Introduction
Why Rawlsian Political Analysis?

Ontology precedes methodology.

New Microfoundations for Social Science

Good and bad and right and wrong are our most fundamental moral concepts. We are instructed from early childhood to “be good,” we want what is good for ourselves and for our families and friends, and we like to be good at our jobs and at other tasks that are important for us. For Americans, the right to the pursuit of happiness, associated with achieving the good, is enshrined in our Declaration of Independence. We all know, too, that we should try to do what is right. Learning right from wrong is an important part of growing up, and to be right with one another is to have our relationships in good order. In our political discourse, no value exceeds doing what is right.

The distinction between the good and the right, and the consequences of this distinction for social analysis, is the topic of this book. Since the 1950s, Anglo-American political scientists, economists, and other social scientists have been trying to found political and social analysis on the analytics of the good, neglecting or relegating to a subsidiary order the analytics of the sense of right. I argue, however, following Immanuel Kant and John Rawls,
that the sense of right is just as fundamental as is the sense of the good,
2. that these two senses are somewhat independent from one another and operate according to different dynamics, and
3. that these two senses and only these two are consistently fundamental to our practical decision making.

These senses and no others form the basis for practical reason. If these claims are true then we need a new political analysis, and this is the analysis this book aims to launch.

Some social scientists have argued that decision making can be modeled in terms of rational utility maximization, which boils down to each person (rationally) maximizing the satisfaction of his or her interests (including the interest we take in the good of others). To satisfy our interests, in this view, is to achieve what we take to be our good. This model lies at the foundation of rational choice theory and of neoclassical economics, and it stands as the only established microfoundations for political or social analysis. When contemporary political scientists or economists speak of microfoundations, this is the model to which they refer.

Among neoclassical economists this model is completely normative and uncontroversial. Almost every economic analysis relies on the assumption of rationality, implying that the decisions and choices of the people under consideration can be modeled as aiming to maximize the satisfaction of their interests. Among political scientists, however, (and among sociologists and historians as far as their disciplines use this model) it is a very different story. Here rational utility maximization as a model of choice is highly controversial, and although most political scientists probably use some findings from rational choice theory, only a minority could be considered adherents to this theory and most consider it inadequate as a basis for the work they do.

I argue that rational choice theorists and economists have simply got their model of choice wrong. To the extent that the model works, it is because the sense of the good (represented by our interests) is one of the two foundations for practical decisions. It is serviceable enough to assume rational utility maximization when analyzing some kinds of voting patterns and power politics, rates of inflation, and changes in prices.
This assumption is not much use, however, for analyzing the social and political dynamics underlying most social problems. In fact, our practical decisions are influenced as much by our principles, the cognitive objects of the sense of right, as by our interests, the main cognitive objects of the sense of the good. For this reason, social analysis based on both the right and the good, on principles as well as interests, will provide accounts of social relations that are more coherent, more satisfying, and more true. Rawlsian political analysis, as I call it, provides the firmer basis for understanding our social world, and, understanding it, for changing it.

I appreciate that these are rather audacious claims. This introductory chapter will lay out the forms of evidence I offer to support them and summarize the main points of my arguments. The model of choice that I propose, however, borrowed from Kant and Rawls, means more for political science than just showing how better empirical arguments can be developed. It also has consequences for how we understand the role of political science, what we are doing when we do political science, and for the ethical frameworks from which we should approach social analysis. That it should have these kinds of consequences will come as no surprise to anyone familiar with the work of Kant and Rawls.

I should state at the outset that adopting their model of choice does not imply accepting their ethics or epistemologies wholesale, much less accepting all of Kantian metaphysics (some of which Rawls rejects). However, I think it does require accepting the main lines of the structures of their ethical theories as plausible. We have to take many of their conclusions seriously even if we do not completely buy them.

In particular, I will argue that adopting his model of choice requires that we accept Kant’s view that our experience of the world is based as much on our own cognitive capacities and on the concepts and categories we bring to perception as on the nature of external reality. When we do political analysis, one thing we are doing is constructing (reinforcing, revising, deepening) our conceptions of right, conceptions that form a basis for our engagement with subsequent practical problems.

I will also argue that accepting this model of choice leads to accepting Rawls’s “original position,” the viewpoint he establishes for selecting principles of justice, as appropriate for approaching problems of social justice generally. In the original position, we consider questions of social
justice as if from behind a “veil of ignorance,” that is, as if we did not know our level of wealth or our position in the distribution of natural talents, political power, and so on. Excluding this kind of knowledge, we arrive at a perspective from which we can consider social problems fairly, free from biases that might arise from our individual interests. As vulnerable social agents with compelling interests and regulative senses of right, when we are going to address questions of social justice, this is the appropriate place to start.

Kant is widely regarded as the preeminent philosopher of the Enlightenment, and Rawls, who largely adopts and extends Kant’s ethics, is the preeminent political philosopher of the last half century. Kant, of course, wrote long before Darwin, but Rawls does not engage with evolutionary theory or the associated science either. Writing on models of choice today, however, in the first decade of the new millennium, as the science advances and evolutionary perspectives have begun to seep into common sense, this connection needs to be made. From an evolutionary perspective, interests can be explained straightforwardly in terms of desires for food, sex, and territory, but how can an equal place in choice for principles or sentiments of right be reconciled with evolution? I build on Marc Hauser’s account of sources for the sense of right particularly in primate evolution to argue that the sense of right is grounded in our cognition of expectations, that is, of the regularity of the natural world. However, this and other components of the sense of right could not produce moral emotions such as guilt and resentment until language came to be well developed. Both interests and principles are basic cognitive objects that we can alter through reflection and that contribute to constructing our world, but although interests are grounded in natural desire, principles are associated with conceptions of the shape and structure of social worlds infused with agency.

When I say that principles and interests are cognitive objects, I mean literally that they are embodied in structures of neurons in our brains. To say they are fundamental to decision making is to say that these structures are engaged when we make decisions, contributing to their content. To say that they operate according to different dynamics is to state the hypothesis that the structures of neurons that embody principles and interests operate according to distinct sets of patterns, probably located in different regions of the brain (and probably with some overlap). Although there
are differences in dynamics among principles and among interests, the differences between the sets are greater than the differences within the sets.

One might have interests in getting a job or a promotion, in reading the *New York Times*, in keeping one’s weight down, or in going to a party. When economists assume that people are interested, say, in maximizing their incomes, formally they may only imply that people’s behavior can be expected (on average, to some extent, in specified contexts) to be consistent with this assumption. This formal assumption, however, must eventually be made good in particular choices and these choices in patterns of neural activity. Also, it is hard to keep the formal use of terms such as “rational self-interest” rigorously distinct from common, everyday self-interest. We sometimes say that someone has an objective interest (in studying to do well in a class, for example) that may differ from their subjective sense of their interests (when they want to go out with friends instead) or that they may not properly understand (say when a young person rides a motorcycle without a helmet). The tension between so-called objective and perceived interests is, of course, one of the basic justifications for scholarship and education.

My use of the term “principle” is broader, and more formal, than the way the term is often used in common speech. Many people might recognize responding aggressively to a threat or giving credit where credit is due as principles, but they might not think of stopping at red lights or greeting neighbors cheerfully when meeting them on the street as principles. But the cognitive processes underlying all these actions fit the pattern of a principle. Principles underlie our senses of right, justice, fairness, appropriateness, legitimacy, and propriety, so when something strikes us as, say, appropriate, it is because an event has conformed to our principle. When we feel indignant, resentful, or vengeful it is due to the violation of a principle, and principles are invoked by sentiments of loyalty, honor, and obligation. The fixed ideas someone might refer to as “my principles” are a subset of that person’s principles as I understand them.

The Kantian Model of Choice

I follow Rawls’s example in appropriating central features of Kant’s model of choice. Whereas Rawls builds a theory of justice on this model,
I build a new form of social analysis. Of course Rawls’s application of the model changes it, and mine does too, but in both cases the essential features of the model remain. Kant takes practical reason to consist of two parts, one empirical and the other pure. Empirical practical reason always involves an object in the world, and it is on this basis that it is defined. For example, if I hurry to a performance in order to find a good place to sit, I am employing empirical practical reason. Since interests normally involve objects (the object in which the interest rests), insofar as rationality supports the pursuit of interests it falls within empirical practical reason. In Kant's lexicon, “pure” refers to an operation of the mind that does not involve impressions from the senses. Our reason is practical when it is employed in guiding our actions. It is practical and pure when we choose an action because it accords with a principle. For example, if after my partner cooks the dinner I choose to wash the dishes because that is fair, I am employing pure practical reason.

