

Chapter 1

Introduction

In *Leviathan*, Thomas Hobbes constructs a theoretical apparatus that promises to unify the fields of human psychology, ethics, and political philosophy. Hobbes's grand project is to found an "objective science of moral philosophy," firmly grounded on truths of human psychology, from which a moral obligation to obey an absolute political sovereign can be generated. Though such an ambitious and far-reaching theory usually attracts considerable and long-lasting attention, Hobbes's social contract theory seems to have been almost entirely dismissed by his contemporaries and largely ignored by political philosophers in the twentieth century until Rawls published *A Theory of Justice* in 1971. Since then, interest in Hobbesian contractarianism has grown steadily. Alongside the critical discussion of Rawls's theory, a substantial body of contemporary Hobbesian scholarship has developed in both political and moral theory. In the last decade, this research has culminated in the publication of three major works inspired by Hobbes's social contract theory. This book attempts to provide an overarching and detailed critical analysis of these works in an effort to explore both the insights and limits of political and moral Hobbesian contractarianism.

Before presenting this analysis, however, I provide the backdrop from which these theories have emerged. My objective in this chapter is to introduce contractarian theory in general, to provide one view of why Hobbesian contractarianism is enjoying a contemporary renaissance in moral and

Introduction

political theory, and to anticipate some of the themes which emerge from the analysis in the subsequent chapters.

A GENERAL CHARACTERIZATION OF CONTRACTARIANISM

Contractarian theories can be versions of either political or moral contractarianism. “Political contractarianism” constitutes a theory of political legitimacy and obligation. Political contractarianism sets forth conditions under which political institutions are morally legitimate and under which individuals may have political obligations.¹ “Moral contractarianism” constitutes a meta-ethical theory which sets forth a criterion of assessment for moral principles. As a meta-ethical theory, moral contractarianism proposes a theory of the nature of moral justification and the meaning of moral terms.

There are two basic categories into which political and moral contractarian theories seem to fall. First, contractarian theories can be *analytical*. Analytical contractarian theories provide a reconstruction of the genesis and continuity of political authority or the genesis of moral principles. Such theories do not make any claims concerning the historically accurate nature of the evolution of political institutions or any actual genesis of moral principles. Rather, they impose an artificial and alternative scenario on political evolution, or adopt a theoretical framework to reconceive of moral principles as the outcome of some hypothetical scenario. Contractarian reconstructions treat a set of political institutions or moral principles *as if* they were the outcome of some

¹ There is an interesting literature which separates the issue of political legitimacy from that of political obligation (e.g., A. John Simmons, *Moral Principles and Political Obligation* [Princeton: Princeton University Press, 1979] and Kavka [p. 399]). For simplicity, I will treat the issues of political legitimacy and obligation *as if* they were the same, and thus I will proceed *as if* political legitimacy entailed political obligation, and *vice versa*. Though this view is controversial, no substantive claims will depend essentially on its truth.

General characterization of contractarianism

contract, agreement, convention, or choice. The aim is to show that political society can be understood in terms of some theoretical, sometimes hypothetical, model, or that moral principles can be understood as resulting from a choice of some kind. Second, contractarian theories may be *normative*.² Normative theories in general conclude that something ought to be the case, is justified, or is legitimate or not. In contrast to analytical contractarian theories, which merely reconstruct or reconceive the origins of political association or moral principles, normative contractarian theories argue in favor of or against political association or moral principles. An analytical reconstruction of political association which holds that political authority can be understood, for example, as the outcome of a bargain is either plausible or not, depending on the nature of the reconstruction offered. But even if plausible, such an analysis by itself is neutral with respect to the legitimacy, desirability, or defensibility of political authority. It is possible to admit that political authority can be reconstructed or modeled as the outcome of some hypothetical scenario or other, but to deny the relevance of this reconstruction for purposes of establishing the legitimacy of political authority. Likewise, it is possible to agree with an analytical theory asserting that certain principles can be understood as the outcome of some hypothetical bargain, but to deny either that such principles qualify as moral principles or that even if moral, such principles are therefore in any sense correct, or ones with which we have reason to conform. Normative contractarian theories go beyond the merely analytical to draw certain normative conclusions about the legitimacy of political authority or the correctness of moral principles.

