins, and screwdrivers require continued use on our part in order to function as bathtubs, coins, and screwdrivers, but hearts and livers continue to function as hearts and livers even when no one is paying any attention. Furthermore, the person actually using some object for an agentive function may not be the agent who actually imposed the function on that object and may even be unaware that the object has that function. Thus most car drivers are probably unaware that the function of the drive shaft is to transmit power from the transmission to the axles, but all the same that is its agentive function.

One more distinction: Within agentive functions we need to identify a special class. Sometimes the agentive function assigned to an object is that of standing for or representing something else. Thus, when I draw a diagram of a football play, I let certain circles stand for the quarterback, the runningback, the offensive linemen, and so on. In this case, the agentive function assigned to the marks on the paper is that of representing or standing for; but because "representing" and "standing for" are just other names for intentionality, in this case we have intentionally imposed intentionality on objects and states of affairs that are not intrinsically intentional. There are names in English for the result of this type of imposition of function: They are called "meaning" or "symbolism." Marks on the paper now have meaning in a way that a screwdriver, for example, does not have meaning, because the marks on the paper now stand for or represent objects and states of affairs independent of themselves. The most famous sorts of meaning are, of course, in language. In the use of language we impose a specific function, namely, that of representing, onto marks and sounds.

I said earlier that the capacity to impose functions on natural phenomena was remarkable, but equally remarkable is the fact that functions may be imposed quite unconsciously, and the func-

tions once imposed are often—so to speak—invisible. So, for example, money may simply have evolved without anyone ever thinking, "We are now imposing a new function on these objects"; and once money has evolved, people may use money to buy and sell without thinking about the logical structure of imposed function. However, for all cases of agentive function, someone must be capable of understanding what the thing is for, or the function could never be assigned. At least some of the participants in the system of exchange must understand, consciously or unconsciously, that money is to buy things with, screwdrivers are for driving screws, and so forth. If we assign a function that is totally apart from human intentions, it would have to fall in the category of nonagentive functions. Thus suppose someone says that the intended agentive function of money is to serve as a medium of exchange and a store of value, but money also serves the hidden, secret, unintended function of maintaining the system of power relationships in society. The first claim is about the intentionality of agentive function. The second claim is about nonagentive function. To see this, simply ask yourself what facts in the world would make each claim true. The first claim is made true by the intentionality with which agents use objects as money. They use it *for the purpose of* buying, selling, and storing value. The second claim, like the claim that the heart functions to pump blood, would be true if and only if there is a set of unintended causal relations and these serve some teleology, even if it is not a teleology shared by the speaker. Some social scientists speak of a distinction between manifest and latent function. If this distinction parallels the distinction I have been making, then manifest functions are agentive functions and latent functions are nonagentive.

To summarize these points, we have discovered three separate categories of the assignment of function. First, nonagentive functions: For example, the function of the heart is to pump blood. In general these nonagentive functions are naturally occurring. Second, agentive functions: For example, the function of a screwdriver is to install and remove screws. Third, within agentive

functions a special subclass, where the function assigned is that of intentionality: For example, the function of the sentence "Snow is white" is to represent, truly or falsely, the state of affairs that snow is white.[5]

Just to keep the terminology straight I will adopt the following conventions.

1. Since all functions are observer relative I will speak of all functions as *assigned* or equivalently as *imposed.*
2. Within the category of assigned functions some are *agentive* because they are matters of the use to which agents put entities, e.g., the function of bathtubs is to take baths in.
3. Within the category of assigned functions some are *nonagentive* because they are naturally occurring causal processes to which we have assigned a purpose, e.g., the function of the heart is to pump blood.
4. Within the category of agentive functions is a special category of those entities whose agentive function is to *symbolize, represent, stand for,* or—in general—to *mean* something or other.

Collective Intentionality

Many species of animals, our own especially, have a capacity for collective intentionality. By this I mean not only that they engage in cooperative behavior, but that they share intentional states such as beliefs, desires, and intentions. In addition to singular intentionality there is also collective intentionality. Obvious examples are cases where *I* am doing something only as part of *our* doing something. So if I am an offensive lineman playing in a football game, I might be blocking the defensive end, but I am blocking only as part of *our* executing a pass play. If I am a violinist in an orchestra I play *my* part in *our* performance of the symphony.