In Kant’s model the imperatives of empirical practical reason are hypothetical and those of pure practical reason are categorical. A hypothetical imperative of empirical practical reason (an interest) is one we follow depending on the configuration of conditions in the world, and as a means to something else. Kant defines a categorical imperative as “one which represents an action as objectively necessary for itself, without any reference to another end.” Here he has in mind an imperative of morality (a principle), which has to do “not with the matter of the action and what is to result from it, but with the form and the principle from which it results.”

Kant’s pure practical reason becomes “the reasonable” for Rawls, while empirical practical reason becomes “the rational.” Rather than defining “the reasonable,” Rawls specifies two of its aspects as virtues of persons. First,

persons are reasonable . . . when, among equals say, they are ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so.

Second, reasonable persons are willing
to recognize the burdens of judgment and to accept their consequences for the use of public reason in directing the legitimate exercise of political power in a constitutional regime.\textsuperscript{8}

Here we see that the cognitive capacity that underlies the reasonable is the sense of fairness.\textsuperscript{9} The reasonable involves the ability to generate an idea of what is fair in particular circumstances (necessarily a generalized idea) and to use such an idea to guide one’s behavior. The second aspect of the reasonable, the willingness to recognize the burdens of judgment, involves the demands that are placed on the sense of fairness in the political life of citizens in a constitutional regime. While this aspect gives an idea of the scope of application of the sense of fairness, and it helps to express the role of the reasonable in Rawls’s idea of political liberalism, it is not essential to the cognitive capacity per se.

The capacity that we call “the sense of fairness” is also expressed in our senses of right, justice, legitimacy, appropriateness, and propriety. We can think of these as a family of senses expressing the same cognitive capacity, or as terms that describe applications of one underlying sense in different contexts. They all operate on the basis of a logic of principles, and their principles work by ruling some actions or states in and others out. If an action strikes us as, say, unjust or illegitimate, we can explain this sense by naming the principle that we take to have been violated. We may also sometimes take an action simply because it is fair or right, and to do so is not to act from self-interest. In such a case it would be incorrect to say that we are acting in order to promote any particular interest, whether ours or another’s, although the action may indeed serve particular interests. One might say that a just action promotes the interest of justice, but this is metaphorically to personify justice.

To include the reasonable within the microfoundations of political analysis is to take motivations associated with the sense of fairness—those discussed above and related sentiments such as guilt, indignation, resentment, and vengefulness—to be no less central to political decisions than motivations associated with interests and notions of the good. This requires an analysis that can account for manifestations of these senses in our political and social world, one in which explanations for political phenomena may be worked out in terms of these senses and

© 2012 University of Notre Dame
their associated motivations. It bears noting that while Kant and Rawls hope to help us to clarify our reflections on principles, perhaps so that we might lead more ethical lives or so that a more just society might be attained, a practical political analysis must come to grips with the actual principles upon which political action is based, however unsavory these may sometimes turn out to be.

In order to model the relationship between principles and interests in cognition, however, it is useful to start from what Rawls calls “ideal theory.” In *Groundwork for a Metaphysics of Morals* Kant presents the categorical imperative, which he thought to be the supreme principle of morality, in several formulations. A familiar one is, “act only in accordance with that maxim through which you can at the same time will that it become a universal law.”\(^{10}\) In his lectures on Kant, Rawls argues that the principles in *Groundwork* are those “from which a fully ideal reasonable and rational agent would act, even against all object-dependent desires, should this be necessary to respect the requirements of the moral law. Such an ideal (human) agent, although affected by natural inclinations and needs, as we must be belonging to the natural world, never follows them when doing so would violate the principles of a pure will.”\(^{11}\) Here Rawls sets up a framework for the relationship between categorical and hypothetical imperatives, or principles and interests. He views one function of principles as essentially negative in the following way: they exclude the pursuit of some interests. We have no moral reason not to act on those interests that the categorical imperative does not exclude.

Rational choice theory conceives of the person as possessing a set of interests, which in a given practical situation yields a schedule of preferences.\(^{12}\) Given the person’s beliefs about a practical situation, this schedule in turn yields an ordering of options and, possibly, the selection of a unique utility-maximizing choice. The sequence can be schematically represented as:

![Figure 1.1 Rational Choice Model](image)

Interests → Preferences → Choice

In the Rawlsian view of the person, however, both the reasonable and the rational, principles and interests, are sources of volition. I follow Rawls in
identifying rational choice theory’s interests as hypothetical imperatives. Let us call them “Interests_h.” It is also possible for the person to adopt the promotion of a principle as an interest (e.g., when one takes an interest in promoting justice). Let us call such interests “Interests_c.” However, the reasonable, as cognitively represented in our principles, also excludes from consideration some interests associated with hypothetical imperatives that we would otherwise affirm. Therefore the significance of our principles for our interests can be represented as:

Figure 1.2  The Reasonable Guides and Restricts the Rational

Principles → Interests_c, Interests_h

Even for Kant’s ideal agent, however, there are some hypothetical imperatives that the categorical imperative does not exclude. Also, the categorical imperative sometimes leads directly to choice, without establishing an interest. Therefore the Kantian model of choice can be represented as:

Figure 1.3  Kantian Model of Choice

Interests_h

Principles → Interests_c, Interests_h

Interests_h (remaining), Interests_c → Preferences → Choice

Principles → Choice

The person starts with interests and principles. Principles lead to the establishment of certain interests and to the exclusion of others. The remaining interests based on hypothetical imperatives as well as those (perhaps few) interests based on categorical imperatives serve as the basis for a schedule of preferences, which (as in rational choice theory), in a particular situation yields a choice. As in rational choice theory, choice is also informed by the person’s beliefs, but this can be omitted here because there is no difference between the two models in this respect. Finally, the Kantian model includes a pathway to choice that is foreign to rational choice theory, which is when a person makes a choice because it is consistent with a principle. Clearly, it is this feature of the model that distinguishes it
absolutely from the rational choice model. The deliberation in the choice made in accordance with a principle is not one of maximization.¹⁶

In cognition, the separation of principles from interests is not always clear and distinct. I think we should take Kant as indicating that principles and interests represent distinct forms of thought or deliberation that are central to practical reason. In some cases a decision may be influenced by both principles and interests. Indeed, Kant argues that reasoning is not necessarily transparent to the reasoner, and an action that one takes to be motivated by a principle may be influenced by an interest in ways that one may not acknowledge or even comprehend.¹⁷ The philosophical challenge has been to demonstrate that principles can be independent of interests. In the well-known scenario he uses to support this point, Kant argues that a man must admit that when threatened with execution if he fails to give false testimony against someone whom his prince would like to destroy under a plausible pretext, it would be possible for him to overcome his love of life.¹⁸ Granting the fundamental and distinct roles of principles and interests in cognition, for the purposes of social analysis we can then relax the constraints on principles imposed by ideal theory; a husband may hold the principle that his wife should both cook and wash dishes because that is a wife’s role, and a parent may hold that a child should obey the parent unquestioningly. If self-interest enters more strongly into the formulation of the principle than perhaps it should, that does not much affect how the principle functions in cognition.

Principles, Norms, Values, and Identity

A defining feature of the model I propose is that principles form the basis for social norms. A norm is nothing more than a widely shared principle. The norms that shape our culture are implicit in our language, and we learn them in childhood by hearing them asserted and by observing how they order social relations (i.e., how they are interpreted and applied). Once we have learned them, they shape our expectations for the behavior of others, and we interpret others’ behavior in terms of the norms/principles we have already adopted. If we think of particular principles as important to our identity, they are specific principles that
we have selected or on which our actions have been based in important life events and that we have chosen to affirm.

In ordinary speech it is not unusual for the terms “values” and “principles” to be used as synonyms. Someone may place a high value on telling the truth, and also affirm telling the truth as a principle. The concept of “values,” however, as it is derived from the concept of value, occludes the distinction between principles and interests. If we hold to a principle, it translates into one of our values, and we also value what we see to be in our interest. We generally have deeper and more considered attachments to values than to interests, and to interests than to mere preferences. (A preference is the residual, so to speak, after applying our relevant principles and interests to a given case, or one of a set of options in the event of a tie.) Compared to a principle, however, a value has a more individualistic connotation. A principle has a warrant to recognition beyond that it belongs to someone. While both values and principles express identity, our links to society are more likely to be expressed by principles.