I have distinguished among political, moral, analytic, and normative contractarian theories. Any given contractarian theory may fall into all or any combination of these categories. But I have not yet indicated what makes a theory,

² See Alan Nelson, "Explanation and Justification in Political Philosophy," *Ethics* 97 (1986):154.

Introduction

whether political, moral, analytic, or normative, a contractarian theory.

THE THREE-STAGE CONTRACTARIAN ARGUMENT SCHEMA

Contractarianism can be characterized by singling out certain structural elements common to all versions of contractarianism, elements which together may be regarded as constitutive or definitive of the contractarian approach. These essential elements of contractarian theory take the form of a three-stage argument schema.

The first stage of the contractarian argument

The first stage of the contractarian argument is the specification of a hypothetical scenario. In this stage, the contractarian merely stipulates a set of hypothetical individuals and characterizes a hypothetical environment in which we are to imagine the hypothetical individuals interacting. There is no definitive set of characterizations of the hypothetical scenario which must be provided. Contractarian specifications of a hypothetical scenario are many and varied, and thus can range from the general and broad to the detailed and narrow. But there are certain parameters within each characterization task to which contractarians typically attend. These parameters will serve to illustrate the nature of the specification of the contractarian hypothetical scenario.

The characterization of a set of hypothetical individuals. A contractarian begins specification of the hypothetical scenario by defining the characteristics of the hypothetical individuals which would significantly affect, if not determine, the nature of their decisions, choices, and interactions in the hypothetical environment. Thus, contractarians characterize their hypothetical individuals in terms of their rationality, risk aversion, future utility discount rate, interests, information, and physical and intellectual capacities.

Three-stage contractarian argument schema

The nature and extent of an individual's rationality will have a profound effect on the choices he will make. Contractarians determine the nature of their hypothetical individuals' rationality by endorsing some definition of rationality. One of the most theoretically fruitful definitions of rationality is the narrow economic definition of rationality as utility maximization. According to this definition, a rational actor seeks to maximize his utility, or in the parlance of economic and game theory, is a maximizer of his utility function. It is standard practice in microeconomic theory to understand this utility function as generated from the actor's set of preferences, which are assumed to be complete, transitive, and asymmetrical.³ The advantage of attributing rationality so construed to the hypothetical actors in the contractarian choice problem is that their choices may be determinable as a matter of mathematical calculation. But contractarians need not claim their hypothetical actors are rational in this economic sense of rationality, and may either attribute a weaker version of the economic definition of rationality to the actors or employ a different notion of rationality entirely.

Whatever the conception of rationality contractarians endorse, they seem to agree as to the extent of rationality which must be attributed to their hypothetical individuals. These individuals must be, in general, at least minimally rational, however one construes rationality. For all contractarians hope to make generalizations concerning the nature of the decisions, choices, and interaction of the individuals

³ If we let x , y , and z be alternatives, and define the relation xP_iy to mean "Person i prefers x to y ," then completeness, transitivity, and asymmetry can be defined as follows:

Completeness: For any pair of alternatives x and y , either xP_iy or yP_ix

Transitivity: For any three alternatives x , y , and z , if xP_iy and yP_iz , then xP_iz .

Asymmetry: For any pair of alternatives x and y , if xP_iy then not yP_ix .

