
Institutions and Their Design

ROBERT E. GOODIN

THE PRINCIPAL AIM of this introduction is to sketch the contours of the existing literature as it touches upon theories of institutional design. It will situate the contributions of the present collection as well as map the domain of the larger series within which this collection is set.

The chapter starts by tracing continuities and discontinuities in the study of institutions, in the first instance, on a discipline-by-discipline basis. I avoid prematurely and preemptorily defining the term “institution,” preferring instead to let each discipline speak with its own voice. Inevitably, each discipline (and indeed subdiscipline) focuses on different institutions as paradigmatic and picks out different characteristics as their defining features. I propose to harness rather than straight-jacket this diversity. Once I have let all these disparate disciplines have their say in their own ways, all of them will then be consolidated into a few broader reflections upon the form and function of social institutions. As part of that, I identify a minimalist definition of “institutions” upon which most institutionalists, old and new and across a range of disciplines, can broadly agree.

That analysis provides a platform from which to address, in the final two sections, questions about the perfectability of social institutions. Key questions there concern the extent to which institutions acciden-

An earlier, very different paper served as background for the ANU conferences from which most other chapters of this book are drawn. I am grateful to Geoff Brennan for helping to develop, and to conference participants for helping us to clarify, those issues. I am also grateful to Paul Bourke for the opportunity to try out these ideas on the Australian Historical Association and for other comments and criticisms from David Austen-Smith, John Braithwaite, John Dryzek, Patrick Dunleavy, Patricia Harris, Barry Hindess, Claus Offe, Philip Pettit, Peter Self, Barry Weingast and, most particularly, from John Ferejohn and Diane Gibson.

tally emerge or automatically evolve, and the extent to which they are subject to intentional design and redesign. Insofar as intentional (re)design of some sort or another is a feasible aspiration in any sense at all, further questions arise about what sorts of principles, both normative and empirical, might properly be employed in those endeavors.

1.1 Institutionalisms, Old and New

Each of the several disciplines that collectively constitute the social sciences contained an older institutionalist tradition. In each case that tradition has recently been resurrected with some new twist. Just as the older institutionalism within each discipline had focused on some slightly different aspects of the phenomenon, seeing social institutions as solutions to the problems which each respective discipline deemed central, so too does the new institutionalism mean something rather different in each of these alternative disciplinary settings. Each of these perspectives has something to contribute to a more rounded view of the ways in which institutions shape social life. But the advantages that come from building toward the larger truth in this way can come only through the realization that the “new institutionalism” is not one thing but many.¹

1.1.1 New Institutionalism in History

Not so long ago, history was principally political history, the study of wars and kings and courtly intrigues. Of course it is also true that history is, and has always been, essentially a matter of storytelling; and a good story requires a good *dramatis personae*. Thus traditional political history has always been highly personalized and in it institutions are always inevitably personified: states by their princes, estates of the realm by their friends at court, and so on.

Although told through stories of striking personalities and individual daring, traditional histories were essentially stories of political institutions, their shaping and reshaping. Kings and courts, states constituted around them, and wars between them are institutional artifacts – the products of political organization. In that sense, history as a discipline has traditionally been highly institutional in its fundamental orientation.

Traditionally, though, it was the specifically political subset of social institutions that attracted most of the historian’s attention. The work-

¹In ways well captured in, e.g., Smith’s 1988 account.

ings of social and economic institutions were certainly also considered, of course, but basically just as they touched upon the political. That focus upon several institutions, but essentially just political institutions, constitutes what I shall here characterize as history's "old institutionalism."

Over the course of this century, the understanding of history as the study of kings and wars has fallen out of fashion. Political history has gradually given way to social history. In that shift – symbolized as much as precipitated by the *Annales* movement – history has increasingly come to be seen as the history of everyday life. And if what counts as central in history is the lived experience of the past, court history is largely irrelevant. Historically, courtly doings have impinged only on the margins of everyday existence – even, historically, when they resulted in war.

