

PENGUIN PHILOSOPHY

Money, marriage, property and government are facts, yet they exist because we believe they exist.

In this fascinating, provocative account, eminent philosopher John Searle shows how our everyday actions and cultural knowledge are of a metaphysical character that is truly staggering. He explores the character of the structures of our world that exist by human agreement and, from this, the nature of objective reality.

For example, how can it be a completely objective fact that coins are money, if something is money only because we believe it is money? And what is the role of language in constituting such facts?

In examining the difference between what can and cannot be socially constructed, he also shows how biology, which offers facts that are independent of human opinion and is often seen in opposition to the social sciences, forms the basis of these cultural and constitutional forms.

'Bursting with plain and necessary right-headedness'. He has performed a genuine service in bringing his rigour to bear on the lazy and pernicious relativism latent in the idea that reality is just what we humans choose it to be'

- Colin McGinn in the *New Republic*

'A brisk, bold and extremely clear book ... This is philosophy in the best modern American manner' - Geoffrey Hawthorn in the *London Review of Books*

Cover photograph © Ernst Haas



PENGUIN
Philosophy

ISBN 0-14-023590-6



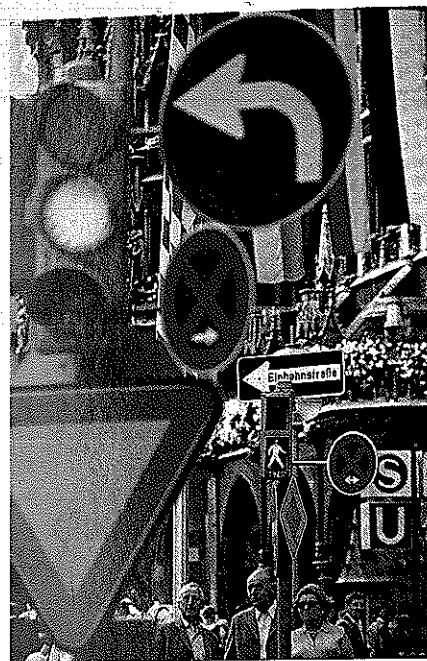
9 780140 235906



9 0201

U.K. £9.99

Ustřední knihovna FSS MU Brno
4240734114



The Construction of Social Reality

'A fascinatingly complex and rewarding discussion of basic and humanly constructed realities'

- Richard Hoggart in the *Guardian*, Books of the Year

JOHN R. SEARLE

Acknowledgments

The first version of these ideas was presented as the Immanuel Kant lectures in Stanford in 1992. Subsequent versions were presented as the Thalheimer Lectures at Johns Hopkins, the Hempel lectures at Princeton, and as a series of lectures at the College de France in Paris. I have also presented this material in seminars in Berkeley and at the University of Graz in Austria. Several of my colleagues read parts of the manuscript and made helpful criticisms. Special thanks are due to Kent Bach, Martin Jones, Lisa Lloyd, Brian McLaughlin, Stephen Neale, and Neil Smelser.

In addition to the lecture series and university courses just mentioned, I have also had the opportunity to try out some of these ideas in several universities in the United States and Europe. We often hear how dreadful contemporary intellectual life is, but I have to say from my own experience that one of the great pleasures of the present era is that one can go just about anywhere in the world and lecture, in English, to audiences that are sympathetic, intelligent, helpful and sophisticated in analytic philosophy. I cannot exaggerate the extent to which I have benefited from the comments of students, friends, colleagues, and total strangers. I really can't thank all of the people who made helpful comments, simply because I do not remember all of them. Among those I do remember, I am especially grateful to Pierre Bourdieu, Herman Capellen, Hubert Dreyfus, Gilbert Harman,

Robert Harnish, Meleana Isaacs, Saul Kripke, Francois Recanati, David Sosa, and Charles Spinosa.

For exceptional hospitality during the course of writing this book, I am grateful to Ann and Gordon Getty and Drue Heinz. Thanks are also due to the entire passenger list of the *Midnight Saga* and the *Rosenkavalier* for putting up so graciously with my relentless pounding on the computer.

Special thanks to my research assistant, Jennifer Hudin, who was helpful at every step of the way, from the earliest formulation of the basic ideas to the final preparation of the index. As always my greatest debts are to my wife, Dagmar Searle, to whom this book is dedicated.

Introduction

We live in exactly one world, not two or three or seventeen. As far as we currently know, the most fundamental features of that world are as described by physics, chemistry, and the other natural sciences. But the existence of phenomena that are not in any obvious way physical or chemical gives rise to puzzlement. How, for example, can there be states of consciousness or meaningful speech acts as parts of the physical world? Many of the philosophical problems that most interest me have to do with how the various parts of the world relate to each other—how does it all hang together?—and much of my work in philosophy has been addressed to these questions. The theory of speech acts is in part an attempt to answer the question, How do we get from the physics of utterances to meaningful speech acts performed by speakers and writers? The theory of the mind I have attempted to develop is in large part an attempt to answer the question, How does a mental reality, a world of consciousness, intentionality, and other mental phenomena, fit into a world consisting entirely of physical particles in fields of force? This book extends the investigation to social reality: How can there be an objective world of money, property, marriage, governments, elections, football games, cocktail parties and law courts in a world that consists entirely of physical particles in fields of force, and in which some of

these particles are organized into systems that are conscious biological beasts, such as ourselves?

Because these questions concern what might be thought of as problems in the foundations of the social sciences, one might suppose that they would have been addressed and solved already in the various social sciences, and in particular by the great founders of the social sciences in the nineteenth century and the early parts of the twentieth century. I am certainly no expert on this literature, but as far as I can tell, the questions I am addressing in this book have not been satisfactorily answered in the social sciences. We are much in debt to the great philosopher-sociologists of the nineteenth and early twentieth centuries—one thinks especially of Weber, Simmel, and Durkheim—but from such acquaintance with their works as I have, it seems to me that they were not in a position to answer the questions that puzzle me, because they did not have the necessary tools. That is, through no fault of their own, they lacked an adequate theory of speech acts, of performatives, of intentionality, of collective intentionality, of rule-governed behavior, etc. This book is an attempt to answer a set of traditional questions using resources that I and others have developed while working on other related questions.

A word about the organization of the book. The main argument is in the first half, Chapters 1 through 5. In these chapters I attempt to develop a general theory of the ontology of social facts and social institutions. The main question is, How do we construct an objective social reality? I apologize for a certain amount of repetition in these chapters, but in the nature of the case I was forced to go over and over the same ground to try to make sure I was getting it right. In Chapter 6 I try to locate the explanatory force of the constitutive rules of human institutions, given the puzzling fact that the agents in question are typically unconscious of the rules. To do that I have to explain my notion of the "Background" of nonconscious nonrepresentational capacities and abilities that enable us to cope with the world. In early drafts of the book I devoted an initial chapter to defending realism, the

idea that there is a real world independent of our thought and talk, and to defending the correspondence conception of truth, the idea that our true statements are typically made true by how things are in the real world that exists independently of the statements. I think that realism and a correspondence conception are essential presuppositions of any sane philosophy, not to mention of any science, and I wanted to make clear some of my reasons for thinking so. But what was originally intended as fairly short introductory material developed a life of its own, as is usually the case with such large philosophical questions. When the first chapter grew to three I decided to move all of this material to the back of the book, lest it overbalance my main argument. Chapters 7 and 8 are discussions of realism, Chapter 9 is a defense of a version of the correspondence conception of truth.