Introduction

in the hypothetical scenario.⁴ The minimal rationality assumption enhances the tractability of analysis of individual choice and collective action problems in the hypothetical scenario. For example, suppose we stipulate that on the economic model of rationality defined above, a minimally rational individual is one who at least seeks to maximize expected utility (even if she does so imperfectly). If we can assume that every individual in the hypothetical scenario is minimally rational in this sense, we can at least deduce the optimal strategies and outcomes for the actors in the scenario, if not the strategies likely to be adopted and the outcomes likely to result. The stronger the rationality assumption, of course, the higher the certainty of the analysis of interaction in the hypothetical scenario. But if the threshold of minimal rationality is not met, it would be extraordinarily difficult even to undertake analysis of the problems in the hypothetical scenario, much less to determine their solutions. If individuals cannot be assumed even to be minimally rational (i.e., and thus cannot be assumed at least to be seeking to maximize their expected utility), then there appears to be no way to generalize about their behavior. Even a complete and accurate description of their beliefs and desires would provide an inadequate basis for predicting their behavior. If, for example, a contractarian's strategy is to construct a hypothetical scenario in which individuals are to make some choice, there would be no straightforward way reliably to predict which choice such individuals would make (or for that matter, whether the individuals would make a choice at all).

In addition, the degree of rationality attributed to the

4 Speaking of the justification of a narrow rationality assumption in economic and game theory, Russell Hardin notes that "the assumption of narrow rationality is almost methodological: it is easy to accommodate in analysis, and it is relatively easy to assess in generalizable behaviors. . . ." Hardin also observes that "often the assumption of narrow rational motivation yields predictions that are the most useful benchmark by which to assess the extent and the impact of other motivations" (Hardin, p. 11).

Three-stage contractarian argument schema

individuals in the hypothetical scenario may also play a substantive role in the ultimate normative argument contractarians make. The contractarian argument might, for example, rely on the claim that the outcome of interaction in the hypothetical scenario is ideally rational because it would result from ideally rational interaction in the hypothetical scenario. Obviously, such a claim can be vindicated only if the contractarian characterizes the hypothetical actors in the hypothetical scenario as ideally rational. The role of the rationality assumption in the overall normative contractarian argument will be discussed in more detail when I consider the construction of the contractarian hypothetical scenario, and again in the course of previewing the specific contractarian theories which will concern us.⁵

The level of rationality attributed to individuals by itself, however, is far from determinative of their decisions, choices, or behavior. Another important factor is the extent to which the hypothetical individuals are characterized as risk averse, risk neutral, or risk inclined, and the extent to which such dispositions toward risk differ among these individuals. For equally rational individuals will make dramatically different choices in the same situation depending on their risk aversion. For the contractarian's analysis of the hypothetical scenario, the specification of risk aversion will affect, for example, what formal rule of choice the hypothetical individuals would employ. Two equally rational individuals, as defined by the economic conception of rationality sketched previously, and in the same situation, may nonetheless employ different choice rules and thus make different decisions. For example, a risk-averse individual might employ the maximin rule in order to maximize his minimum possible gain (or minimize his maximum possible loss), while forfeiting the possibility of higher gains, while a risk-inclined individual might employ the maximax rule in order to maximize his possible gain, at the expense of risking a lesser gain (or greater loss) than he could otherwise guarantee himself.

⁵ See the section "Construction and Fit" in Chapter 1.

Introduction

An equally important factor in determining individual decisions and behavior is the rate at which individuals discount future utility. The “discount rate” for individuals is the economic term for what is colloquially referred to as “farsightedness” or “shortsightedness.” Equally rational individuals may differ with respect to their current or present valuation of future gains. One rational individual may value, for example, one dollar today four times as much as one tomorrow, while another equally rational individual may value one dollar today only twice as much as one tomorrow. As a result, these individuals will calculate their utility gains very differently and make very different choices. Whether and to what extent individuals are shortsighted or farsighted will play a crucial role in the theories we will later consider.