In recent years, the focus of history as a discipline has shifted once again. The everyday life of the ordinary person, we have come to appreciate, does not stand alone and utterly apart from the rest of the larger society. There has, accordingly, been a shift back to the study of larger social institutions. The focus of these new histories is not necessarily upon kings and courtly doings (or their contemporary equivalents: presidents and cabinets, legislatures and judges, financiers and media magnates). Certainly newer forms of historical inquiry do not focus on that to the exclusion of all else, in a way that older forms of historical inquiry might once have done.

The newer focus is, at one and the same time, both broader and narrower. The focus is both upon social institutions more broadly (like churches, the family, and the labor market) and upon organs of the state more narrowly (like the workings of social relief agencies, the Children's Bureau, or public works or public health agencies).² As a result of this reorientation, history is once again largely a story about the workings of social structures, albeit now with a new focus upon the actual impact of those structures on real people's ordinary lives.

The peculiarly historical contribution to institutionalism, old or new, lies in history's fixation upon the past. If in social scientific terms each discipline "owns" one particular variable, time is history's. Insofar as it has social scientific aspirations, history is just the study of the way in which the past shapes the present and the future. Or, in less Whiggish mode, we might say that history just amounts to the telling of stories

²For good samplers, see Evans et al. (1979) and Steinmo et al. (1992). For exemplary studies, see Theda Skocpol's *Protecting Soldiers and Mothers* (1992) or Karen Orren's *Belated Feudalism* (1991).

about the past which we internalize as our own and which, in the telling and retelling, shape us and our future actions.³

That the past exerts this sort of influence over us is the central claim of history as a discipline, and that is the peculiar emphasis it imparts to its various forms of institutionalism. Questions of why and how the past exerts that sort of sway are essentially left to other disciplines to resolve. Good historians naturally speculate upon the psychological or sociological or political dynamics at work. They will gesture toward the satisfaction that comes from fitting one's life into some large narrative structure, or to the historical construction of the *conscience collective*, or to organization as both the mobilization and ossification of bias. But gestures these typically remain.

Even when historians' attention is caught by the workings of social institutions, their interest is in the particular tale surrounding the particular institutions at the particular historical juncture. The peculiarly historical interest tends not to lie in what broader theory can be constructed around those and other cognate cases. Indeed, those of a peculiarly historical cast of mind often shy away from such larger generalizations, thinking that they necessarily do violence to the historical particularity surrounding each of the specific cases that together, in very stripped-down form, constitute the more scientific social scientist's "data."⁴

1.1.2 New Institutionalism in Sociology

From its beginnings, sociology too was essentially concerned with the study of social institutions. At the outset, this concern took the form of fixating rather unimaginatively upon a standard catalogue of institutions. Herbert Spencer's 1879 *Principles of Sociology*, for example, progresses through a tedious array of ceremonial institutions, political institutions, professional institutions, industrial institutions, and so on.

The ensuing classics of modern sociology imposed far richer theoretical overlays upon such pedestrian partitionings of the sociological problematic. But it is fair to say that all the masters of modern sociology – Pareto, Mosca, and Michels; Tönnies and Durkheim; Simmel

³That alternative formulation crucially differs in that it acknowledges the ways in which we read the past in light of the present. But centrality of "the past" remains: what gives these largely fictitious reconstructions the power of "history" is precisely the fact that they are fictions about the past.

⁴Contrast, for example, Tilton's (1990) study of the peculiar circumstances surrounding the foundation of the Swedish welfare state with Jackman's (1972) sixty-nation study of the emergence of welfare states in general; for a nice compromise between the two, see Esping-Anderson 1990.

and Weber; and, most especially, Marx – were all centrally concerned with the ways in which collective institutions subsume and subordinate the individual. All those classic studies – of “organic solidarity” as much as studies of “mechanical solidarity,” of identification theory as much as of organization theory, of “base” as much as of “superstructure” – essentially amount to stories about mechanisms for effecting social control over individual volition.⁵