The Building Blocks of Social Reality

The Metaphysical Burden of Social Reality

This book is about a problem that has puzzled me for a long time: there are portions of the real world, objective facts in the world, that are only facts by human agreement. In a sense there are things that exist only because we believe them to exist. I am thinking of things like money, property, governments, and marriages. Yet many facts regarding these things are "objective" facts in the sense that they are not a matter of your or my preferences, evaluations, or moral attitudes. I am thinking of such facts as that I am a citizen of the United States, that the piece of paper in my pocket is a five dollar bill, that my younger sister got married on December 14, that I own a piece of property in Berkeley, and that the New York Giants won the 1991 superbowl. These contrast with such

facts as that Mount Everest has snow and ice near the summit or that hydrogen atoms have one electron, which are facts totally independent of any human opinions. Years ago I baptized some of the facts dependent on human agreement as "institutional facts," in contrast to noninstitutional, or "brute," facts.¹ Institutional facts are so called because they require human institutions for their existence. In order that this piece of paper should be a five dollar bill, for example, there has to be the human institution of money. Brute facts require no human institutions for their existence. Of course, in order to *state* a brute fact we require the institution of language, but the *fact stated* needs to be distinguished from the *statement* of it.

The question that has puzzled me is, How are institutional facts possible? And what exactly is the structure of such facts? But in the intervening years some curious things have happened. Many people, including even a few whose opinions I respect, have argued that all of reality is somehow a human creation, that there are no brute facts, but only facts dependent on the human mind. Furthermore, several people have argued against our common-sense idea that there are facts in the world that make our statements true and that statements are true because they correspond to the facts. So after attempting to answer my original question, How is a socially constructed reality possible? I want also to defend the contrast on which the question rests. I want to defend the idea that there is a reality that is totally independent of us (Chapters 7 and 8). Furthermore, because my method of investigation is to examine the structure of the facts that make our statements true and to which they correspond when they are true, I will also defend (a version of) the correspondence theory of truth (Chapter 9). The last three chapters, therefore, are concerned with defending certain general assumptions about reality, representation, knowledge, and truth.

Some of the questions I am trying to answer in the main argument of the book (Chapters 1–6) are, How can there be an objective reality that exists in part by human agreement? For example,

how can it be a completely objective fact that the bits of paper in my pocket are money, if something is money only because we believe it is money? And what is the role of language in constituting such facts?

To give you a feel for the complexity of the problem, I want to begin by considering the metaphysics of ordinary social relations. Consider a simple scene like the following. I go into a café in Paris and sit in a chair at a table. The waiter comes and I utter a fragment of a French sentence. I say, "*un demi, Munich, à pression, s'il vous plaît.*" The waiter brings the beer and I drink it. I leave some money on the table and leave. An innocent scene, but its metaphysical complexity is truly staggering, and its complexity would have taken Kant's breath away if he had ever bothered to think about such things.* Notice that we cannot capture the features of the description I have just given in the language of physics and chemistry. There is no physical-chemical description adequate to define "restaurant," "waiter," "sentence of French," "money," or even "chair" and "table," even though all restaurants, waiters, sentences of French, money, and chairs and tables are physical phenomena. Notice, furthermore, that the scene as described has a huge, invisible ontology: the waiter did not actually own the beer he gave me, but he is employed by the restaurant, which owned it. The restaurant is required to post a list of the prices of all the *boissons*, and even if I never see such a list, I am required to pay only the listed price. The owner of the restaurant is licensed by the French government to operate it. As such, he is subject to a thousand rules and regulations I know nothing about. I am entitled to be there in the first place only because I am a citizen of the United States, the bearer of a valid passport, and I have entered France legally.

*Kant did not bother to think about such things because in his era philosophers were obsessed with knowledge. Much later, for a brief, glorious moment, they were obsessed with language. Now this philosopher at least is obsessed with certain general structural features of human culture.

Notice, furthermore, that though my description was intended to be as neutral as possible, the vocabulary automatically introduces normative criteria of assessment. Waiters can be competent or incompetent, honest or dishonest, rude or polite. Beer can be sour, flat, tasty, too warm, or simply delicious. Restaurants can be elegant, ugly, refined, vulgar, or out of fashion, and so on with the chairs and tables, the money, and the French phrases.

If, after leaving the restaurant, I then go to listen to a lecture or attend a party, the size of the metaphysical burden I am carrying only increases; and one sometimes wonders how anyone can bear it.

The Invisible Structure of Social Reality

One reason we can bear the burden is that the complex structure of social reality is, so to speak, weightless and invisible. The child is brought up in a culture where he or she simply takes social reality for granted. We learn to perceive and use cars, bathtubs, houses, money, restaurants, and schools without reflecting on the special features of their ontology and without being aware that they have a special ontology. They seem as natural to us as stones and water and trees. Indeed, if anything, in most cases it is harder to see objects as just natural phenomena, stripped of their functional roles, than it is to see our surroundings in terms of their socially defined functions. So children learn to see moving cars, dollar bills, and full bathtubs; and it is only by force of abstraction that they can see these as masses of metal in linear trajectories, cellulose fibers with green and gray stains, or enamel-covered iron concavities containing water.

The complex ontology seems simple; the simple ontology seems difficult. This is because social reality is created by us for our purposes and seems as readily intelligible to us as those purposes themselves. Cars are for driving; dollars for earning, spending, and saving; bathtubs for taking a bath. But once there is no

function, no answer to the question, What's it for? we are left with a harder intellectual task of identifying things in terms of their intrinsic features without reference to our interests, purposes, and goals.

The invisibility of the structure of social reality also creates a problem for the analyst. We cannot just describe how it seems to us from an internal "phenomenological" point of view, because money, property, marriages, lawyers, and bathtubs do not seem to have a complex structure. They just are what they are, or so it seems. Nor can we describe them from the external behaviorist point of view, because the description of the overt behavior of people dealing with money, property, etc., misses the underlying structures that make the behavior possible. Nor, in turn, can we describe those structures as sets of unconscious computational rules, as is done by contemporary cognitive science and linguistics, because it is incoherent to postulate an unconscious following of rules that is inaccessible in principle to consciousness. And besides, computation is one of those observer-relative, functional phenomena we are seeking to explain.²

If neither the internal phenomenological nor the external behaviorist point of view is adequate, what then is the correct stance, the correct methodology, for describing the *structure* of social reality? To start with, in this chapter and the next, I will use a first-person intentionalistic vocabulary to try to lay bare certain elementary features of social ontology. Later, in Chapter 6, I will show how some, though not all, of the intentionalistic apparatus can be explained in terms of, and ultimately eliminated in favor of, what I have elsewhere called the "Background" of capacities, abilities, tendencies, and dispositions.