Despite their connecting us with others, we think of our principles as very much our own. When we identify a norm, however, we are acknowledging that it is held by “most people,” or by a particular group. There are two ways in which we are likely to encounter a norm, from the outside and from the inside. As a visitor to a foreign country, we are likely to remark on how local norms differ from the ones with which we are familiar. “To think that the drivers stop here when they see a pedestrian in the cross walk!” or “If you don’t ‘tip’ the customs officer here you are likely to spend several hours at the airport!” When we encounter norms from the outside, it is like learning the rules of a new game (particularly if we plan to participate for some time) and/or it is part of coming to understand a complex external phenomenon (a society, an organization).

When we encounter a norm from the inside, it is likely that we are becoming aware of its present force or significance for us or for someone else. It may be a minor realization leading to routine compliance, such as when I recall the way I should address the president of my university when I meet him in the library. Or it may involve a twinge of guilt and an effort to compensate, such as when I realize how long it has been since I extended hospitality in my circle of friends. Often our compliance with social norms is barely conscious; we know and follow many complex
rules that we might not be able to articulate. Although it may strike us that they are externally imposed (e.g., by expressions of favor or disfavor), and we may find them burdensome, it is clear that cognitively we impose them on ourselves. Unless we are told what to do, it is our own conception of the norm with which we comply.

It would be a mistake to think of identity narrowly or reductionistically, given that it seems to involve our entire person as it is expressed in our various relationships, the projects we undertake, and our experience up to the present. Nevertheless, our principles make up a substantial part of our identity, perhaps the most important part. Kant has a particular idea of freedom that emphasizes not being a subject to material causality. He takes it that it is in adherence to the principles of a moral law of which we are ourselves the author that we find our freedom. One can admit the centrality of principles to identity, however, without accepting the whole cloth of Kantian ethics. Note that we have to learn many of our society’s principles (norms) before we are in a secure position for identifying our own, and to a considerable extent we choose and define our principles over and against the context constituted by those of our family, communities, and society. Although we habitually comply with many principles, it is the ones we choose ourselves, those that become central to our decision making, that we often value most highly. This is not to say that we normally adopt principles *qua* principles with heroic clarity of choice. On the contrary, we often adopt them in the course of the give-and-take of a relationship, or we make a choice and then gradually adopt the principles that follow from it. The way in which principles become layered with one another and with experience over time contributes importantly to our identity.

**Interests and Norms Are Widely Accepted Bases for Social Analysis**

In order to demonstrate the plausibility of Rawlsian political analysis I pursue several strategies. First, in this introductory chapter, I show that the distinction between interests and norms is widely accepted as fundamental in the social sciences, with each forming the basis for different
schools of social analysis. If one accepts that norms are shared principles, then principles and interests really are the uniquely correct microfoundations for social analysis, which suggests that Rawlsian analysis could unify hitherto disparate strands. Following this discussion I outline the principal arguments in the remaining chapters.

The noted organizational theorist James G. March argues that there are two approaches to analyzing decisions, one based on a logic of consequences (i.e., interests) and the other on a logic of appropriateness (i.e., norms). For March these are simply two different approaches, the former associated with rational choice theory and the latter with certain traditions in sociology. Each approach is found in a systematic and established tradition of social analysis, so it seems plausible that each of their respective logics captures something that is fundamental to practical reason. Yet surely our reason is the same, regardless of the perspective from which it is analyzed. If each logic is indeed fundamental, should not social analysis be responsive to both?

As distinguished a rational choice scholar as Jon Elster finds himself similarly compelled to take recourse to two social theories. In discussing problems of bargaining and of collective action, he employs two main conceptual tools: rational choice theory and the theory of social norms. He notes that he has “come to believe that social norms provide an important kind of motivation for action that is irreducible to rationality or indeed to any other form of optimizing mechanism.” These conceptual tools, in turn, correspond to the two sides of “[one] of the most persisting cleavages in the social sciences . . . the opposition between . . . lines of thought conveniently associated with Adam Smith and Emile Durkheim, between homo economicus and homos sociologicus. Of these, the former is supposed to be guided by instrumental rationality, while the behavior of the latter is dictated by social norms.” Elster therefore adopts an eclectic view, taking it that some behaviors “are best explained by the assumption that people act rationally, whereas others can be explained by something like the theory of social norms,” or else that many actions are influenced by both rationality and social norms. It is certainly better to take an eclectic approach than to ignore the reasonable, but from a theoretical perspective this is unsatisfactory. We should aim for a unified theoretical apparatus based on the best available conception of social reason.
In her influential book *Governing the Commons*, Elinor Ostrom develops a model of rational choice that incorporates norms. Rational action, she argues, involves an “internal world of individual choice” consisting of four variables: expected benefits, expected costs, discount rates, and internal norms. In a given situation, the first three variables, which interpret the present value of the net benefits from a choice, correspond to interests in Kant’s model, and the fourth variable, internal norms, corresponds to principles. Ostrom argues that communities that develop norms involving high levels of trust and reciprocity possess social capital, and these communities are more likely to succeed in building institutions that resolve dilemmas associated with common pool resources (such as groundwater or fisheries). She reverts to a maximization model, however, to account for how people change their rules.

Recall that one aspect of the reasonable involves the readiness to propose principles and standards as fair terms of cooperation and to abide by them willingly. Given Rawls's idea of the person, one can no longer imagine choice—even, or particularly, choices about rules for managing common resources—merely in terms of a process of individual maximization. Rather we see that such decisions are routinely and often sincerely based at least in part on notions of fairness. This leads to a view of social institutions that is quite different from those found in rational choice theories; the analyst must now attend to the specific nature of the principles of fairness that are employed.

This distinction between rational choice and Rawlsian views of institutions is a fundamental one, with implications not only for how social analysis should be carried out but also for how citizens in a constitutional regime should (or are likely to) regard one another. Thus, as Jack Knight develops theoretical foundations for rational choice institutionalism, he is bound to view individuals as locked in conflict. Like Elster (whom he cites), Knight sees a basic cleavage in the social sciences, although Knight places Adam Smith along with David Hume and Herbert Spencer on the side of theorists who emphasize the collective benefits produced by social institutions, while he counts Karl Marx and Max Weber as theorists who emphasize conflict in the distribution of these benefits. Like Elster, Knight takes it that it is “reasonable to assume that both norms and rational calculation motivate action in different contexts.” Knight argues,
however, that “most social outcomes are the product of conflict among actors with competing interests,” and that “rational choice theory is better able to capture the strategic aspects of that social conflict.”

Knight takes it that social institutions are formed out of a “standard bargaining problem,” a situation where “there are benefits to be gained from social actors working together, sharing resources, or coordinating their activities in some way. These actors need rules to structure their independent activities. More than one set of rules can satisfy this requirement, and the rules differ in their distributional properties. Because of this, people have conflicting preferences regarding the institutional alternatives.” Since there is nothing in Knight’s model to ameliorate these conflicts and each resolution yields ongoing inequalities, the conflict will persist. Yet this is the very same problem that Rawls takes to define the circumstances of justice:

There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to try to live solely by his own efforts. There is a conflict of interests since men are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share. Thus principles are needed for choosing among the various social arrangements which determine this division of advantages and for underwriting an agreement on the proper distributive shares.

While Rawls acknowledges the conflict that Knight identifies, he thinks people are capable of adopting principles that are fair as the basis for a constitutional regime. Hence, we need not perpetually regard one another as potential adversaries. We can hope to establish a society in which all benefit from one another’s natural gifts and productive projects.

Although Knight initially acknowledges that norms influence choice independently from interests, he later attempts to derive norms from interests, saying he aims to build “microfoundations for the informal network of rules, conventions and norms that capture some of the principal ideas of macro-level accounts in the Weberian and Marxian tradition.” He argues that informal institutions arise in a decentralized manner as “a
by-product of strategic conflict over substantive social outcomes,” the results of which hinge on “the fundamental relationship between resource asymmetries, on the one hand, and credibility, risk aversion, and time preference, on the other.” Principles play such a major role in social relations, however, that any social theory will be substantially defined by how it accounts for them. Since rational choice theory is based on interests, it is perhaps inevitable that it eventually finds interests to provide the basis for principles.

Although a Rawlsian view does not discount the force of interests, it implies that even principles that affect large-scale social outcomes can arise from our free reason. Moreover, principles and interests have distinct roles in cognition that require different forms of analysis. Principles are significantly independent from interests, and their sources and the way they influence choice are different. They have different links with the emotions, and certain sentiments (e.g., indignation, resentment) depend directly upon principles. If principles and interests follow different logics, and they are each fundamentally involved in choice, and if neither principles nor interests can be reduced to the other, then both should be included in the microfoundations of social and political analysis.

