

Introduction

Let's do an experiment: let's look around and describe what we see.

If I look to the right, I see my colleague Antonio reviewing a paper for a scientific journal. On the left, a map of China hangs on the wall. In front of me, past the door of my office, two students are walking in the corridor of the Department of Economics. Out of the window, I can see the second floor of the Faculty of Social Science of the University of Milan.

You and I, of course, are surrounded by different things. But if we compare our lists, they will have something in common: most of the things that we see are institutional entities. An “institutional entity” is an object with properties or characteristics that depend on the existence of an institution. Antonio, for example, is a colleague of mine because we are both employees of the same university, and the University of Milan is an institution.

When I drew my list I could have used a different language, describing all the things that I saw in noninstitutional terms. I could have focused on their physical, chemical, or biological properties, for example. But undoubtedly such descriptions would have been incomplete: most of the things that surround us are not just physical or biological entities. A description of the world that ignored institutional entities and properties would miss much of what constitutes our world.

But what are these entities then? Institutions are as mysterious as they are ubiquitous. We can mention some examples:

the church, democracy, the army, the public school system. But as soon as we try to theorize, we face many difficult questions. What are these things? What do they have in common? What are they made of, and how do they work?

In the chapters that follow I will try to answer these questions. Philosophers sometimes refer to this sort of inquiry using the term “ontology.” This pompous name comes from the Greek words *on* and *logos*, meaning literally “theory or discourse about being.” Ontology asks what there is in the world, and social ontology, in particular, studies what there is in the social world.

Social ontology has been characterized for decades by a Babel of different approaches. This has certainly made it a stimulating field of research, but at the same time has also created major problems of communication. Researchers have found it difficult to appreciate the value of alternative approaches, and often have decided to start from scratch, ignoring decades if not centuries of work on the same topics.

Problems of communication emerge frequently at the crossroads between social science and philosophy. Institutions are a major topic of research for political scientists, economists, sociologists, and anthropologists, so we should expect them to be the experts in this area. But philosophers have often found the theories of social science unsatisfactory. The author of one of the most influential philosophical books of the past twenty years, for example, has stated boldly that he cannot find anything helpful in the *entire* social science literature. The “tradition” is inadequate, “the classical theorists have the direction of analysis back to front.” And this is true “not only of such foundational figures as Max Weber, Emil Durkheim, Georg Simmel, and Alfred Schutz, but of the whole Western tradition of discussing political and social institutions that goes back to Aristotle’s *Politics*, if not earlier” (Searle 2005: 2).

The dissatisfaction is reciprocal: social scientists have struggled to appreciate what the contribution of philosophy might

be. The book that I just mentioned for instance has been deemed “quite literally indifferent as sociology” (Osborne 1997: 98). And according to another reviewer, it shows “how big the hiatus between philosophy and the social sciences has become” (Kno-blauch 1996: 1461).

It is an odd situation. It is certainly possible that different scholars have different concerns. Perhaps they approach the topic from different angles. But the general questions are clearly the same: What is an institution? What is the social world made of? How many kinds of social entities are there? So either the answers are substantially different, or they must be somehow compatible. In the first case, we must try to figure out which answers are right and which ones are wrong. In the second case, we must try to understand how seemingly different answers fit together; we must figure out whether they focus on different aspects of social reality, or whether they are expressing the same ideas using different vocabularies and theoretical frameworks.

In this book I propose a theory to unify the main traditions in the field of social ontology and explore the implications of this unification. In the course of the book I focus mostly on *human* sociality. This may seem a contentious decision, because humans are not the only social animals. From bees to hyenas, swallows, and chimps, many nonhuman animals live in groups and have interesting forms of social organization. But there is something special in humans: our societies are more complex and much more diverse than the societies of any other species. Throughout history, humans have experimented with many types of social organization, and there are other arrangements that are potentially feasible but have never been tried until now.

Consider the myriads of ways in which human beings have organized their family life. Anthropologists classify family structures as monogamous and polygamous, poligenous and polyandrous, exogamous and endogamous, matrilineal and patrilineal, matrilocal, patrilocal, bilocal, and neolocal, consanguinal, affinal,

affiliative, and fictive (the list could be longer). And notice that the family is one of the social institutions that are most closely related to biological functions like mating and reproduction. Other institutions are even more autonomous from biological constraints, and have been shaped in thousands of different ways during the history of humanity.