Fundamental Ontology

Since our investigation is ontological, i.e., about how social facts exist, we need to figure out how social reality fits into our overall ontology, i.e., how the existence of social facts relates to other

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, H₂O 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 ob-

jective 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 ob-

server 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 yourself, 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 assess 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 article³ 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 pump-

ing 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 preser-

*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 in-

vation 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.

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*.

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*.

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.

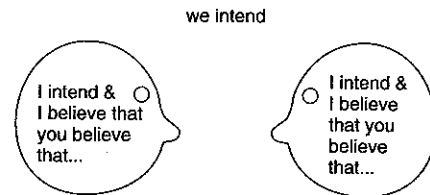
We can see these differences quite starkly if we contrast the case where there is genuine cooperative behavior with the cases where, so to speak, by accident two people happen to find that their behavior is synchronized. There is a big difference between two violinists playing in an orchestra, on the 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

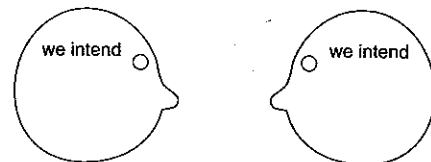
The traditional picture of "we intentions" looks like this:

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 language⁸ 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.

Creating Institutional Facts

In this chapter I describe the elementary construction of social facts and the logical structure of the development of institutional facts from simpler forms of social facts. To do so, I will use the apparatus of agentive functions, collective intentionality, and constitutive rules. I will also attempt to explain several puzzling features of social reality.

Some Apparent Features of Social Reality

To begin, let us identify some of the apparent features of social reality we would like to explain. Because I believe philosophical investigations should begin naively (how they proceed and conclude is another matter), I will simply list half a dozen of what appear to be naive, intuitive features of social reality, including features of institutional facts, such as, for example, the fact that I am an American citizen, as well as features of those social facts that do

not require institutional structures, such as, for example, the fact that two men are pushing a car together to try to get it started.

1. *The Self-Referentiality of Many Social Concepts*

The concepts that name social facts appear to have a peculiar kind of self-referentiality. As a preliminary formulation we can say, for example, in order that the concept "money" apply to the stuff in my pocket, it has to be the sort of thing that people think is money. If everybody stops believing it is money, it ceases to function as money, and eventually ceases to be money. Logically speaking, the statement "A certain type of substance, *x*, is money" implies an indefinite inclusive disjunction of the form "*x* is used as money or *x* is regarded as money or *x* is believed to be money, etc." But that seems to have the consequence that the concept of money, the very definition of the word "money," is self-referential, because in order that a type of thing should satisfy the definition, in order that it should fall under the concept of money, it must be believed to be, or used as, or regarded as, etc., satisfying the definition. For these sorts of facts, it seems to be almost a logical truth that you cannot fool all the people all the time. If everybody always thinks that this sort of thing is money, and they use it as money and treat it as money, then it is money. If nobody ever thinks this sort of thing is money, then it is not money. And what goes for money goes for elections, private property, wars, voting, promises, marriages, buying and selling, political offices, and so on.

In order to state this point precisely we need to distinguish between institutions and general practices on the one hand and particular instances on the other, that is, we need to distinguish between types and tokens. A single dollar bill might fall from the printing presses into the cracks of the floor and never be used or thought of as money at all, but it would still be money. In such a case a particular token instance would be money, even though no one ever thought it was money or thought about it or used it at all. Similarly, there might be a counterfeit dollar bill in circulation

even if no one ever knew that it was counterfeit, not even the counterfeiter. In such a case everyone who used that particular token would think it was money even though it was not in fact money. About particular tokens it is possible for people to be systematically mistaken. But where the *type* of thing is concerned, the belief that the type is a type of money is constitutive of its being money in a way we will need to make fully clear.

For some institutional phenomena, such as money, what I say applies more to types than tokens, for others, such as cocktail parties, it applies to each individual token. For the sake of simplicity I will assume that the reader is aware of the distinction, and I will speak of the self-referentiality of institutional concepts in general, without making the distinction in every case. Later I will try to explain the difference between self-referentiality as applied to types and as applied to tokens.

But if the *type* of thing in question is money only because people believe it to be money, if "money" implies "regarded as, used as, or believed to be money," then philosophers will get worried, because the claim seems to produce either a vicious infinite regress or a vicious circle. If part of the content of the claim that something is money is the claim that it is believed to be money, then what is the content of that belief? If the content of the belief that something is money contains in part the belief that it is money, then the belief that something is money is in part the belief that it is believed to be money; and there is, in turn, no way to explain the content of that belief without repeating the same feature over and over again. Later on, I will try to show how to avoid this infinite regress. At this point, I am just calling attention to a peculiar logical feature that distinguishes social concepts from such natural concepts as "mountain" or "molecule." Something can be a mountain even if no one believes it is a mountain; something can be a molecule even if no one thinks anything at all about it. But for social facts, the attitude that we take toward the phenomenon is partly constitutive of the phenomenon. If, for example, we give a big cocktail party, and invite everyone in Paris, and if

things get out of hand, and it turns out that the casualty rate is greater than the Battle of Austerlitz—all the same, it is not a war; it is just one amazing cocktail party. Part of being a cocktail party is being thought to be a cocktail party; part of being a war is being thought to be a war. This is a remarkable feature of social facts; it has no analogue among physical facts.

2. *The Use of Performative Utterances in the Creation of Institutional Facts*

One of the most fascinating features of institutional facts is that a very larger number, though by no means all of them, can be created by explicit performative utterances. Performatives are members of the class of speech acts I call "declarations."¹ In declarations the state of affairs represented by the propositional content of the speech act is brought into existence by the successful performance of that very speech act. Institutional facts can be created with the performative utterance of such sentences as "The meeting is adjourned," "I give and bequeath my entire fortune to my nephew," "I appoint you chairman," "War is hereby declared," etc. These utterances create the very state of affairs that they represent; and in each case, the state of affairs is an institutional fact.

3. *The Logical Priority of Brute Facts over Institutional Facts*

Intuitively it seems there are no institutional facts without brute facts. For example, just about any sort of substance can be money, but money has to exist in some physical form or other. Money can be bits of metal, slips of paper, wampum, or entries in books. In fact, most of our money in the past couple of decades underwent a revolutionary physical transformation that we did not even notice. Most money is now in the form of magnetic traces on computer disks. It does not matter what the form is as long as it can

function as money, but money must come in some physical form or other.