The nature and scope of the interests of the hypothetical individuals also significantly affect their decision process and behavior. Equally rational individuals may have radically different interests which range over radically different areas. Contractarians sometimes characterize the nature of their hypothetical individuals’ interests in a highly theoretical manner. They sometimes, for example, endorse the economic conception of rationality we have seen, but further define “true” or “ideal” rationality as the maximization of a utility function generated from a set of purified or idealized interests.⁶ Contractarians may have different theories regarding the purification or idealization of these interests. They must, then, specify whether their hypothetical individuals have these idealized or nonidealized sets of interests. When contractarians specify the scope of the interests of their individuals, they must explain, for example, the extent to which the individuals have interests only in themselves, or what we might call “self-regarding” interests, or interests

6 Cf. David Brink’s description of a purified desire-satisfaction theory: “A counterfactual desire-satisfaction theory claims that what is valuable is what would satisfy one’s desires in some preferred epistemic position, for instance, if one’s beliefs and desires were consistent and one’s desires were based on full (nonevaluative) information” (Brink, p. 221).

Three-stage contractarian argument schema

in the well-being of others, or what we might call “other-regarding” interests.⁷ In economic terms, the issue is whether individuals have interdependent utility functions. Here, the contractarian specifies whether the hypothetical individuals are, for example, egoistic, altruistic, or some combination of both.

Contractarians typically also characterize the individuals in the hypothetical scenario in terms of the nature and scope of their information and uncertainty. The information typically specified in the hypothetical scenario concerns self-knowledge, social knowledge, and general knowledge. Self-knowledge includes individuals’ knowledge of their personal characteristics, such as height, weight, sex, physical and mental abilities and disabilities, and so on. Social knowledge includes knowledge of their cultural values and norms, social and political institutions, economic strata and institutions. General knowledge includes knowledge of economics, sociology, criminology, political science, medicine, and all other standard areas of academic and common knowledge.

Finally, contractarians will characterize the physical and mental capacities of the hypothetical actors. Traditional contractarians typically follow Hobbes in characterizing the hypothetical individuals as roughly equivalent in physical and intellectual capacities, or at least sufficiently so to make it very unlikely that any one individual would, on her own, be capable of systematically dominating the others. And with the exception of those who attribute ideal or perfect rationality to the hypothetical actors, most contractarians attribute roughly the same intellectual capacities to their hypothetical actors as ordinary people have.

The characterization of the hypothetical environment. Once a set of hypothetical individuals has been defined, the contractarian then constructs a hypothetical environment in

⁷ The distinction between “self-regarding” and “other-regarding” interests is drawn nicely in Joel Feinberg’s *Harm to Others: The Moral Limits of the Criminal Law* (New York: Oxford University Press, 1984), pp. 70–9.

Introduction

which these individuals are to interact. The characterization of the hypothetical environment includes such features as the availability of resources, the availability of certain technologies, and the size of the population of individuals.⁸ Contractarians often characterize their environment as one in which the existence of various technologies does not affect the approximate physical equality of individuals in it, and the population is large enough to create moderate scarcity of resources and significant coordinative and informational problems which confront large groups of people. Further, the physical characterization also determines whether the environment is supposed to be a prepolitical setting designed to model a hypothetical anarchy preexisting any political association, a postpolitical setting designed to model a hypothetical anarchy ensuing upon the end of a preexisting political association, or more of a “thought experiment,” which is neither.

Summary

The first stage of the contractarian argument, characteristic of contractarian theories in general, requires the specification of a hypothetical scenario. This requires the characterizations of a set of hypothetical individuals and a hypothetical environment in which they are to interact. The contractarian typically attributes to the hypothetical individuals a certain level of rationality, risk aversion, farsightedness, self- and other-regarding interests, information, and physical and mental capacities. The hypothetical environment is characterized according to the resources and technologies available, and the size of the population of individuals in it. Once contractarians provide these characterizations, they begin the second stage of the contractarian argument.

⁸ See Hampton, pp. 254–5 for the significance of the existence of varying degrees of technologies in the contractarian choice problem.