So most of this book is devoted to understanding what human institutions are, how they work, why they differ, and what they can do for us. Since these questions can be tackled in different ways, it is important to pitch the inquiry at the right level. At the bottom of the scale, going from the particular to the general, we could study specific institutions like the British monarchy, the Catholic Church, or the Gandhi family. Following an established philosophical jargon, we shall call them *token* institutions. The main evidence that we are dealing with a token institution is that such entities have a history and a geographical location. They are situated in space and time.

Although historians and social scientists spend a lot of time studying token institutions, they theorize mostly about sets or classes of institutions. The point of theorizing is to generalize beyond specific cases, to explain by means of general models the functioning of more than one token institution. So the study of institutions will require that we shift one level up in the scale of generality. Terms like “monarchy,” “political party,” “bank,” “firm,” “union,” “church,” “cult,” “family,” “golf club” are used to refer to entire classes of institutions that share important properties. This is the level where one can hope to find models and categories of general theoretical interest.

There are many possible levels of analysis, which can be ordered hierarchically in terms of generality. “Church” refers to a kind of institution that is more general than “Protestant Church,” which is in turn more general than “Reformed Church,” and much more general than the “Dutch Reformed Church.” A scholar of institutions may theorize at each of these levels, depending on her goals and interests. And in some cases she may

even formulate general principles that hold for all institutions. This is the point where social science meets philosophy, in the field of social ontology. Social ontologists investigate the social world at the highest level of abstraction, and devise theories that are supposed to hold for all social institutions irrespective of their individual features.

Theorizing of course is easy, if unconstrained. The difficult task is to separate good from bad theories, those that are explanatory from those that are not. In order to do that, philosophers and scientists usually test their theories against particular cases. When a political scientist theorizes about types of democracy, for example, she tests her theories using token examples of historically existing democratic institutions. When doing social ontology we must proceed in a similar manner, except that the examples and counterexamples are usually picked from one level up: we test our theories against types of institutions.

There is obviously an enormous variety of types of institutions that can be used for this purpose. And the danger of picking examples to support one's favorite theory looms large. Fortunately, however, there is a set of cases that are considered paradigmatic and that must be accounted for by any theory that wants to be taken seriously. In the course of the book I will refer frequently to three types of institutions, either for illustrative purposes or to test specific theoretical hypotheses. My paradigmatic institutions are *marriage*, *private property*, and *money*. I will also occasionally refer to the rules of traffic, a simple institution that we are all familiar with, and that is analogous in many ways to more complex institutions like marriage, property, and money.

Each one of these institutions is a high-level type, which can be analyzed hierarchically into lower-level types and, low and behold, tokens. Marriage for example can be monogamous or polygamous, temporary or permanent, chosen or arranged, same-sex or different-sex, open or closed. And there are of course

historical institutions with their peculiarities: there are Catholic, Muslim, Judaic, and Hindu marriages. There is my marriage with my wife, Caesar and Cleopatra's marriage, or the marriage of the gay couple who live next door. The interesting questions are, what do all these types and tokens have in common? What is *the institution of marriage* at the most general level of description? Answering these questions now would be premature. But as an appetizer, we can begin to notice that institutions are usually grouped by scientists according to their *functions*. The token institutions that anthropologists classify in the "marriage" category, for example, usually regulate activities aimed at procreation, the rearing of children, the care of the elderly, inheritance, and economic cooperation between the spouses. Similarly, money is defined by economists as whatever entity or type of entity is used as a store of value, medium of exchange, and unit of accounting ("money is what money does," as the saying goes).

The advantage of functional definitions is that they abstract away from the innumerable ways in which a goal may be achieved in different contexts. For this reason functions are used for classificatory purposes not only by social scientists, but also by biologists when they theorize about physiological traits. An eye, for example, is an organ that perceives and represents the environment through the detection of light. Eyes come in different guises, eyes can be classified according to different types, and each token eye (the eye of a wasp as opposed to the eye of a mammal) may exploit different light-detection mechanisms. Nevertheless, there are general theoretical principles that hold for eyes across the species. And similarly, there are interesting generalizations that apply to different marriages, regardless of the specific ways in which each token institution works.