What is true of money is true of chess games, elections, and universities. All these can take different forms, but for each there must be some physical realization. This suggests what I think is true, that social facts in general, and institutional facts especially, are hierarchically structured. Institutional facts exist, so to speak, on top of brute physical facts. Often, the brute facts will not be manifested as physical objects but as sounds coming out of peoples' mouths or as marks on paper—or even thoughts in their heads.

4. *Systematic Relationships Among Institutional Facts.*

An institutional fact cannot exist in isolation but only in a set of systematic relations to other facts. Thus, for example, in order that anybody in a society could have money, that society must have a system of exchanging goods and services for money. But in order that it can have a system of exchange, it must have a system of property and property ownership. Similarly, in order that societies should have marriages, they must have some form of contractual relationships. But in order that they can have contractual relationships, they have to understand such things as promises and obligations.

Furthermore, quite apart from the logical or conceptual requirement of interrelationships of institutional facts, it just turns out that in any real life situation one will find oneself in a complex of interlocking institutional realities. The restaurant scene described in Chapter 1 illustrates this: at any instant in the scene, one is (at least) a citizen, an owner of money, a client, a bill payer; and one is dealing with property, a restaurant, a waiter, a bill.

It might seem that games are counterexamples to this general principle, because, of course, games are designed to be forms of activity that do not connect with the rest of our lives in a way that

institutional facts characteristically do. Today's philosophy department softball game need have no consequences for tomorrow, in a way that today's wars, revolutions, buyings, and sellings are intended precisely to have consequences for tomorrow and into the indefinite future.* Nonetheless, even in the case of games, there are systematic dependencies on other forms of institutional facts. The position of the pitcher, the catcher, and the batter, for example, all involve rights and responsibilities; and their positions and actions or inactions are unintelligible without an understanding of these rights and responsibilities; but these notions are in turn unintelligible without the general notion of rights and responsibilities.

5. *The Primacy of Social Acts over Social Objects, of Processes over Products*

It is tempting to think of *social objects* as independently existing entities on analogy with the objects studied by the natural sciences. It is tempting to think that a government or a dollar bill or a contract is an object or entity in the sense that a DNA molecule, a tectonic plate, or a planet is an object or entity. In the case of social objects, however, the grammar of the noun phrases conceals from us the fact that, in such cases, process is prior to product. Social objects are always, in some sense we will need to explain, constituted by social acts; and, in a sense, *the object is just the continuous possibility of the activity*. A twenty dollar bill, for example, is a standing possibility of paying for something.

*To the extent that professional sports have such consequences, they cease to be just games and become something more, e.g., big business.

6. *The Linguistic Component of Many Institutional Facts*

Related to features 1 and 2 is the further apparent feature that only beings that have a language or some more or less language-like system of representation can create most, perhaps all, institutional facts, because *the linguistic element appears to be partly constitutive of the fact*.

It is common, for example, to read that certain ant colonies have slaves or that beehives have queens. I think such manners of speaking are harmless metaphors, especially where the so called "social insects" are concerned, but it is important to keep reminding ourselves that for a community literally to have slaves or literally to have a queen, the participants would have to have the apparatus necessary to represent something as a queen or as a slave. Just behaving in certain ways, where behavior is construed solely in terms of bodily movements, is not sufficient for a community to have a queen or to have slaves. In addition, there would have to be a certain set of attitudes, beliefs, etc., on the part of the members of the community, and this would seem to require a system of representation such as language. Language seems to be essential not only to represent these facts to ourselves; but in a way that we will have to explain, the linguistic forms in question are partly constitutive of the facts. But what *exactly* is the role of language in the constitution of institutional facts? This is not an easy question, and we will devote the next chapter to answering it.

**From Collective Intentionality to Institutional Facts:
The Example of Money**

The simplest form of social facts involves simple forms of collective behavior. As I said earlier, I think the capacity for collective behavior is biologically innate, and the forms of collective intentionality cannot be eliminated or reduced to something else. For example, it takes no cultural apparatus, cultural conven-

tions, or language for animals to move together in a pack or to hunt together. When hyenas move in a pack to kill an isolated lion, no linguistic or cultural apparatus is necessary, even though the behavior of the hyenas is very skillfully coordinated and the hyenas are responsive not only to the lion but to each other. The selectional advantage of cooperative behavior is, I trust, obvious. Inclusive fitness is increased by cooperating with conspecifics.

The only tricky feature of assimilating collective animal behavior into a general theory of intentionality derives from the fact that in any complex form of behavior, such as the example of hyenas attacking a lion, each animal's individual contribution to the collective behavior will have a different intentional content from the collective intentionality. In the case of humans, for example, if our team is executing a pass play, and my assignment is to block the defensive end, then my individual intentionality is, "I am blocking the defensive end"; but that has a different content from the collective intentionality, "We are executing a pass play," even though I am blocking the defensive end only as part of our executing the pass play. The content of the individual intentionality, then, may vary from the content of the collective intentionality, even though the individual's intentionality is part of the collective. It takes two to tango and more than two to execute a pass play.² As a step in developing a hierarchical taxonomy of social and institutional reality, I have already stipulated that any fact involving collective intentionality is a social fact. Thus, for example, hyenas hunting a lion and Congress passing legislation are both cases of social facts. Institutional facts, it will turn out, are a special subclass of social facts. Congress passing legislation is an institutional fact; hyenas hunting a lion is not.

The next step is the introduction of agentive functions of a collective sort. Given an apparatus that includes both collective intentionality and the intentional imposition of agentive functions on physical objects, it is no big step to combine the two. If it is easy

to see how a single person might decide to use some object as a chair or a lever, then I believe it is not difficult to see how two or more people together could decide to use some object as a bench on which they can all sit or to use something as a lever to be operated by several people, rather than just one. Collective intentionality can generate agentive functions as easily as individual intentionality.

The next step is more difficult because it involves the collective imposition of functions on objects where the function assigned to the object cannot be performed solely in virtue of the object's intrinsic physical features, as was the case for a log used as a bench, or a stick used as a lever. In this next type of case, the function is itself performed only as a matter of human cooperation. We will see in some detail that this step, the collective imposition of function, where the function can be performed only in virtue of collective agreement or acceptance, is a crucial element in the creation of institutional facts.

Consider for example a primitive tribe that initially builds a wall around its territory. The wall is an instance of a function imposed in virtue of sheer physics: the wall, we will suppose, is big enough to keep intruders out and the members of the tribe in. But suppose the wall gradually evolves from being a physical barrier to being a symbolic barrier. Imagine that the wall gradually decays so that the only thing left is a line of stones. But imagine that the inhabitants and their neighbors continue to *recognize* the line of stones as marking the boundary of the territory in such a way that it affects their behavior. For example, the inhabitants only cross the boundary under special conditions, and outsiders can only cross into the territory if it is acceptable to the inhabitants. The line of stones now has a function that is not performed in virtue of sheer physics but in virtue of collective intentionality. Unlike a high wall or a moat, the wall remnant cannot keep people out simply because of its physical constitution. The result is, in a very primitive sense, symbolic; because a set of physical objects now

performs the function of indicating something beyond itself, namely, the limits of the territory.* The line of stones performs the same *function* as a physical barrier but it does not do so in virtue of its physical construction, but because it has been collectively assigned a new *status*, the status of a boundary marker.