Grounding and Demonstrating Rawlsian Political Analysis

As this brief discussion has shown that it is common for theorists to base different forms of social analysis on principles and on interests, chapter 2 grounds the concepts of principles and interests in the work of Kant and Rawls and discusses their roles in their respective philosophies. This is to give a richer understanding of the meaning of principles and interests and of their roles in cognition. Also, given Kant’s and Rawls’s positions in Western social thought, to see the architectonic place of principles and interests in their corpuses aids an appreciation of the importance of these concepts. Little attention in the scholarship on Rawls has been devoted to the way the reasonable and the rational are represented in the original position, so it is helpful, too, to bring this representation out.

Chapter 2 proceeds to discuss several accounts of the origins and development of principles. One set of accounts, as noted above, takes an
evolutionary perspective. Interests are so clearly central to cognition in the context of evolution that the task here is to explain how principles could possibly be coequal to interests in the microfoundations of social analysis. Scientists have recently uncovered considerable evidence that some primates and other mammals possess many component parts of a sense of fairness. Expectations, a sense of status, the recognition that other animals’ actions are affected by what they can see (implying some kind of appreciation that other animals can think), and other cognitive capacities of primates provided materials for what became the sense of right. In this chapter I argue, however, that the development of language deepened humans’ appreciation of consequences and their understanding that other people think and feel like them. Language also made it possible for humans to make commitments. All of these capacities together provided the platform from which the sense of right emerged.

Granted evolutionary sources for the sense of right, the questions remain how the sense of right develops in each person and how we should understand the historical development of norms that embody complex political principles. These are deep and complicated questions that can only be addressed here in a limited way. In regard to the individual person’s development of a sense of right, chapter 2 discusses three stages of morality, from childhood through adulthood, that Rawls presents in *A Theory of Justice*. The roles and significance of principles change as a person passes through what Rawls describes as the morality of authority and the morality of association to, in full maturity, the morality of principles. It is notable that this discussion lies at the intersection of philosophy and psychology, as Rawls links his account not only with those of Jean-Jacques Rousseau, Kant, and John Stuart Mill, but also with that of child psychologist Jean Piaget.

For another view on the development of principles in children and for an account of the historical development of norms, chapter 2 turns to Axel Honneth’s notion of reciprocal recognition. A student of Jürgen Habermas and Georg Wilhelm Friedrich Hegel, Honneth’s ideas are not easily encapsulated. However, he presents a threefold schema based on love, rights, and solidarity that is both psychologically and historically grounded. Although each schema is based on a different form of recognition, I argue that they can be reframed in terms of principles. Linking our discussion of the evolutionary sources of principles to Honneth’s account
of the transition from traditional (medieval) to postconventional morality helps to round out this account of the development of principles.

After chapter 2 fleshes out the ideas of principles and interests, chapter 3 addresses the legacy of rational choice theory. Given the widespread applications of rational choice theory, if rational choice theorists have indeed got their model of choice wrong they must have found ways to accommodate principles outside the core of their analyses. This chapter presents several examples of rational choice theorists encountering evidence of the sense of right and shows how they have accommodated it in rationality-centered accounts. Recall that microfoundations provide models of choice that are said to underlie decisions and social events that we observe in the world. There are two possible patterns: either (1) rational choice theorists will exclude patterns of reasonable choice (including resentful and vengeful choice as well as choices based on ideas of fairness or appropriateness) from their core analyses, or (2) they will incorporate them but reframe them in terms of rationality. In the first case, they will either (a) account for observed choices wrongly or (b) include reasonable choices in their discussions but fail to explain them. In the second case (such as Ostrom’s model of choice described above), as they force the reasonable into models based on the rational, their models will lack a certain elegance and at some point they must inevitably impose a consequentialist framework onto deontological material (actions based on principles). In either case, they will fail to reach fully adequate explanations for social patterns based on principles.

Chapter 3 begins with the historical record of prisoner’s dilemma experiments, where the behavior of human subjects apparently motivated by a sense of fairness diverged consistently from predictions informed by rational choice theory. This is an example of pattern 1a: because their models excluded reasonable choice, the experimenters explained certain behaviors that they observed wrongly. Next, the chapter considers Robert Bates’s account of the Mau Mau Rebellion in Kenya and Margaret Levi’s account of compliance and noncompliance with military conscription in Canada. The former is an example of 1b: Bates identifies behavior arising from resentment, which he actually explains quite well, but he fails to ground his explanation in his theory. The latter is a second example (after Ostrom’s model of choice) of pattern 2: Levi builds a model including
ideas based on fairness but forces it to conform to a utilitarian framework by way of the so-called dual utility hypothesis.

Besides addressing rational choice theory’s limits as a positive theory, chapter 3 also discusses what we might describe as the theory’s constructive limitations. From a Rawlsian perspective, social and political analysis cannot be merely positive, since to analyze the social world is also to take part in constructing it. Rational choice theory’s focus on interests leads it to overemphasize conflict, and in light of the constructive role of social analysis this is deeply problematic. Moreover, although Knight argues that rational choice theory is better at explaining the strategic aspects of social conflict, this theory also fails to recognize many of the grounds for cooperation (as well as long-standing resentments that sometimes undermine it). But as vulnerable social beings in changing environments, we need an analysis that can help us to thrive collectively as well as individually. Rational choice theory’s limitations in this regard constitute a serious analytic shortcoming.

Chapters 4 through 6 present different kinds of applications of Rawlsian political analysis—a program analysis, a political history, and an analysis of a social problem, in each case comparing them to neoutilitarian alternatives. Changing microfoundations leads to two kinds of changes in social analysis. First, although economics and rational choice theory (in their internally consistent applications) explain social events and patterns of social change based only on interests, Rawlsian analysis explains events and patterns based also on principles. At the micro level, Rawlsian analysis can register dynamics involving the promotion of new principles—and resentments and power plays based on old ones—to which neoutilitarian analyses tend to be blind. At the macro level, the Rawlsian approach conceives of patterns of social relations and social change that are based on principles, and combinations of principles and interests, as well as patterns based mainly on interests. Since principles are internal to the person and interests have external objects (usually subject to change from various forces), analyses of social patterns based on principles tend to require longer time frames. These patterns are also prone to violent eruptions. Norms and principles often define and are expressed by orientations and frames of reference, and when these conflict they seem to be harder to negotiate than conflicting interests.
The second structural change in social analysis due to the change in microfoundations is found in its evaluative standards and frameworks. Neoutilitarian analyses, not surprisingly, involve utilitarian evaluations of one kind or another, that is, evaluations based on some idea of maximizing the satisfaction of interests (sometimes including interests in the well-being of others). In addition to the satisfaction of interests, Rawlsian analyses, by contrast, take account of changes in principles, and are oriented to autonomy and social justice. Thus, Rawlsian political analysis rests directly on Rawls's original position.

This change in evaluative standards works out differently in program analyses from the way it does in political history or problem-based analyses. Program analyses orient strategies of governments or of other agencies with public purposes. Whatever the evaluative standard, these agencies should aim to maximize their impacts subject to their resource constraints (and without violating ethical prohibitions). Rawlsian analysis adds a second spectrum of impacts to those a neoutilitarian program analysis would recognize. While neoutilitarians measure and count up gains such as in income, health, and security, Rawlsian analysts also take account of changes in principles in favor of autonomy and social justice, such as when members of a previously excluded group are empowered to participate in decisions about the allocation of public resources, or when groups build what neoutilitarians refer to as social capital. Changes in principles and in the satisfaction of interests are brought together (possibly summed) and compared to costs, and the result anchors the assessment of the program and its component parts.

A political history or a problem-based analysis, by contrast, orient an individual person or a group, contributing to that individual’s or that group’s members’ conceptions of the social world. It responds to the general assumption of agency and potential engagement, although of course a particular individual who reads and affirms an analysis may not act specifically in response to it. Kantian and Rawlsian ethics, however, do involve an imperfect (wide) obligation to promote social justice (see chapter 2). Political histories and problem-based analyses can be understood as supporting the orientations required to fulfill this obligation.