The notion of function is strictly related to the idea of purpose or goal. So what is the purpose of institutions? As a first approximation, it seems that institutions facilitate coordination and cooperation. They help groups of individuals to do things

that are better done together. Sometimes these collective activities are not particularly problematic, and coordination takes place easily. But at other times, the same goal can be achieved in different ways, each way implies a different division of labor, and it is not clear which is the best way to do it. In such cases, we shall say that there is a *problem of coordination*.

For example, it is easier to organize a dinner party if each host takes a specific role: you go shopping and I cook, I wash the dishes while you entertain the guests. For the party to be a success it is important that we all do our job. It would be a disaster if we both went shopping and no one cooked. But who is going to do what? Perhaps we both like cooking and we both hate shopping. At the same time, each one of us would be willing to do the shopping, if she knew that the other was doing the cooking. The problem is in part a problem of assurance, of being confident that the others are going to do their part in a complex collective task.

How can institutions build this assurance? Answering this question will take up the first part of the book. I will survey different views of institutions, analyze them critically, and explain how they relate to each other. I will begin by drawing a distinction between those theories that view institutions as rules, and those that view institutions as equilibria of strategic games. Then, I will argue that these two approaches are complementary, and that they can be unified within a single framework.

The *equilibria* approach spans across the divide between philosophy and social science. The seminal theory in this tradition was proposed by David Lewis in a justly celebrated book on *Convention* (1969), but over the past four decades several other philosophers and social scientists have proposed equilibrium-based accounts of social institutions. Theories within the equilibria approach view institutions as behavioral patterns that tend to persist because individuals have no incentive to deviate from the pattern unilaterally (unless everyone else does the same).

In spite of its explanatory achievements and its mathematical elegance, the equilibria approach has not been universally endorsed however. According to an equally popular alternative, institutions should rather be conceived as *rules* that guide the actions of individuals engaged in social interactions.

The rules account is close to our vernacular, prescientific understanding of institutions: intuitively, institutions regulate behavior, making certain actions appropriate or even mandatory in specified circumstances. The institution of private property, for example, regulates the use of resources by indicating who has access to them. The institution of money regulates the use of paper certificates in economic transactions. And the institution of marriage regulates the behavior of two or more individuals who pool their resources to raise kids, manage property, and help each other in many different ways.

But if institutions are rules, how do they influence behavior? Stating a rule is clearly insufficient to bring about an institution. To realize why, consider that there are plenty of *ineffective rules*: rules that are officially or formally in existence but that are nevertheless ignored by the majority of people. Traffic lights in Milan are regulation, in Rome they are a suggestion, and in Naples they are just decoration, as the saying goes. But since the rules are formally the same in Milan, Naples, and Rome, there must be something else going on. There must be some special ingredient that makes people follow the rules in some circumstances and ignore them in others.

The equilibria account of institutions tells us what the special ingredient is: effective institutions are backed up by *a system of incentives and expectations* that motivate people to follow the rules. An equilibrium in game theory is a profile of actions or strategies, one for each individual participating in a strategic interaction. Each action may be described by a simple sentence of the form “do X” or “do Y.” The defining characteristic of an equilibrium—which distinguishes it from other profiles—is that each strategy must be a best response to the actions of the

other players or, in other words, that no player can do better by changing her strategy unilaterally. If the others do their part in the equilibrium, no player has an incentive to deviate.

Since the actions of a strategic game can be formulated as rules, equilibrium-based and rules-based accounts of institutions are compatible. From the point of view of an external observer, an institution takes the form of a regularity that corresponds to the equilibrium of a coordination game. But each equilibrium strategy also takes the form of a rule that dictates each player what to do in the given circumstances. By combining the rules account with the equilibria account we obtain a unified theory that I call the *rules-in-equilibrium* approach to the study of institutions. Rules by themselves lack the power to influence behavior, but together with the right system of incentives and beliefs, they can influence the behavior of large groups of individuals. Institutions, in a nutshell, are rules that people are motivated to follow.