I would like this step to seem a most natural and innocent development, but it is momentous in its implications. Animals can impose functions on natural phenomena. Consider, for example, the primates that use a stick as a tool to get bananas that are out of reach.³ And some primates have even developed traditions of agentive functions that are transmitted from one generation to the next. Thus, most famously Imo, a Japanese macaque, used water to get the sand off her potatoes and eventually salt water both to get the sand off and to improve the flavor. Thanks to Imo, "today," writes Kummer, "potato-washing in salt water is an established tradition which infants learn from their mother as a natural adjunct of eating potatoes."⁴ Anthropology texts routinely remark on the human capacity for tool using. But the truly radical break with other forms of life comes when humans, through collective intentionality, impose functions on phenomena where the function cannot be achieved solely in virtue of physics and chemistry but requires continued human cooperation in the specific forms of recognition, acceptance, and acknowledgment of a new *status* to which a *function* is assigned. This is the beginning point of all institutional forms of human culture, and it must always have the structure X counts as Y in C, as we shall see later.

*In an earlier version of this argument, I used the ethologists' example of groups of animals marking limits to their territory. In such a case, as in the example of the primitive tribe, the barrier is not a sheer physical obstacle like a wall or a moat but is, in some sense, symbolic. But I am not certain that the ethologists are justified in attributing so much collective intentionality to the animals, so I have substituted the tribal example to make the same point. When we discuss the role of language in the next chapter we will see that the distinction between the linguistic and the prelinguistic is important.

Our aim is to assimilate social reality to our basic ontology of physics, chemistry, and biology. To do this we need to show the continuous line that goes from molecules and mountains to screwdrivers, levers, and beautiful sunsets, and then to legislatures, money, and nation-states. The central span on the bridge from physics to society is collective intentionality, and the decisive movement on that bridge in the creation of social reality is the collective intentional imposition of function on entities that cannot perform those functions without that imposition. The radical movement that gets us from such simple social facts as that we are sitting on a bench together or having a fistfight to such institutional facts as money, property, and marriage is the collective imposition of function on entities, which—unlike levers, benches, and cars—cannot perform the functions solely by virtue of their physical structure. In some cases, paper currency, for example, this is because the structure is only incidentally related to the function; in other cases, licensed drivers, for example, it is because we do not allow people to perform the function of driving unless they have been *authorized*.

The key element in the move from the collective imposition of function to the creation of institutional facts is the imposition of a collectively recognized *status* to which a function is attached. Since this is a special category of agentive functions, I will label these *status functions*. In the case of the boundary, we imagined a causally functioning physical object, a wall, evolving into a symbolic object, a boundary marker. The boundary is intended to function in the same way that the wall did, but the means by which it performs this function is the collective recognition of the stones as having a special status to which the function is attached. In the extreme case, the status function may be attached to an entity whose physical structure is only arbitrarily related to the performance of the function. As an illustration, consider the case of money and especially the evolution of paper currency. Standard textbook accounts of money identify three kinds: *commodity money*, such as gold, is regarded as valuable, and hence as money,

because the commodity itself is regarded as valuable; *contract money* consists of bits of paper that are regarded as valuable because they are promissory notes to pay the bearer in valuable commodities such as gold; and *fiat money* consists of bits of paper that are declared to be valuable as money by some official agency such as a government or a central bank. So far, though, it is not clear what the relationship among these three is, or even what fact about all three makes it the case that they are all money. In the case of commodity money the stuff is a medium of exchange because it is valuable; in the case of fiat money the stuff is valuable because it is a medium of exchange.

The logical relations among these three can be illustrated by the standard account of the evolution of paper currency in medieval Europe. I will assume this account is true, but it does not really matter much for our present purposes. I am using the account only to illustrate certain logical relations, which do not depend on its historical accuracy. Here is how it goes. The use of commodity money, such as gold and silver, is, in effect, a form of barter, because the form that the money takes is regarded as itself valuable. Thus the substance in question performs the function of money solely because of its physical nature, which will typically already have some function imposed on it. Thus, gold coins are valuable not because they are coins but because they are made of gold, and the value attached to the coin is exactly equal to the value attached to the gold in it. We impose the function of "value" on the substance gold because we desire to possess that kind of substance. Because the function of value has already been imposed on gold, it is easy to impose the function of money on top of the function of value. And that is just a fancy way of saying that because people already regard gold as valuable because of its physical nature, they are willing to accept it as a medium of exchange. We thus have a system of exchange where objects are held for the purposes of barter, even though the people holding those objects may have no interest in them or use for them, as such. A similar situation existed, by the way, in the former Soviet Union at the

time of its collapse. In Moscow, in 1990 and 1991, packs of Marlboro cigarettes had attained the status of a kind of currency. People would accept payment in Marlboros, even though they did not themselves smoke. The combination of paper and tobacco already had an agentive function, named by the word "cigarette," and on top of that function was imposed the agentive function named by "medium of exchange."

The story told about medieval Europe is that bankers would accept gold and store it for safekeeping, and in return for the gold they issued paper certificates to the depositors of the gold. The certificates then could be used as a medium of exchange, just as the gold itself was. The certificate was a kind of substitute for the gold. It had complete credibility as an object of value, because at any point, it was exchangeable for gold. Commodity money had thus been replaced by contract money.

A stroke of genius occurred when somebody figured out that we can increase the supply of money simply by issuing more certificates than we have gold. As long as the certificates continue to function, as long as they have a collectively imposed function that continues to be collectively accepted, the certificates are, as they say, as good as gold. The next stroke of genius came when somebody figured out—and it took a long time for people to figure this out—we can forget about the gold and just have the certificates. With this change we have arrived at fiat money, and that is the situation we are in today. On old Federal Reserve notes it said we could take the bill to the Treasury and they would "pay the bearer" the equivalent in "dollars." But suppose we gave them a twenty dollar Federal Reserve note, what exactly would they give us? Another twenty dollar Federal Reserve note!⁵