Nowhere is this blending of concerns clearer than in the work of Talcott Parsons, on some accounts the greatest contemporary sociologist. His early work on *The Structure of Social Action* was officially concerned with the sources of voluntaristic individual action. But his collaborative *General Theory of Social Action* then formalized what in that earlier work were merely inchoate notions of “the social system” as a control mechanism, notions which achieved their fullest (and most rococco) elaboration in his later book of that same name.⁶

Inevitably there then came a reaction against what was seen as an overemphasis upon the way individuals’ volitions were shaped by collective social structures. Critics complained of “the oversocialized conception of man” in the dominant structural-functionalist sociology, entering pleas instead for “bringing man back in.”⁷ The ways in which these midcentury critics proposed doing that varied, from essentially phenomenological stories about the “social construction of reality” to social-psychologically inspired behavioralism and “action theory.”⁸

Whatever precise form the countervailing theories took, their basic thrust was to downgrade (but without ever totally denying) the importance of collective social structures and institutions in determining the actions and choices of individuals and groups within any given society. The emphasis within this midcentury sociological reaction against old institutionalism was upon the role of individual and collective choice as against social-structural determinism of all outcomes of social (or anyway sociological) consequence.

The “new institutionalism” within contemporary sociology is a reaction against that reaction, in turn. In part it amounts to little more than picking up older institutionalist-cum-structuralist themes and giving them a different normative spin. Old structuralist-institutionalists tended to be conservatives: observing structures, they tended simply to assume that they made some functional contribution to social stability;

⁵One of the best early renderings of that increasingly submerged theme in sociology is E. A. Ross’s (1901) little book on *Social Control*, originally published as a series of articles spread across the first three volumes of the *American Journal of Sociology*.

⁶Parsons 1937. Parsons and Shils 1951. Parsons 1952.

⁷Wrong 1961; Homans 1964.

⁸Berger and Luckman 1966. Manis and Meltzer 1967. Goffman 1970.

and they were inclined to celebrate that fact, thoroughly approving of the various ways in which the collective conscience got a grip on individuals.⁹ One group of new sociological institutionalists are basically radicals who, observing the same phenomenon (“structuration,” and domination of individual and group agency by structural determinants more generally), bemoan the ways in which such institutions exercise hidden power over helpless social agents, be they individuals or marginalized groups.¹⁰

Other new institutionalists within sociology, often no less radical in their politics, confine themselves to more narrowly analytic points. Individual action, they observe, is “embedded” within the context of collective organizations and institutions. Those actions are shaped, and their effects affected and deflected, by the institutional contexts in which they are set.¹¹ New sociological institutionalists of this stripe point, in particular, to the important role that intermediate organizations can and do play in shaping and reshaping both individual actions and collective outcomes emanating from them. The family is one obvious example very much at the center of many current controversies.¹² But theories of civil society, the density of institutional networks and mediating structures quite generally also figure largely in such accounts.¹³

Feminist accounts of the family and cognate social institutions combine these two approaches. Such mediating structures, they say, do so much mediating that certain institutionalized patterns of dominance and subordination disappear altogether from view. Feminists focus in particular upon the “public/private dichotomy,” which is so central to the self-conception of liberal societies; and they proceed to show the various ways in which collective, institutional relations of power and dominance reach into what were supposed to be purely privately ordered spheres. Thus, feminists say, opponents of social oppression need to examine the power relations embodied in the mediating structures of the putatively “private” sphere just as much as they do those embodied in more straightforwardly public institutions.¹⁴

⁹Contested though this familiar charge may be, as against American structural-functionists, it is frankly and forthrightly true of the most influential postwar German institutionalist, Arnold Gehlen (Berger and Kellner 1965).

¹⁰Giddens 1984. Lukes 1974. Dahrendorf’s classic 1958 essay “Out of Utopia” is a calmer precursor in broadly the same spirit.

¹¹Granovetter 1985; 1992.

¹²Berger and Berger 1983.