Institutional rules sometimes simply state that we must “do X” or “do Y.” In many cases, however, they are *conditional* statements that prescribe different actions depending on the occurrence of certain events (“if X then do Y”). For example, the rules of traffic state that you must stop at the crossroads if the traffic light is red, proceed if it is green. Similarly, in many societies the actions of individuals are regulated according to their identities—there are rules of courtesy like “ladies first,” as well as hierarchical rules like “give orders if you are the husband, follow them if you are the wife.” Biological traits in such cases are used as signals that facilitate coordination, pretty much as traffic lights help us drive around smoothly. (If you are perplexed by this statement, let me clarify that these arrangements are not necessarily good equilibria: perhaps we would be better off if women gave orders and men obeyed; similarly, we could stop when the light is green and proceed when it is red.)

Traffic lights and biological traits are *correlation devices*, and the actions of people who use these signals constitute *correlated*

equilibria. Correlation devices multiply the number of ways in which we can try to coordinate. Suppose, to use my trite example, that you and I want to organize a dinner party. To simplify, let us suppose that we do not have strong preferences regarding the division of labor. To make sure that we coordinate, I text you a message: “I shop and you cook.” The main purpose of this signal is to create the expectation that I will go shopping. Because if you believe that I will go shopping, then you will do the cooking, and the party will be a success. But of course this is just one of many possible signals that we could have used to coordinate. Had I told you “I cook and you shop,” the opposite equilibrium would have been implemented. So language is a tremendously versatile device to create institutions, by sending signals that people use to converge on new equilibria. Humans are special in the animal kingdom in large part because they have language, and because they can use it to create a wide range of different social arrangements.

This point has not passed unnoticed of course. The most original and systematic attempt to place language at center stage in social ontology is the theory of constitutive rules proposed by John Searle. Although this theory is a variant of the rule-based account of institutions, it attempts to explicate institutions using a very different kind of rule that, instead of merely regulating behavior, creates the possibility of new types of behavior. Constitutive rules according to Searle are statements of the form “X counts as Y in C,” where Y denotes an institutional entity or fact or property, X is a preinstitutional entity, and C is a set of circumstances or conditions of instantiation. In the case of money for example a constitutive rule is: “Bills issued by the Bureau of Engraving and Printing (X) count as money (Y) in the United States (C)” (Searle 1995: 28).

Searle contrasts constitutive rules to regulative rules that have as their syntax “do X,” or “if X do Y.” The actions or strategies that appear in game-theoretic accounts of institutions, as we have seen, have precisely this form, so Searle’s distinction

suggests that there is a deep hiatus between his own approach and the accounts of institutions found in the social science literature. But if this were true, then the attempt to unify different approaches to social ontology would fail: not all institutions would be systems of (regulative) rules in equilibrium.

There are good reasons, however, to believe that Searle's distinction between regulative and constitutive rules does not hold. Using an argument originally devised by Frank Hindriks, I will show that constitutive rules have a much more limited role than the one envisaged by Searle: they are term-introducing principles that state the conditions of application of the theoretical terms that we use to label institutions. They are, first and foremost, naming devices for regulative rules.

The constitutive rule of money, for example, specifies the conditions that have to be satisfied for something to be money (it must be a paper bill issued by the Bureau of Engraving and Printing), and implicitly specifies what to do with paper certificates of that kind (use them to trade commodities, save them for future purchases, etc.). Hindriks's view that regulative rules can be transformed into constitutive rules via the introduction of theoretical terms highlights the fact that constitutive rules do not add anything that cannot be expressed by means of simple regulative rules. In principle they could even be eliminated from our theoretical vocabulary, without causing any substantial ontological loss. The constitutive rule of money for example can be translated in a regulative rule such as: "if a bill has been issued by the Bureau of Engraving and Printing, then use it to purchase commodities or save it for the future," and so forth.

The unified theory thus helps attain ontological parsimony and at the same time offers an explanation of the pragmatic function of institutional terms (why they are useful and how they help us coordinate). Having accomplished that, the remaining part of the book will be devoted to articulate the theory in more detail, and to explore its philosophical implications. In

particular, I will focus on the implications of the unified theory for the explanatory and predictive ambitions of social science.