Constitutive Rules: X counts as Y in C

I think we can better understand what is going on in the evolution of money if we explore the relation of constitutive rules to the creation of institutional facts. I said that the form of the constitutive

rule was "X counts as Y in C"; but as I am using this locution, that only determines a set of institutional facts and institutional objects where the Y term names something more than the sheer physical features of the object named by the X term.⁶ Furthermore, the "counts as" locution names a feature of the imposition of a status to which a function is attached by way of collective intentionality, where the status and its accompanying function go beyond the sheer brute physical functions that can be assigned to physical objects. So, for example, as I am using this formula, it would not be a statement of a constitutive rule to say "objects that are designed and used to be sat on by one person count as chairs," because satisfying the X term is already sufficient for satisfying the Y term, just from the definition of the word "chair." The "rule" does not add anything but a label, so it is not a constitutive rule. Furthermore, it does not express a constitutive rule to say "objects of a certain shape count as chairs," because the functions assigned can be assigned independently of any human agreement. If it has a certain kind of shape, we can use it as a chair regardless of what anyone else thinks. But when we say that such and such bits of paper count as money, we genuinely have a constitutive rule, because satisfying the X term, "such and such bits of paper," is not by itself sufficient for being money, nor does the X term specify causal features that would be sufficient to enable the stuff to function as money without human agreement. So the application of the constitutive rule introduces the following features: The Y term has to assign a new *status* that the object does not already have just in virtue of satisfying the X term; and there has to be collective agreement, or at least acceptance, both in the imposition of that status on the stuff referred to by the X term and about the function that goes with that status. Furthermore, because the physical features specified by the X term are insufficient by themselves to guarantee the fulfillment of the assigned function specified by the Y term, the new status and its attendant functions have to be the sort of things that can be constituted by collective agreement or acceptance.

Also, because the physical features specified by the X term are insufficient to guarantee success in fulfilling the assigned function, there must be *continued* collective acceptance or recognition of the validity of the assigned function; otherwise the function cannot be successfully performed. It is not enough, for example, that we agree with the original assignment, "This stuff is money"; we must continue to accept it as money or it will become worthless.

Our sense that there is an element of magic, a conjuring trick, a sleight of hand in the creation of institutional facts out of brute facts derives from the nonphysical, noncausal character of the relation of the X and Y terms in the structure where we simply *count* X things as Y things. In our toughest metaphysical moods we want to ask "But is an X really a Y?" For example, are these bits of paper really *money*? Is this piece of land really somebody's *private property*? Is making certain noises in a ceremony really *getting married*? Even, is making noises through the mouth really making a *statement* or a *promise*? Surely when you get down to brass tacks, these are not real facts. We do not have this sense of giddiness where the agentive function is performed entirely in virtue of physical features. Thus, we do not have any metaphysical doubts about whether or not this is really a screwdriver, or this is really a car, because the sheer physical features of the objects in question enable them to function as screwdrivers or cars.

At this point I am simply describing the structure whereby institutional reality actually works in real human societies. Because this step is crucial for my argument, I will go through it slowly, using the example of U.S. paper money; and since I hope to be able to generalize certain features of the example, I will list its most salient general characteristics. Certain sorts of bits of paper are widely circulated in the United States. These pieces of paper satisfy certain conditions that constitute satisfying the X term. The pieces must have particular material ingredients, and they must match a certain set of patterns (five dollar bill, ten dollar bill, etc.).

They must also be issued by the Bureau of Engraving and Printing under the authority of the U.S. Treasury. Anything that satisfies these conditions (X term) counts as money, i.e., U.S. paper currency (Y term). But to describe these bits of paper with the Y term "money" does more than provide a shorthand label for the features of the X term; it describes a new status, and that status, viz. money, has a set of functions attached to it, e.g., medium of exchange, store of value, etc. In virtue of the constitutive rule, the paper counts as "legal tender for all debts public and private." And the imposition of this status function by the Y term has to be collectively recognized and accepted or the function will not be performed.

Some of the most salient generalizable features of this example are as follows:

FIRST, collective intentionality assigns a new status to some phenomenon, where that status has an accompanying function that cannot be performed solely in virtue of the intrinsic physical features of the phenomenon in question. This assignment creates a new fact, an institutional fact, a new fact created by human agreement.

SECOND, the *form* of the assignment of the new status function can be represented by the formula "X counts as Y in C." This formula gives us a powerful tool for understanding the form of the creation of the new institutional fact, because the form of the collective intentionality is to impose that status and its function, specified by the Y term, on some phenomenon named by the X term. The "counts as" locution is crucial in this formula because since the function in question cannot be performed solely in virtue of the physical features of the X element, it requires our agreement or acceptance that it be performed. Thus, we agree to count the object named by the X term as having the status and function specified by the Y term. The sorts of functions and sta-

tuses that can be assigned by the Y term, therefore, are seriously limited by the possibilities of having functions where the performance of the function contains an element that can be guaranteed simply by collective agreement or acceptance. This is, perhaps, the most mysterious feature of institutional facts, and I will have a good deal to say about it later.

THIRD, the process of the creation of institutional facts may proceed without the participants being conscious that it is happening according to this form. The evolution may be such that the participants think, e.g., "I can exchange this for gold," "This is valuable," or even simply "This is money." They need not think, "We are collectively imposing a value on something that we do not regard as valuable because of its purely physical features," even though that is exactly what they are doing. There are two points about the relation of this process to consciousness. First, obviously, for most institutions we simply grow up in a culture where we take the institution for granted. We need not be consciously aware of its ontology. But second, and more to the point here, in the very evolution of the institution the participants need not be consciously aware of the form of the collective intentionality by which they are imposing functions on objects. In the course of consciously buying, selling, exchanging, etc., they may simply evolve institutional facts. Furthermore, in extreme cases they may accept the imposition of function only because of some related theory, which may not even be true. They may believe that it is money only if it is "backed by gold" or that it is a marriage only if it is sanctified by God or that so and so is the king only because he is divinely authorized. Throughout the history of the United States, literally millions of Americans have thought that the Constitution was divinely inspired. As long as people continue to recognize the X as having the Y status function, the institutional fact is created and maintained. They do not in addition have to recognize that they are so recognizing, and they may hold all sorts of

other false beliefs about what they are doing and why they are doing it.

FOURTH, where the imposition of status function according to the formula becomes a matter of general policy, the formula acquires a normative status. It becomes a constitutive *rule*. This is shown by the fact that the general rule creates the possibility of abuses that could not exist without the rule, such as counterfeit money (objects are designed to look as if they satisfy the X term, when they do not) and hyperinflation (too much money is issued, so that the objects satisfying the X term can no longer perform the function specified by the Y term). The possibility of such forms of abuse is characteristic of institutional facts. Thus, for example, the fact that attorneys have to be certified creates the possibility that those who are not certified can pretend that they are and thus pretend that they are attorneys. They are, so to speak, "counterfeit" attorneys. But even a person qualified as an attorney can abuse the position and so fail to perform the functions properly (malpractice). Another illustration is provided by the decay of the institution of knighthood during the Middle Ages. At first knights were required to be competent warriors, in charge of many men and owning a lot of horses, etc. When decay set in, many people who did not meet the criteria (X term) for becoming knights asked the king to make them knights (Y term) anyway. Though they didn't pass the tests, they, for example, insisted that because they came from such a good family, the requirements should be waived in their case. Furthermore, many people who did rightfully acquire the status of knight became unable to carry out the functions of knighthood. They no longer had the required number of horses, or the required sort of armor, or they were not in the physical condition necessary to carry out the tasks of knighthood.