Even most forms of human conflict require collective intentionality. In order that two men should engage in a prizefight, for

example, there has to be collective intentionality at a higher level. They have to be cooperating in having a fight in order for each of them to try to beat the other up. In this respect, prizefighting differs from simply beating up someone in an alley. The man who creeps up behind another man in an alley and assaults him is not engaging in collective behavior. But two prizefighters, as well as opposing litigants in a court case, and even two faculty members trading insults at a cocktail party, are all engaged in cooperative collective behavior at a higher level, within which the antagonistic hostile behavior can take place. An understanding of collective intentionality is essential to understanding social facts.

What is the relation between singular and collective intentionality, between, for example, the facts described by "I intend" and "We intend"? Most efforts I have seen to answer this question try to reduce "We intentionality" to "I intentionality" plus something else, usually mutual beliefs. The idea is that if we intend to do something together, then that consists in the fact that I intend to do it in the belief that you also intend to do it; and you intend to do it in the belief that I also intend to do it. And each believes that the other has these beliefs, and has these beliefs about these beliefs, and these beliefs about these beliefs about these beliefs . . . etc., in a potentially infinite hierarchy of beliefs. "I believe that you believe that I believe that you believe that I believe. . . . ," and so on. In my view all these efforts to reduce collective intentionality to individual intentionality fail. Collective intentionality is a biologically primitive phenomenon that cannot be reduced to or eliminated in favor of something else. Every attempt at reducing "We intentionality" to "I intentionality" that I have seen is subject to counterexamples.[6]

There is a deep reason why collective intentionality cannot be reduced to individual intentionality. The problem with believing that you believe that I believe, etc., and you believing that I believe that you believe, etc., is that it does not add up to a sense of *collectivity.* No set of "I Consciousnesses," even supplemented with beliefs, adds up to a "We Consciousness." The crucial element in collective intentionality is a sense of doing (wanting, believing,

etc.) something together, and the individual intentionality that each person has is derived *from* the collective intentionality that they share. Thus, to go back to the earlier example of the football game, I do indeed have a singular intention to block the defensive end, but I have that intention only as part of our collective intention to execute a pass play.

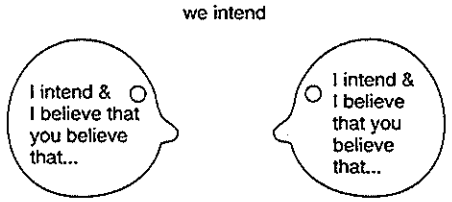
We can see these differences quite starkly if we contrast the case where there is genuine cooperative behavior with the cases where, so to speak, by accident two people happen to find that their behavior is synchronized. There is a big difference between two violinists playing in an orchestra, on the one hand, and on the other hand, discovering, while I am practicing my part, that someone in the next room is practicing her part, and thus discovering that, by chance, we are playing the same piece in a synchronized fashion.

Why are so many philosophers convinced that collective intentionality must be reducible to individual intentionality? Why are they unwilling to recognize collective intentionality as a primitive phenomenon? I believe the reason is that they accept an argument that looks appealing but is fallacious. The argument is that because all intentionality exists in the heads of individual human beings, the form of that intentionality can make reference only to the individuals in whose heads it exists. So it has seemed that anybody who recognizes collective intentionality as a primitive form of mental life must be committed to the idea that there exists some Hegelian world spirit, a collective consciousness, or something equally implausible. The requirements of methodological individualism seem to force us to reduce collective intentionality to individual intentionality. It has seemed, in short, that we have to choose between reductionism, on the one hand, or a super mind floating over individual minds, on the other. I want to claim, on the contrary, that the argument contains a fallacy and that the dilemma is a false one. It is indeed the case that all my mental life is inside my brain, and all your mental life is inside your brain, and so on for everybody else. But it does not follow from that that all my mental life must be expressed in the form of a singular

noun phrase referring to me. The form that my collective intentionality can take is simply "we intend," "we are doing so-and-so," and the like. In such cases, I intend only as part of our intending. The intentionality that exists in each individual head has the form "we intend."[7]

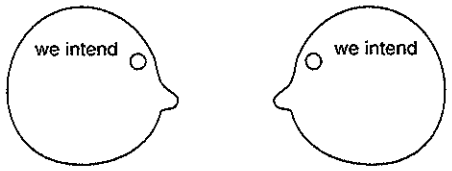
The traditional picture of "we intentions" looks like this:

Figure 1.1



The alternative that I am proposing looks like this:

Figure 1.2



By stipulation I will henceforth use the expression "social fact" to refer to any fact involving collective intentionality. So, for example, the fact that two people are going for a walk together is a social fact. A special subclass of social facts are institutional facts; facts involving human institutions. So, for example, the fact that this piece of paper is a twenty dollar bill is an institutional fact. I will have a great deal more to say about institutional facts.