¹³Theoretical speculations (Hirst 1994; Cohen and Arato 1992) are powerfully confirmed by Putnam’s (1993) painstaking empirical analysis of the causes and consequences of civic traditions in modern Italy.

¹⁴Macintosh and Barrett 1982. See also Elshtain 1981, Pateman 1983/1989, and Okin 1989.

Were we assigning “key variables” to disciplines, the one owned by sociology might be said to be “the collective.”¹⁵ The old institutionalism within sociology focused upon ways in which collective entities – the family, the profession, the church, the school, the state – create and constitute institutions which shape individuals, in turn. The new institutionalism focuses, more modestly perhaps, upon ways in which being embedded in such collectivities alters individuals’ preferences and possibilities. But it is the hallmark of sociological institutionalism, whether old or new, to emphasize how individual behavior is shaped by (as well, perhaps, as shaping) the larger group setting.

1.1.3 New Institutionalism in Economics

Within economics, the dominant tradition has long centered around a neoclassical paradigm involving idealized free agents interacting in an idealized free market. Within that model, order and patterns emerge out of those interactions: they do not prefigure it. But there has long been, both within European public finance and American economics more generally, an “institutionalist” tradition predating and (especially in America) crystalizing explicitly into an opposition movement against that neoclassical orthodoxy.

The original notion of American institutional economics, as promulgated by John R. Commons and his followers, was to examine ways in which collective action can be institutionally embodied and in that form shape and constrain subsequent individual choice.¹⁶ The more positive and constructive side of the project was to study institutions and mechanisms – property law and the rules of the courts enforcing them, particular exchanges and the practices governing them, credit institutions and merchant banks, trade unions and trade associations, and so on – which create and control economic life. The more negative aim was to undermine the neoclassical orthodoxy by showing the many ways in which its idealized notions of “free markets” misrepresent the institutional reality of any actual economy, which is in fact dominated by actors (corporations, classes, central banks, or whatever) with precisely the sort of power to shape market outcomes (especially but not exclusively by altering relative prices of commodities and resources) that is assumed away in fictitiously idealized free markets.¹⁷

Still, the neoclassical paradigm remained much the dominant orthodoxy within twentieth-century economics. As that paradigm consoli-

¹⁵“Class” is, of course, just a special case of that more general formulation.

¹⁶Commons 1931, 1934.

¹⁷The papers collected in Samuels 1988, and especially the editor’s introduction, provide a good overview of these themes.

dated its hold on the profession, its ambitions ranged ever more widely. From the outside, it seemed to be a clear case of microeconomics run amok, staging a takeover bid not only for macroeconomics but also (in its “public choice” guise) exercising imperial ambitions over large areas of explicitly “nonmarket” (especially political) decision-making as well.¹⁸

The highest aspiration and continual quest, within this neoclassical program, was to provide “microfoundations” for macro-level phenomena in economics and elsewhere.¹⁹ The aim was to reduce all economic behavior – and, ultimately, all social behavior more generally²⁰ – to the interaction of individual preferences under conditions of scarcity (meaning just that not all of everyone’s preferences can simultaneously be completely satisfied). Smooth and more-or-less instantaneous response functions were simply assumed; “stickiness,” deriving from institutions or otherwise, was essentially just assumed away. Constrained maximization was the basic analytical device, equilibrium the preferred solution concept.

The attempted reduction never quite came off. But along the way the essentially anti-institutional program of neoclassical economics was powerfully sustained, both positively and negatively. Among the most crucial positive breakthroughs was Arrow and Hahn’s *General Competitive Analysis*, providing a proper microeconomic proof of the fundamental theorems of welfare economics (roughly speaking, that Adam Smith was right: the invisible hand really will, under idealized assumptions, work precisely the magic claimed for it).²¹ Negatively, too, Arrow contributed powerfully to the loss of faith in power of collective decisions to decide anything (that is the way his General Impossibility Theorem, showing that majority voting can lead us around in circles, was standardly – if not altogether correctly – interpreted²²) and to the loss

¹⁸To borrow a phrase from the title of the first issues of what later became *Public Choice*, the flagship journal of this movement.