Where money is concerned cultures vary with their emphasis on the X or the Y aspect. United States currency is explicit on the Y aspect. It says, "This note is legal tender for all debts public and private," but it says nothing about counterfeiting. French cur-

rency, on the other hand, contains a long statement about the X aspect, specifically about the illegality of and punishment for counterfeiting.* Italian currency makes the same X aspect point, but more succinctly: "La legge punisce i fabbricatore e gli spacciatori di biglietti falsi."

FIFTH, the relation of rule and convention, at least in this case, is reasonably clear. That objects can function as a medium of exchange is not a matter of convention but of rule. But *which objects* perform this function is a matter of convention. Analogously, in chess, the powers of the king are not a matter of convention but of rule. But *which shape* to impose those powers on is a matter of convention. Because in these cases the conditions laid down by the X term are only incidentally related to the function specified by the Y term, the selection of the X term is more or less arbitrary; and the resulting policy as to which types of things shall be used as, e.g., money or a king in chess, is a matter of convention. As we will see in later examples, often the features necessary for the applicability of the X term are essential to the performance of the Y term. Thus, for example, when it comes to being a certified surgeon, the authorization to perform surgery (Y term) has to be based on meeting certain medical criteria (X term). Nonetheless, even in these cases, there is an addition marked by the Y term that is not already present in the X term. The person in question now has the status, e.g., of certified surgeon.

It might seem that there are obvious counterexamples to the claim that the features of the X term are insufficient to guarantee the function named by the Y term. For example, when the president or a state governor declares an earthquake or a major fire to be a "disaster," surely, one might say, the brute facts about the

*L'article 139 du code pénal punit de la réclusion criminelle a perpétuité ceux qui auront contrefait ou falsifié les billets de banque autorisés par la loi, ainsi ce que ceux qui auront fait usage de ces billets contrefaits ou falsifiés, ceux qui les auront introduits en France seront punis de la même peine.

earthquake or fire are sufficient to qualify them as disasters in virtue of their physical features. There is nothing conventional about being an earthquake or a holocaust. But if one looks closely at these cases, even they illustrate the point. The function of a declared disaster is that the local victims qualify for such things as financial aid and low-interest loans, whereas fires and earthquakes by themselves do not generate money in virtue of their brute physical features and consequences.

A similar point can be made about the criminal law. The whole point of the criminal law is regulative, not constitutive. The point is to forbid, for example, certain antecedently existing forms of behavior such as killing. But to make the regulations work, there must be sanctions, and that requires the imposition of a new status on the person who violates the law. Thus the person who kills another (X term), under certain circumstances (C term), and is found guilty of so doing is now assigned the status of "convicted murderer" (Y term, and hence, institutional fact); and with that new status come the appropriate punishments. Thus the regulative "Thou shalt not kill" generates the appropriate constitutive "Killing, under certain circumstances, counts as murder, and murder counts as a crime punishable by death or imprisonment."

In many cases the X term is chosen precisely because it is supposed to have the features necessary to perform the function specified by the Y term. Thus, for example, each of the expressions "attorney," "physician," "president," and "cathedral" names a status with a function imposed on entities—graduates of law school or medical school, winners of certain sorts of elections, and large buildings capable of accommodating big church services and acting as the seat of a bishopric—precisely because they are supposed to be able to perform the Y functions implied by the status labels "attorney," "physician," "president," or "cathedral." But even in these cases, something is added by the Y term. The features specified by the X term are not themselves enough to guarantee the additional status and function specified by the Y term. The difference between attorneys and screwdrivers, for ex-

ample, is that the screwdriver just has the sheer physical structure to enable it to perform its function, but for the law school graduate to be an attorney, an additional authorization or certification is required to confer the status of attorney. Collective agreement about the possession of the status is constitutive of having the status, and having the status is essential to the performance of the function assigned to that status.

An interesting class of cases are those where the entity in question has *both* a causal agentive function and correlated status-function. Consider, for example, the actual fence on portions of the border between Mexico and the United States. It is supposed to function causally as a physical barrier to crossing the border. But it is also supposed to mark a national boundary, something one is not supposed to cross unless authorized. Even in this case the status-function is in addition to the physical function, even though they both have the same ultimate objective.

The point is that the Y term must assign some new status that the entities named by the X term do not already have, and this new status must be such that human agreement, acceptance, and other forms of collective intentionality are necessary and sufficient to create it. Now, you might think, that is not much of an apparatus to work with, but in fact, as we will see in detail, the mechanism is a powerful engine in the generation of social reality.

SIXTH, finally there is a special relation between the imposition of these status-functions and language. The labels that are a part of the Y expression, such as the label "money," are now partly constitutive of the fact created. Odd as it may sound, in the creation of money, the linguistically expressed concepts, such as "money," are now parts of the very facts we have created. I will explore this feature in the next chapter.

Why Self-Referentiality Does Not Result in Circularity

In my list of six apparent features of social reality that needed explanation, the first was a puzzle about how we can define "money," if part of the definition is "being thought of, or regarded as, or believed to be money." I asked: does this not lead to a circularity or infinite regress in any attempt to define the word, or even to give an explanation of the concept of money? But the resolution of the paradox is quite simple. The word "money" marks one node in a whole network of practices, the practices of owning, buying, selling, earning, paying for services, paying off debts, etc. As long as the object is regarded as having that role in the practices, we do not actually need the word "money" in the definition of money, so there is no circularity or infinite regress. The word "money" functions as a placeholder for the linguistic articulation of all these practices. To believe that something is money, one does not actually need the word "money." It is sufficient that one believes that the entities in question are media of exchange, repositories of value, payment for debts, salaries for services rendered, etc. And what goes for money goes for other institutional notions such as marriage, property, and speech acts such as promising, stating, ordering, etc. In short, the fact that a set of attitudes is partly constitutive of the truth conditions of a certain concept, and the fact that those attitudes would normally be summarized by using that very concept (e.g., thinking that something is money, thinking that those people are married), does not have the consequence that the word expressing that concept cannot be defined without circularity or infinite regress.

Although we do not need the concept "money" to define "money," and thus we avoid an immediate circularity, to explain the concept we do need other institutional concepts such as "buying," "selling," and "owing," and thus we avoided the vicious circularity only by expanding the circle by including other institutional

concepts. We are not trying to reduce the concept "money" to noninstitutional concepts.

I mentioned that there is a distinction between the self-referentiality of the concept as applied to types and as applied to tokens. Where money is concerned a particular token could be money even if no one thought it was money, but where cocktail parties are concerned if no one thinks of a particular event that it is a cocktail party, it is not a cocktail party. I think the reason we treat cocktail parties differently from money in this regard has to do with codification. In general, if the institution in question is codified in an "official" form, such as in the laws concerning money, then the self-referentiality in question is a feature of the type. If it is informal, uncoded, then the self-referentiality applies to each token. Codification specifies the features a token must have in order to be an instance of the type. Hence a token may have those features even if no one thinks about it, but the type is still defined in this self-referential way.