The program analysis carried out in chapter 4 focuses on the Grameen Bank of Bangladesh, mainly in the period from the late 1980s through the 1990s. The Grameen Bank, established in the early 1980s,
provides loans and other financial services to over three million poor people, predominantly rural women. Consistently reporting repayment rates around 98 percent, in the 1990s it gained worldwide acclaim as one of the few successful large-scale microfinance programs serving the poor. In this period, international financial institutions such as the World Bank were promoting so-called market friendly approaches to development in low-income countries. Microfinance programs were often established as a partial substitute for welfare programs that made handouts to the poor, and the Grameen Bank was often proposed and taken as a model.

The Grameen Bank is an appropriate illustration of Rawlsian program analysis for several reasons. First, it is very well studied. Neoclassical economics provided the dominant theoretical orientation within the international development community of the 1990s, and hence the standard framework for analyzing the Grameen Bank, but there are also many rigorous studies informed by other theoretical frameworks. Second, the two central elements of the Grameen Bank’s strategy—microfinance and a focus on poor women—offer good materials for substantiating neoutilitarian and Rawlsian approaches both in their causal analyses and in their impact assessments. Third, because the Grameen Bank was widely taken as a model, mistakes and omissions in the neoutilitarian analyses are particularly important. This is a good case for demonstrating the significance of adopting Rawlsian program analysis.

Economists attributed the Grameen Bank’s success largely to the structure of incentives resulting from the bank’s famous group-lending policy with its so-called social collateral. The bank would lend only to individuals in groups of five, with new loans to any member of the group contingent on all members repaying their loans. Under this policy each borrower has an incentive to help the members of her group to repay their loans and to discourage them from defaulting. While social collateral certainly contributed to the Grameen Bank’s success, it is unlikely to have had the central importance often attributed to it. This is partly because the policy was seldom enforced, and indeed it has now been abandoned. Additionally, Rawlsian analysis reveals measures the Grameen Bank took to encourage borrowers and also bank personnel to adopt principles from the bank’s ideology. These principles bound the bank’s borrowers and personnel in a joint project: poverty reduction and empowerment.
Impact studies of the Grameen Bank informed by neoutilitarian assumptions conceive of the bank’s impacts as consisting of a sum of utilities. These include changes in borrowers’ incomes, health, nutritional standards, and other goods that enhance the welfare of the borrowers and their families. In neoclassical economic analyses, the borrower’s motivation to participate in the program is thought to be a function of her (and her husband’s) expected utility gains, such as in these areas.

The logic of accounting for impacts in terms of autonomy and justice is complicated, because Kant’s idea of autonomy centers on acting in conformance to (freely generated principles of) the moral law. A heteronomous action, by contrast, is one that seeks to achieve an end or objective. Although material goods often are not immediately necessary for autonomy, the lack of the requirements for a secure subsistence undermines the development of autonomy. It is partly for this reason that Rawls’s theory of justice places so much emphasis on securing for all citizens what he calls “primary goods.” Having adequate income, education, and health care, as well as rights and liberties, creates conditions that support the development of autonomy. By this somewhat circuitous route Rawlsian analysis accounts for the importance of the program impacts that a neoutilitarian analysis recognizes, and the Rawlsian approach provides the stronger account of the importance of reducing poverty.

Rawlsian analysis also counts certain changes in principles as impacts. While some microfinance programs reinforce a paternalistic status quo, the Grameen Bank helps women who have traditionally been excluded from many household and societal decisions to participate in these decision-making processes. The development community’s discourse on the Grameen Bank has given a prominent place to the bank’s empowerment of women, so economists who study the bank have also recognized impacts in this area. The notion of empowerment, however, involves women changing their orientations to their families and to their society. These changes of orientation consist of changes in principles, but the neoutilitarian model does not acknowledge such changes. For this reason, economic accounts of the Grameen Bank’s impacts have had a hard time dealing with empowerment. In Rawlsian analysis, however, participation in collective will formation is an important part of autonomy. Empowerment has a natural place in Rawlsian analysis, and chapter 4 reveals and explains the Grameen Bank’s impacts in this area.
After chapter 4 addresses the nature, causes, and consequences of a famously successful institution, chapter 5 examines an instance of deep government failure. This chapter offers a Rawlsian analysis of the causes of poverty in Bihar, one of India’s poorest, and, with a population over one hundred million, second most populous state (prior to Bihar’s recent division). Political history is the approach that Rawlsian analysis necessarily takes to explain the conditions of a whole society. Rawls would consider Bihar to be a burdened society, one with historical, social, and economic circumstances that impede the establishment of a public conception of justice that supports the good of all citizens. In light of Bihar’s increasing poverty and worsening social conditions, the Rawlsian question is Why has it not been reasonably and rationally governed? To answer this question we look to Bihar’s moral traditions and to the structure of the state’s political and social institutions.

From this perspective we find that competent governance in Bihar has been undermined by the unfolding politics of caste at the state and village levels in a formally democratic context. Institutions based on an assumption of legal equality were superimposed on a society segmented by caste’s hierarchical principles. Governance has been undermined by the aggregation of political loyalties along caste lines and the resulting difficulty of aggregation along functional lines of material interest. For example, relations among the political elite in the twenty years following India’s independence in 1947 often took the form of conflict among the upper castes, while they neglected to promote their common interests as landlords and farmers. The application of caste principles sometimes displaced a focus on productivity, such as when some upper-caste landlords’ disdain for manual labor left them more likely to squeeze their tenants than to invest in improvements on the land.

Particularly divisive conflicts have arisen from caste resentments. Neoutilitarian views of the person provide no purchase on resentment, because it is not a rational emotion. Resentment arises, rather, from a sense of injury due to the uncorrected or uncompensated violation of a principle. On the face of it, given Bihar’s strong caste prohibitions and shockingly inequalitarian social relations, democratic practices were likely to engender resentment among dalits (former untouchables) and the lower castes as they were awakened to their democratic rights. It seems to have been the rage engendered by attempts of recently enriched
middle-caste farmers to dominate those below them, however, and by subsequent dalit and lower-caste resistance, that has prompted the most violence in Bihar’s countryside. For decades Bihar has suffered marauding caste armies, and the Indian public has become inured to reported massacres of men, women, and children in dalit hamlets and villages. Also, while many factors have contributed to fragmentation and corruption in Bihar’s state government, the most precipitous decline in the government’s competence has followed from a charismatic politician’s mobilization of caste resentments.

This chapter illustrates Rawlsian analytics at both macro and micro levels. We have noted that at the macro level, because a Rawlsian analysis must account for the interacting dynamics of principles as well as interests it tends to adopt a more panoramic perspective than neoclassical economic or rational choice approaches would require. Norms tend to reproduce themselves, and conflicts in Bihar, for example, between hierarchical principles of caste and egalitarian principles of democracy, have evolved over the better part of a century. The causes of the violence in Bihar, similarly, extend deep into the state’s colonial history. Examples of analysis at the micro level include investigations of caste principles, civil service appointments based on loyalty rather than merit, and cognitive-level factors influencing Bihar’s widespread social violence.

Program analysis lends itself to sharp distinctions between Rawlsian and neoutilitarian approaches, but explanations for a society’s social and economic conditions constitute a more diffuse field. To explain conditions in Bihar this Rawlsian analysis takes a narrative approach, establishing the context and developing the various themes that carry out and link the micro- and macro-level parts of the analysis. The analysis overall can be seen as framing Bihar’s conditions in terms of social justice, but the chapter does not include a point-by-point comparison with neoutilitarian approaches. This narrative could be compared to the analytic narratives of rational choice theory. The chapter does briefly compare its recommendations for Bihar with those from neoclassical economists.

Whereas the Grameen Bank is very well studied, Bihar’s conditions have not received adequate scholarly attention. I became familiar with this case when I taught a class on the political economy of Tanzania and Bihar. The socioeconomic problems in Bihar seemed to lend themselves
to Rawlsian analysis; economic analyses of Bihar’s conditions have been particularly superficial, and I could not find a good overall synthesis. Although the specific problems would of course be different, this chapter’s analytic approach could be applied to countries, and the contrast in analytic method will be apparent to readers familiar with economic and rational choice country studies.

While chapter 4 takes as its point of departure a famously successful program and chapter 5 sets out to explain the extreme poverty of a particular group of people, chapter 6 aims to come to grips with a looming social problem: global climate change. Rawlsian political analysis generally provides an orientation to its subject in terms of social justice on the assumption that it can orient and partially constitute the relevant portion of a person’s conception of the social world, and it is always at least implicitly forward looking. In the case of a problem like climate change that involves all living persons, its role is particularly foundational. Rawlsian analysis aims to offer a framework from which a person can approach his or her activities that involve the production of greenhouse gasses and that can orient political reform and institutional construction to respond to climate change. It provides a first answer to the questions What is the problem of climate change? and What should be done about it?