For well over a century social scientists have been discussing the methodological foundations of their discipline. On the one hand, methodological “monists” have been arguing that the social sciences must follow the same approach as the natural sciences. On the other hand, methodological “pluralists” have argued that the very nature of social reality makes it impossible for social scientists to attain the same explanatory and predictive success of the natural sciences. Social scientists should adopt a different approach and give up the traditional goals of naturalistic scientific inquiry.

What ontological differences may license this kind of skepticism? A classic cause of concern has been the *mind-dependence* of social reality. The idea is that social entities differ from natural entities in that the former, but not the latter, depend essentially on our representations. The nature of a dollar bill, the fact that it is money, for example, depends on a collective belief or recognition that it is money—that it can be used to buy certain commodities and services. (Otherwise, it would be just a piece of paper with a picture of George Washington printed on it.) In contrast, a molecule of water is water regardless of what anybody believes about it. It does not have to be represented as water, in order to be what it is.

The thesis of mind-dependence has been used by many theorists to challenge the scientific ambitions of social science. The challenge can take different forms, however, depending on how the concept of dependence is interpreted. So part of the book will be devoted to distinguish between different versions of the dependence thesis. In particular, it will be useful to distinguish between *causal* and *noncausal* dependence on representations.

I will argue that the thesis of causal dependence is true, but that its philosophical consequences have been exaggerated. Mind-dependence, in particular, when it is interpreted causally does not constitute a threat to the scientific ambitions of social

science. The thesis of noncausal dependence, in contrast, is just false. I will try to give a precise formulation of the thesis using the notion of ontological dependence, and show that it is inconsistent with the functionalist understanding of institutions that is common in the social sciences. This, as we shall see, implies that we can entertain a realist and fallibilist attitude toward the entities studied by social scientists, just as we do with the entities studied by natural scientists.

The final two chapters are devoted to an issue that is currently hotly debated in many countries, concerning the design and identity of one of our most important institutions. The issue is whether to reform the institution of marriage so as to make it possible for partners of the same sex to get married. As we shall see, traditionalists have claimed that the institution of marriage is intrinsically or necessarily limited to heterosexual couples, and that the inclusion of same-sex couples would turn it into a different institution. The claim has often been backed up by sophisticated semantic arguments, and philosophers have been engaged in the battle on both sides of the field.

My own view is that it is perfectly legitimate to use the term “marriage” to refer to the contracts that regulate the relationships between individuals of the same sex. However, the debate on marriage highlights an interesting problem: it suggests that it is difficult to be simultaneously a realist and a reformist about institutions. Some philosophers have argued that the identity of institutions depends not on the rules that people actually follow, but on those that they *should* follow—that is, on the normative targets that we set for ourselves as a community. This “ameliorative” approach (a term used by Sally Haslanger) however is incompatible with realism. So I will propose a different solution based on the unified theory, to save both the realist principle that institutions do not depend noncausally on our intentions, and the reformist intuition that the rules of the game can be redesigned without changing the identity of an institution.

This is, more or less, the content of this book. Although I do not expect anyone to be persuaded by a short summary, I hope the appetizer will make you want to read more. There are gaps in the argument, and I will try to fill them in the chapters that follow. But even if I fail to convince, I hope the book will help philosophers and scientists appreciate how different projects in social ontology are related to one another. I hope that it will foster communication across research programs, and collaboration between scholars who adopt different approaches. Social ontology has been disunified for too long, and it is time that we put it together again.

REFERENCES AND FURTHER READINGS

Social ontology is a lively field, and the literature as a consequence is quite large. The volumes edited by Schmitt (2003), Mantzavinos (2009), and Gallotti and Michael (2014) include essays by many influential philosophers and are a good point of entry in contemporary debates. The diversity of human institutions and their relative independence from biological constraints are central topics in contemporary social science and have generated heated debates. For a view that emphasizes biological constraints, influenced by evolutionary psychology, see Boyer and Petersen (2012). On functionalism in social science and biology, I tend to follow Pettit (1996). Since most of the topics mentioned in this introduction are analyzed in more detail in the rest of the book, I refer the reader to the literature reviews at the end of the subsequent chapters.