The self-referentiality we have been discussing is an immediate consequence of the nature of agentive functions. It is not peculiar to institutional facts. So, for example, in order that something be a chair, it has to function as a chair, and hence, it has to be thought of or used as a chair. Chairs are not abstract or symbolic in the way that money and property are, but the point is the same in both cases. Where agentive functional concepts are concerned, part of satisfying a description is being thought to satisfy that description. This does not lead to circularity or infinite regress for the reason just stated: We can cash out the description in terms of the set of practices in which the phenomenon is embedded. Chairs are for sitting in, money is to buy things with, tools are for manipulating objects in various ways, etc.*

*In the *Random House Dictionary*, one of the definitions given for "tool" is: "anything that can be used as tool." As a definition, that seems pretty dumb, but it is not quite as dumb as it looks. You could not define "screwdriver" as "anything that can be used as a screwdriver," because lots of things can be used as screw-

The Use of Performative Utterances in the Creation of Institutional Facts

The second apparent feature we need to explain concerns the role of performative utterances in the creation of many, though not all, institutional facts. The explanation is provided by the structure of constitutive rules. *In general, where the X term is a speech act, the constitutive rule will enable the speech act to be performed as a performative declaration creating the state of affairs described by the Y term.* Because saying certain things *counts as* entering into a contract or adjourning a meeting, you can perform those acts by saying you are performing them. If you are the chairman, then saying in appropriate circumstances "The meeting is adjourned" will make it the case that the meeting is adjourned. Saying, in appropriate circumstances, "I appoint you chairman" will make it the case that you are chairman. The same words said by the wrong person or in the wrong circumstances will have no such effect. Because the constitutive rule enables the function to be imposed on a speech act, then just performing that speech act in appropriate circumstances can constitute the imposition of that function, and thus will constitute a new institutional fact.

It is said that in Moslem countries a man can divorce his wife by simply saying "I divorce you" three times while throwing three white pebbles. This is clearly a performative use of the verb "divorce," which does not exist in other countries. Those who think that meaning is use would have to conclude that the word "divorce" has a different meaning for Moslems than it does for others. But that is not the case. What has happened is that a new status-function has been imposed on an existing sentence form. The sentence form "I divorce you" does not change its meaning when a new status-function is added; rather, it is now simply used

drivers that definitely are not screwdrivers, for instance, coins. But since "tool," unlike "screwdriver," names a very large class of agentive functions, anything that can be used as a tool is, roughly speaking, a tool.

in the creation of a new institutional fact, namely, the particular divorce, in virtue of a new constitutive rule according to which the husband's saying "I divorce you" three times with the appropriate throwing gestures counts as divorcing his wife. Thus the performative utterance creates a new institutional fact, the divorce.

Even the statement on the twenty dollar bill, though it contains no performative verbs, is a declaration. It says, "This note is legal tender for all debts, public and private." But that utterance is not an empirical claim. It will not do, for example, to ask the Treasury, "How do you know it is legal tender?" or "What's the evidence?" When the Treasury says it is legal tender, they are *declaring* it to be legal tender, not announcing an empirical fact that it already is legal tender.

The possibility of creating institutional facts by declaration does not hold for every institutional fact. You cannot, for example, make a touchdown just by saying you are making it.

To summarize this point: performatives play a special role in the creation of institutional facts, because the status-function marked by the Y term in the formula "X counts as Y" can often, though not always, be imposed simply by declaring it to be imposed. This is especially true where the X term is itself a speech act.

The Logical Priority of Brute Facts over Institutional Facts

The third apparent feature we need to explain concerns the priority of brute facts over institutional facts. As with feature two, this is explained by the structure of constitutive rules. The structure of institutional facts is the structure of hierarchies of the form "X counts as Y in context C." That hierarchy has to bottom out in phenomena whose existence is not a matter of human agreement. This is just another way of saying that where there is a status-function imposed on something, there has to be something it is imposed on. If it is imposed on another status-function, eventually

one has to reach a rock bottom of something that is not itself any form of status-function. So, for example, as I said earlier, all sorts of things can be money, but there has to be some physical realization, some brute fact—even if it is only a bit of paper or a blip on a computer disk—on which we can impose our institutional form of status function. Thus there are no institutional facts without brute facts.

This discussion anticipates a discussion of realism I will present in Chapters 7 and 8. It could not be the case, as some antirealists have maintained, that all facts are institutional facts, that there are no brute facts, because the analysis of the structure of institutional facts reveals that they are logically dependent on brute facts. To suppose that all facts are institutional would produce an infinite regress or circularity in the account of institutional facts. In order that some facts be institutional, there must be some other facts that are brute. This is a consequence of the logical structure of institutional facts.

Systematic Relations and the Primacy of the Act over the Object

Our fourth question was, Why are there always certain sorts of systematic relations among institutional facts? And the fifth was, Why do institutional acts seem prior to institutional objects?

The most obvious reason why there are systematic relationships among the various sorts of social facts of the type that I tried to describe is that the facts in question are designed for precisely that purpose. Governments are designed to impact on our lives in all sorts of ways; money is designed to provide a unit of value in all kinds of transactions. Even games, which are explicitly designed to be insulated from the rest of our lives, nonetheless employ an apparatus—of rights, obligations, responsibilities, etc.—that, as I remarked earlier, is intelligible only given all sorts of other social facts.

The explanation for the apparent primacy of social acts over so-

cial objects is that the “objects” are really designed to serve agentive functions, and have little interest for us otherwise. What we think of as social *objects*, such as governments, money, and universities, are in fact just placeholders for patterns of *activities*. I hope it is clear that the whole operation of agentive functions and collective intentionality is a matter of ongoing activities and the creation of the possibility of more ongoing activities.

Unconsciously, we have throughout this discussion been acknowledging this point by our talk of institutional *facts* rather than institutional *objects*. Such material objects as are involved in institutional reality, e.g., bits of paper, are objects like any others, but the imposition of status-functions on these objects creates a level of description of the object where it is an institutional object, e.g., a twenty dollar bill. The object is no different; rather, a new status with an accompanying function has been assigned to an old object (or a new object has been created solely for the purpose of serving the new status-function), but that function is manifested only in actual transactions; hence, our interest is not in the object but in the processes and events where the functions are manifested.

The priority of process over product also explains why, as several social theorists have pointed out, institutions are not worn out by continued use, but each use of the institution is in a sense a renewal of that institution. Cars and shirts wear out as we use them but constant use renews and strengthens institutions such as marriage, property, and universities. The account I have given explains this fact: since the function is imposed on a phenomenon that does not perform that function solely in virtue of its physical construction, but in terms of the continued collective intentionality of the users, each use of the institution is a renewed expression of the commitment of the users to the institution. Individual dollar bills wear out. But the institution of paper currency is reinforced by its continual use.

The sixth and final feature we need to explain concerns the role of language in institutional reality, and to that topic I devote the next chapter.