Since Rawlsian analysis shares the neoutilitarian focus on interests, one analytic source for approaching climate change is the theories of externalities and common pool resources already well developed by economists and rational choice theorists. Global warming involves costs of productive activities that producers may not have to bear and that are imposed on others besides those who consume the goods produced. It also involves the atmosphere surrounding the earth (and other goods of nature such as the oceans) as a global commons. Like herdsmen who share a pasture, we all share the atmosphere, so the greenhouse gasses that each person contributes can collectively undermine its value for everyone. While economists approach the resulting regulatory problem by taking existing institutions for granted and applying values based on the concept of utility, the Rawlsian approach starts from the original position and applies values based largely on primary goods. Given the nature of the causes of climate change and the roles of these causes in our productive and political systems, Rawlsian analysis assumes that we need significant
institutional reform to reach a just solution. It takes the original position to be the appropriate starting point for considering these reforms, a vantage point from which threats to the livelihoods of certain groups in the present and the future take on particular importance.

Chapter 6 develops Rawlsian analysis of climate change in contrast to analyses by economists Nicholas Stern and William Nordhaus. They both aim to balance marginal costs and benefits, but Nordhaus, following growth economics, arrives at a much higher target for restraining global warming, a temperature rise of about 3.5°C compared to preindustrial times. He discounts harms to future generations quite heavily, but Stern counts future harms the same as current ones, so he arrives at a target range of 2°C to 2.8°C. Both Nordhaus and Stern count up harms in economic terms based on the price system, but Rawlsian analysis aims to defend the livelihoods of those most threatened by climate change (while minimizing the compromise to economic growth). Rawlsian analysis also articulates the challenge of responding to climate change as one of changing principles and institutions. Nordhaus and Stern discuss strategies for reaching targets, but Rawlsian analysis identifies institutional barriers and what it would take to overcome them. Hence, agents in the original position accept a 2°C target for a maximum rise in temperatures, preferring a lower target but finding it institutionally implausible.

The Rawlsian approach analyzes the temperature target in terms of justice between generations. It analyzes the division of rights and responsibilities for reducing greenhouse gas emissions and for addressing harms caused by climate change in terms of justice between the rich and the poor (taking account of different contributions to causing climate change). In this context the international system of nation-states presents significant barriers to a just resolution. Nordhaus and Stern identify carbon taxes as efficient means to reduce greenhouse gas emissions, and Rawlsian analysis concurs. Given the association between greenhouse gas emissions and a country’s long-term economic prospects, however, the Rawlsian approach also aims for equal average per capita emissions on an international basis by 2050. Climate change turns out to be (among other things) a mechanism by which wealthy individuals impose harms on poor individuals, and averting and ameliorating these harms is a matter of justice. Chapter 6 explores the institutional barriers to a just solution.
and indicates directions for reforms, in particular, (1) political constraints on the United States achieving adequate reductions in its greenhouse gas emissions and adequate contributions to developing countries’ adaptation costs, and (2) institutional constraints on the effective use of adaptation and transition funds in developing countries.

Political constraints can be understood partly through political history. Chapter 6 discusses, for example, how in the United States, Republican governments and the Senate have usually blocked significant responses to climate change. The substantial and costly changes needed for an adequate response will need widespread public support, however—significantly more than has been demonstrated at this writing. Chapter 6 reflects briefly on some of the factors involved in generating and restraining such public support. The greatest challenge in the United States is probably to reduce greenhouse gas emissions. Even if the United States and other wealthy countries did provide adequate funds for poor countries to adapt to harms from climate change and to support transitions to low-carbon development pathways, however, at present these funds could not be used cost effectively. Adaptation and transition present technical and political challenges largely due to the difficulty vulnerable populations have in securing their interests in the public forum. Given that such difficulties cannot be overcome in the short run, I propose institutional innovations to defend the interests of vulnerable populations. This involves new ways to hold implementing agencies accountable for employing adaptation resources cost-effectively. Poor country governments may be no better positioned to use transition funds effectively than adaptation funds, but agents in the original position would nevertheless place primary responsibility for transitions to low carbon economies on poor-country governments while responsibility for supporting adaptation to climate change would be more widely shared.

Chapter 7 concludes the book by discussing some of the implications of the Kantian model of practical reason for the social sciences. This model views the practice of social science as constructing the social world, so although it must conform to empirical standards, it is inherently engaged in the ethical domain. Political economy and mainstream comparative politics already analyze macrodynamics of principles as well as interests, but they are likely to be enriched by building on microfoundations that
heretofore have only been implicit. Also, their leading practitioners have
often understood their task as merely empirical and articulated their ques-
tions in a critical dialogue with neoclassical economics (and sometimes
with rational choice theory). The touchstone for social analysis as an ethi-
cal practice, however, is the concept of social justice; social analysis can
only fulfill its constructive potential when it is grounded in this idea.
Chapter 7 assesses the consequences of this shift in relation to important
works by Robert Wade and Atul Kohli.
This chapter aims to secure the place of principles as coequal with interests in the microfoundations of political analysis, and to consider how this recasts the role of political analysis. In order to do these things, it first shows how the proposed political analysis is grounded in the work of Kant and of Rawls and explains how that analysis builds on Rawls’s project. Although I noted above that Kant and Rawls are engaged in ideal theory, given the Kantian moral psychology as a foundation, a certain applied analysis becomes both necessary and possible. In order to explore our cognition of principles and the relation of the reasonable to the rational and to other forms of cognition, this chapter considers the role of principles in Kant’s and Rawls’s philosophies. While Rawls asserts that “justice as fairness” (his conception of social justice) does not rely on Kantian metaphysics,¹ the Kantian moral psychology is at the core of Rawls’s constructive procedure. Showing how the reasonable and the rational are represented in the original position and their role in Rawls’s idea of political liberalism at once clarifies the concepts of principles and interests and also helps us to see what is required from what I am calling “Rawlsian analysis.”

In order to move from ideal theory to political analysis we need to grasp a more prosaic conception of principles than one might normally associate with Kant or Rawls. For this purpose, and also to establish
principles as coequal with interests, we consider three accounts of the sources and development of principles; the first is from Rawls, the second is from Honneth, and the third is my own account of possible origins of principles in human evolution. Kant and Rawls work with ideological frameworks that include theistic worldviews, although Rawls does not adopt such a view himself and Kant generally keeps the working area of his philosophy free from deistic considerations. Contemporary common sense, however, is beginning to take evolution more or less for granted. While an evolutionary perspective might appear to suggest that interests are more fundamental than principles, I show how the independence of principles from interests can be consistent with such a perspective.

The sparest justification for a Rawlsian political analysis can be drawn from basic features of Kant’s ontology and Rawls’s ethics. Given that (1) social relations are constructed from principles and interests and (2) we have an obligation to promote social justice, the need for such an analysis becomes apparent. We can link this spare account with Rawls’s project as follows. Rawls views the importance of justice as given. He builds on the contractarian tradition, drawing from John Locke, Rousseau, and Kant, to provide an alternative conception of social justice to those of the prevailing utilitarian tradition. His theory is loosely grounded in history, in that it appeals to what he refers to as “our” considered judgments and political traditions, and he explains a general sequence of implementation (i.e., by way of a constitutional convention), but he does not address how his theory might be adopted, for example, by any particular society.

I would like to emphasize that achieving a society that is just in Rawlsian terms is contingent and perhaps unlikely. Social justice is, after all, the ultimate public good (to borrow from economic theory). As the obligation of all, it is the responsibility of none, and steps to building a just society offend powerful interests. Rawls offers his theory of justice as cohering with and helping to order our considered judgments, but whether it does so depends on the principles that constitute a particular person’s political identity. The analysis I propose takes the question of what principles actually constitute our political identities to be an empirical one.
Kant and the Sense of Right

Kant regarded his own philosophy as having solved, once and for all, many long-standing problems in ontology, epistemology, and ethics. A central feature of his solutions in all three areas is the distinction between things in themselves and things as objects of experience, and the concepts from Kant on which Rawls bases the reasonable and the rational employ this distinction in Kant’s ethics. The distinction initially arises in Kant’s first great book, the *Critique of Pure Reason*, in which he addresses what he calls “theoretical reason,” the reason that supports the kind of knowledge found in the natural sciences. According to Kant we can gain knowledge of things in the world only through features revealed by our senses, and it is on this basis that things become objects of experience. There is no way for us to know what things in themselves might be like separate from our sensory experience. Our knowledge of things in the world, however, is based not only on sensory experience but also on how our minds interpret evidence from the senses. For instance, Kant takes it that we bring senses of space and time and the concept of causation to experience. Spatial and temporal relations and notions of cause and effect are forms of order imposed on experience by our minds.