Constitutive Rules and the Distinction Between Brute and Institutional Facts

In my work on the philosophy of language[8] I suggested the beginnings of an answer to the question concerning the relationships between those features of the world that are matters of brute physics and biology, on the one hand, and those features of the world that are matters of culture and society, on the other. Without implying that these are the only kinds of facts that exist in the world, we need to distinguish between *brute facts* such as the fact that the sun is ninety-three million miles from the earth and *institutional* facts such as the fact that Clinton is president. Brute facts exist independently of any human institutions; institutional facts can exist only within human institutions. Brute facts require the institution of language in order that we can *state* the facts, but the brute facts *themselves* exist quite independently of language or of any other institution. Thus the *statement* that the sun is ninety-three million miles from the earth requires an institution of language and an institution of measuring distances in miles, but the *fact stated,* the fact that there is a certain distance between the earth and the sun, exists independently of any institution. Institutional facts, on the other hand, require special human institutions for their very existence. Language is one such institution; indeed, it is a whole set of such institutions.

And what are these "institutions"? To answer this question, I introduced another distinction, the distinction between what I call "regulative" and "constitutive" rules.[9] Some rules regulate antecedently existing activities. For example, the rule "drive on the right-hand side of the road" regulates driving; but driving can exist prior to the existence of that rule. However, some rules do not merely regulate, they also create the very possibility of certain activities. Thus the rules of chess do not regulate an antecedently existing activity. It is not the case that there were a lot of people pushing bits of wood around on boards, and in order to prevent

them from bumping into each other all the time and creating traffic jams, we had to regulate the activity. Rather, the rules of chess create the very possibility of playing chess. The rules are *constitutive* of chess in the sense that playing chess is constituted in part by acting in accord with the rules. If you don't follow at least a large subset of the rules, you are not playing chess. The rules come in systems, and the rules individually, or sometimes the system collectively, characteristically have the form

"X counts as Y" or "X counts as Y in context C."

Thus, such and such counts as a checkmate, such and such a move counts as a legal pawn move, and so on.

The claim I made was, institutional facts exist only within systems of constitutive rules. The systems of rules create the possibility of facts of this type; and specific instances of institutional facts such as the fact that I won at chess or the fact that Clinton is president are created by the application of specific rules, rules for checkmate or for electing and swearing in presidents, for example. It is perhaps important to emphasize that I am discussing *rules* and not *conventions*. It is a rule of chess that we win the game by checkmating the king. It is a *convention* of chess that the king is larger than a pawn. "Convention" implies arbitrariness, but constitutive rules in general are not in that sense arbitrary.

The context "X counts as Y in C" is intensional-with-an-s. It is referentially opaque in that it does not permit of substitutability of coextensive expressions *salva veritate*. Thus, for example, the statements:

1. Bills issued by the Bureau of Engraving and Printing(X) count as money(Y) in the United States(C).

and

2. Money is the root of all evil.

do not imply

3. Bills issued by the Bureau of Engraving and Printing count as the root of all evil in the United States.

As always the discovery of referential opacity is a crucial point. In this case it provides a clue that there is a mental component in institutional facts. The intentionality-with-an-s of the verbal formulation is a clue that the phenomena represented are intentional-with-a-t. A great deal hangs on this, as we will see in subsequent chapters.

Various social theorists have attacked my account of the distinction between regulative and constitutive rules,[10] but I think my account is right as far as it goes. The problem is that for our present purposes it does not go far enough. We still need a more thorough account of rules and institutions. And we need to answer a lot of questions. Are all social facts institutional facts? Are there constitutive rules of, for example, wars and cocktail parties? What makes something into a "constitutive rule" anyway? Hardest of all, how do we make the connection between the fundamental ontology of conscious biological beasts like ourselves and the apparatus of social facts and human institutions?

I will have more to say later about the form of constitutive rules and how they relate to the ontology of institutional facts. My aim in this chapter is to assemble the pieces, and I now have the three I need: the imposition of function on entities that do not have that function prior to the imposition, collective intentionality, and the distinction between constitutive and regulative rules. With these in hand we can now turn to the construction of institutional reality.