When it comes to objects in the world, the idea of a “thing in itself” expresses an absolutely impenetrable barrier. There is just no way for us to know what a thing in itself might be like. In Kant’s view, though, when it comes to our own persons, we can and do have knowledge of ourselves, of our own personalities, distinct from the persons we are as belonging to the world of sense. As it turns out in Kant’s philosophy, although he does not express it in these terms, our personalities as “things in ourselves” are constituted by principles, while our natures as persons belonging to the world of sense are constituted largely by interests. Theoretical reason involves our knowledge of external objects; practical reason involves how we decide what to do. As Rawls puts it, “practical reason is concerned with the production of objects according to a conception of those objects—for example, the conception of a just constitutional regime taken as the aim of political endeavor—while theoretical reason is concerned with the knowledge of given objects.” And just as our knowledge of things in the world is conditioned by our notions of causation, space, and time,
so our experience of practical situations is conditioned by our principles (but also, in the case of practical reason, by our interests).

As noted in chapter 1, practical reason, according to Kant, consists of two parts—one pure, one empirical—and it is in the course of working out this ontological distinction that he develops the main concepts that make up his ethics. Pure reason, whether practical or theoretical, involves concepts or constructs we bring to experience from our own minds; empirical reason involves the activity of and evidence from the senses. As persons belonging to the world of sense, we find ourselves with natural inclinations and desires, and it is empirical practical reason that is moved by these. Pure practical reason, however, has the capacity to establish principles that can guide action without any admixture of an empirical motive, and this describes the central dynamic in Kant’s ethics. In Kant’s view we should always act in a manner that fulfills or at least that is not inconsistent with our duty, and we know our duty, in the first instance, by the principles that constitute our pure practical reason.

Two central and closely related concepts in Kant’s ethics—the categorical imperative and autonomy—are based on the distinction between pure and empirical practical reason. He takes it that every person naturally has a sense of duty, but that philosophy can sharpen or guide this sense and defend it from being undermined by various forms of skepticism. The form of the imperatives that constitute the sense of duty, or morality generally, is categorical, such that, a “categorical imperative would be that one which represented an action as objectively necessary for itself, without any reference to another end.” And also, “It has to do not with the matter of the action and what is to result from it, but with the form and the principle from which it results; and what is essentially good about it consists in the disposition, whatever the results may be.” Kant offers a universal formula for morality, the categorical imperative, which in one statement reads, “Act only in accordance with that maxim through which you can at the same time will that it become a universal law.” Thus when we are considering an action, we are to ask if we could affirm everyone in similar circumstances acting according to the same maxim or principle. Note that in establishing principles based on the categorical imperative, we of course take account of consequences, and part of comprehending (and framing) a situation in which we might apply such principles is to appreciate the consequences of different courses of action. To apply the
The categorical imperative is to match a principle to a case. Once we see that the situation calls for a particular maxim, say, of fairness, we simply apply it, at this point without regard to consequences.

The categorical imperative falls within pure (practical) reason because both the establishment and application of its principles abstract from sensory impulses. By applying principles, we impose an order on the social world that we ourselves have chosen, separate from imperatives arising from our nature as embodied beings. Our natural desires are included in the information involved in forming principles, but since we are choosing as though for a universal law we account for the natural desires of others on an equal footing with our own.

While the moral quality of principles that apply the categorical imperative arises from their suitability for universal application, their cognitive character arises largely from their independence from our nature as persons belonging to the world of sense. These two features—moral quality and cognitive character—together establish the suitability of principles as the basis for autonomy, which then becomes the central ideal for the person in Kantian ethics.

If the will seeks that which should determine it anywhere else than in the suitability of its maxims for its own universal legislation, hence if it, insofar as it advances beyond itself, seeks the law in the constitution of any of its objects, then heteronomy always comes out of this. Then the will does not give itself the law but the object through its relation to the will gives the law to it. Through this relation, whether it rests now on inclination or on representations of reason, only hypothetical imperatives are possible: “I ought to do something because I will something else.” By contrast, the moral, hence categorical imperative says: “I ought to act thus-and-so even if I did not will anything else.”

According to Kant our dignity as persons is due to our capacity to choose and to act from principles consistent with the categorical imperative, and in doing so we establish autonomy from our nature as sensible beings. He takes it that humans, like all embodied or sensible beings, necessarily desire happiness, such as from the satisfaction of interests, yet the principle of one’s own happiness is “most reprehensible.”
The absolutely good will, whose principle must be a categorical imperative, will therefore, undetermined in regard to all objects, contain merely the form of volition in general, and indeed as autonomy, i.e., the suitability of the maxim of every good will to make itself into a universal law is itself the sole law that the will of every rational being imposes on itself, without grounding it on any incentive or interest in it.\footnote{14}

This contrasts markedly with utilitarianism, which would orient morality to maximizing the good, largely in the satisfaction of interests (and which Kant critiques in its early form represented by Hume). Utility maximization as understood in economics and rational choice theory is often heteronomous.

Kant uses the terms "principles" and "interests" somewhat differently from my use of them here. He does not make the point that our sense of right is constituted by principles that represent duty and morality for us, but I argue this is evidently the case. Rawls (whom I follow in this matter) is right to associate the rational, which is constituted by interests, with hypothetical imperatives and empirical practical reason.\footnote{15} Also, Kant’s discussions of political principles generally take some form of monarchy more or less for granted, and distributive justice does not appear in his enumeration of duties. He asserts an obligation of beneficence but not to promote social justice.\footnote{16} Nevertheless, Kant is the classical source for the idea expressed in shorthand as “the right has priority over the good,” and so heavily does Rawls draw on Kant (although they are separated by almost two centuries) that it is hardly an overstatement to call Rawls’s work an extension of Kant’s.

The Reasonable and the Rational in Rawls’s Theories of Justice and Political Liberalism

While Kant’s ethics are largely oriented to providing a secure account of and grounding for morality, Rawls aims to provide an account of social justice. He argues that the social contract tradition, and Kant in particular, provides a better basis for this task than either the utilitarian
tradition, which dominated Anglo-American political philosophy before his *Theory of Justice*, or the other established alternatives. While utilitarians aim to maximize happiness or “the good,” perhaps as the satisfaction of rational desire, contractarians take it that we should adopt principles of justice that would be selected in some initial situation of equality. Rawls develops the original position to characterize such a situation, revising the categorical imperative as it applies to the problem of social justice. The original position is to carry the contract doctrine to a higher level of generality and to separate it from features of earlier approaches that were widely taken to be fatal weaknesses.

Principles of justice provide a basis for describing institutions that would constitute a just society, and “laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust.” Once one takes the problem of injustice seriously it is clear that the obligation to beneficence translates into an obligation to promote social justice, and Rawls takes this obligation for granted, although how it applies to each individual depends on her own circumstances and opportunities. To identify what should count as a just law or institution in a modern society, however, in the context of multiple and competing claims, can be quite difficult. Yet it is straightforward that just laws and institutions must be based on just principles. Rawls constructs the original position to serve as a framework for identifying such principles, and he does this by applying the reasonable and the rational to the problem of social justice.

Using Rawls’s well-known concepts as shorthand, we can say that the reasonable is applied through the veil of ignorance and the rational through the idea of primary goods. In the original position, imaginary agents representing us as citizens are to choose the principles of justice that will then serve as the basis for society’s constitution, laws, and institutions. These agents’ deliberations take place, however, behind a veil of ignorance, in that although they know that their society is reasonably well off, they do not know which particular individual they represent—and not the person’s wealth or income, gender, ethnicity, or natural talents, nor the person’s particular aspirations or plan of life. The idea is that when you and I select principles of justice, it would lead to bias if we took factors such as these into account. Through the veil of ignorance the
original position excludes all information that would permit any kind of heteronomy in the selection of principles of justice, so the principles selected will be reasonable.

On what basis, then, are agents in the original position to make their decision? They are to select the principles of justice that secure the best combination of primary goods for the person they represent, whoever this person turns out to be. They assume that this person has some plan of life and they select the principles that lead to the best goods to support it, including such goods as income, education, and health care; freedoms of belief, speech, and association; and opportunities to fill society’s various positions of responsibility. In securing such primary goods, the original position represents the rational, since justice requires society to provide the wherewithal for each person to secure her interests.

A central argument in *A Theory of Justice* is that agents in the original position would select the following two principles of justice (Rawls’s) rather than the principle to maximize the total or average utility of the people they represent.

1. Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.
2. Social and economic inequalities are to be arranged so that they are both:
   a. to the greatest benefit of the least advantaged, consistent with the just savings principle, and
   b. attached to offices and positions open to all under conditions of fair equality of opportunity.

These principles provide a secure basis for each person to develop and pursue a life plan, while under utilitarian principles there is always the danger that a lesser utility for some may be justified by a greater utility for others.

Thus, the original position provides a handy means of representing the reasonable and the rational as they apply to the problem of selecting principles of justice. Moreover, when we act on principles selected in this way, we are acting autonomously. We could clearly will that such
principles should become universal laws (if freely chosen by citizens of reasonably well-off societies), and by acting on them we express and consolidate our nature as free and equal rational beings alongside other citizens.

This brief summary of Rawls’s arguments is drawn from A Theory of Justice. I emphasize Rawls’s continuity with Kant and the centrality of the reasonable and the rational for both philosophers. In the years following the publication of Theory, however, Rawls found that he needed to distance himself from Kant in a particular way, and this is a central purpose for his second great book, Political Liberalism. It became apparent to Rawls that Theory expresses, or is grounded in, what he comes to call a “Kantian comprehensive doctrine,” and that this undermined the chances that his two principles would be adopted as a basis for political reform. The problem is that one normally finds in contemporary democracies several competing comprehensive doctrines, and indeed this is a natural consequence of sustained freedom of thought in a modern democratic society. Alongside the comprehensive liberalisms of Kant and Mill, one is likely to find Protestant, Catholic, Muslim, Jewish, Hindu and/or other religious doctrines, as well as some worldviews that do not rest on any comprehensive philosophical or religious doctrine, and some of their various adherents are unlikely to accept Kant’s autonomy (or Mill’s individualism, on the side of utilitarianism) as central values. Hence, in Political Liberalism Rawls presents his two principles of justice as the object of an overlapping consensus among adherents to competing—but reasonable—comprehensive doctrines.

After this development in Rawls’s thought, the original position still represents the reasonable and the rational as applied to the problem of selecting principles of justice, but now in a circumscribed way. Also, now the idea that to act from principles chosen in the original position is to act autonomously no longer carries the weight that it did in Theory. In Political Liberalism, Rawls distinguishes between rational autonomy, full autonomy, and ethical autonomy, and he argues that selecting principles of justice from the original position expresses rational and full autonomy. Adherents to religious and to non-Kantian philosophical comprehensive doctrines and to other worldviews are likely to have their own central values in the place Kant reserves for (what Rawls now calls)
ethical autonomy, but it is nevertheless decisive that Rawls’s principles of justice provide reliable support for them to advance their own notions of the good.

Not all comprehensive doctrines are reasonable, of course. Some promote coercive proselytizing and would use state power to impose their own views on others, and some doctrines are simply mad. Adherents to reasonable doctrines seek to deal fairly with adherents to other doctrines, so they want society’s laws and institutions to be based on reasons all can support. In *Political Liberalism*, Rawls emphasizes the role of the original position as the basis for an overlapping political consensus among adherents to reasonable but competing comprehensive doctrines. While each person’s idea of the good (and therefore what is rational for her) may be specified in terms of her own comprehensive doctrine, Rawls’s two principles of justice provide a reasonable basis for political institutions that will support her liberty and her capacity to advance her idea of the good, whatever it turns out to be. Adherents to reasonable comprehensive doctrines adopt the reasonable constraints imposed by the original position to establish laws that all reasonable people can support.

A central difference between the systems of ideas in *Theory* and *Political Liberalism* is expressed in the distinction in the latter work between rational and full autonomy on one side and ethical autonomy on the other. In *Political Liberalism*, Rawls argues that the original position models citizens’ rational and full autonomy and in doing so provides a particular way for citizens to think of themselves as free and equal. Rational autonomy, according to Rawls, is shown in persons “exercising their capacity to form, to revise, and to pursue a conception of the good . . . to deliberate in accordance with it . . . [and] to enter into agreement with others [to support it],” and full autonomy is demonstrated when citizens not only comply with principles of justice but also act from these principles as just. In the original position there is no external standard, such as God’s law or a historically established constitution, for determining what is just. The parties themselves have to decide what principles are best, considering what is most likely to advance the good of the citizens they represent. Moreover, the parties recognize that citizens possess a reasonable and a rational capacity, a sense of justice and an idea of the good, and they (the parties) aim to select principles that will support the
complete development and full exercise of these capacities. In these ways the original position models citizens’ rational autonomy.

Rawls argues that the original position models full autonomy through its structure, that is, with the parties deliberating behind a veil of ignorance so they are fairly situated to select principles for basic laws and institutions. The structure of the original position also fixes what are to count as appropriate reasons for selecting principles of justice and requires that (if possible) the parties should select principles “that may be stable, given the fact of reasonable pluralism; and hence . . . that can be the focus of an overlapping consensus of reasonable doctrines.” By establishing conditions in which its parties act justly the original position models the full autonomy of citizens in society.

Note, however, that full autonomy for citizens, as opposed to that modeled by the original position, “is realized in public life by affirming the political principles of justice and enjoying the protections of the basic rights and liberties; it is also realized by participating in society’s public affairs and sharing in its collective self-determination over time.” In *Political Liberalism*, Rawls emphasizes the distinction between full autonomy and ethical autonomy, “which may apply to the whole of life, both social and individual, as expressed by the comprehensive liberalisms of Kant and Mill.” *Political Liberalism* “affirms political autonomy for all but leaves the weight of ethical autonomy to be decided by citizens severally in light of their comprehensive doctrines.”

As noted above, the way rational and full autonomy are modeled in the original position supports citizens in thinking of themselves as free and equal. We have seen that citizens appreciate that they are recognized, in the device of representation for selecting principles of justice, “as having the moral power to form, to revise, and rationally to pursue a conception of the good,” and that the parties make it their priority to support all citizens in securing the requisite primary goods. They see that all are recognized equally as having reasonable and rational powers “and other capacities that enable us to be normal and fully cooperating members of society[, and that] all who meet this condition have the same basic rights, liberties, and opportunities, and the same protections of the principles of justice.” Hence, citizens see that to live in a society with laws and institutions based on principles selected in the original position

© 2012 University of Notre Dame
would support their self-respect, and that it is a worthy endeavor to work toward establishing such a society.

In order to clarify Rawls’s arguments it may be helpful to respond to some challenges to them from Amartya Sen. Rawls’s aim in _Theory_ and _Political Liberalism_ is to identify a few basic principles to guide the design of the main social institutions in reasonably well-off states, aiming to improve on those of the most prominent alternatives, utilitarianism and perhaps perfectionism. He holds that the original position defines the appropriate perspective from which to select such principles, and that from this perspective his two principles would be chosen. Sen, however, is not convinced that there would be any unique choice in the original position. “There are genuinely plural, and sometimes conflicting, general concerns that bear on our understanding of justice. They need not differ in the convenient way—convenient for choice, that is—that only one such set of principles really incorporates impartiality and fairness, while the others do not.” Sen argues, in fact, against all “transcendental” approaches that propose particular principles to uniquely define an ideally just society. Rather, he argues, our plural ideas of justice often permit the identification of unambiguous improvements over existing states of affairs, and theories of justice should support just such incremental improvements.

To respond to Sen we need to rehearse the justification for Rawls’s two principles and the boundaries to what Rawls expects from them. Given that each of us has a set of interests—a conception of the good, and, in due course, a rational plan of life to achieve it—we need to acknowledge our vulnerability. Casual perusal of current affairs is sufficient to see how easily and often hopes are dashed by poverty, authoritarian government, or domination by powerful economic interests or by ethnic or religious factions. The defense of liberty and of economic opportunity found in Rawls’s principles provides critical guarantees, perhaps the best general guarantees society can offer, that each person will have the opportunity to develop and carry out his or her life plan. In a polity dominated by organized interests, the radical demand of the original position is to set aside our individual and group interests when it comes to political fundamentals and to develop a political program based on this conception of our common good, in order that a just society should be built and sustained. Rawls’s two principles are favored in the original position because