¹⁹Weintraub 1979.

²⁰The work of Gary Becker (1976, 1981) best epitomizes this vaunting ambition.

²¹Arrow and Hahn 1971.

²²Arrow (1951), as interpreted particularly by, e.g., Riker (1980, 1983). That interpretation is variously flawed, however. All that Arrow’s proof actually shows is that cycling can happen, not that it necessarily will. Furthermore, that is true only on certain further assumptions which may not actually obtain (e.g., on the assumption that voters’ preferences are not substantially “single-peaked” across the community). Cycling is not in any case inevitable, because giving up on transitivity is only one among many possible responses (accepting someone’s preferences as “dictatorial” is another). Finally, note that the Arrow theorem applies as much to markets as to politics: the same results should apply equally in aggregating preferences there, too; and the main way we get equilibria in market economies is, in practice, precisely

of faith in the power of collectivities to organize to implement anything (which is the explicit thrust of his influential lectures on *The Limits of Organization*²³).

Latterly, however, there has been a resurgence of interest in institutional economics, reacting against the hyperindividualism of this microeconomic putsch. This resurgence is represented most notably, perhaps, in the work of Nobel laureate Douglass North. But because his work is in economic history, the power of the general points he makes – about the importance of institutional frameworks as background conditions for the emergence and operation of markets as we know them – might be blunted.²⁴ Perhaps a better representative for this purpose might be Oliver Williamson, whose transaction-cost economics generalizes Coase's observations about the nature of the firm into a larger theory, *The Economic Institutions of Capitalism*.²⁵

Whomever we choose as our representative, the basic point of this new institutionalism in economics is to show the various ways in which the actual functioning of the free markets of neoclassical economics requires and presupposes a fair bit of prior institutional structure. Most especially, institutions reduce transaction costs and in that way facilitate exchange. They promote trades, and hence trade.²⁶

Neoclassical economists value unfettered trade because it helps people to realize their desires and give effect to their choices, to the maximal extent possible within the limits imposed by scarce resources. Institutions facilitate trade. The way they do so, though, is precisely by constraining choice. If we could not sign a binding contract (or its social equivalent: stake our reputation on a binding promise) then we could never enter into any deferred-performance exchanges, in which one party has to act first, trusting the other to act later. Putting ourselves in a position to be sued, should we fail to keep our contract, is putting ourselves under a constraint – but it is one that we welcome, for absent

through imperfect (monopolistic or oligopolistic) competition creating concentrations of power that make some actors' preferences to a greater or lesser extent "dictatorial" in just this way. I am grateful to David Austen-Smith for forcefully impressing all these points upon me.

²³Arrow 1974. See similarly Wolf 1990.

²⁴North 1990. For a splendid case study – but also an example of the sort of thing that leads people to dismiss such work as consisting of purely historical curiosities – see his marvelous coauthored paper on the origins of the Law Merchant in the medieval trade fairs (Milgrom, North, and Weingast 1990).

²⁵Coase 1937. Williamson 1985. For a sociological perspective on the same themes, see Granovetter (1985, 1992).

²⁶Coleman (1990, esp. chap. 3) offers a compelling example of this, building up to a central clearing house and money economy from the increasing efficiency of mediated exchanges over barter ones.

such a system of constraints a raft of mutually beneficial exchanges could never take place.²⁷

In saying that institutions facilitate trade, it must also be recognized that institutions facilitate some trades more than (and in a world driven by relative prices, at the expense of) others. Institutions similarly facilitate certain trading channels and partnerships at the expense of other possible ones. In that way, institutions do indeed introduce and reinforce biases in favor of some interactions and interacting agents and against others. None of that should come as a surprise. After all, institutions are in essence just ossified past practices and the power imbalances and bargaining asymmetries embodied in them.²⁸

Emphasizing the importance of economic institutions thus amounts to emphasizing the importance of things past as determinants of present economic choices. That is something that the neoclassical paradigm, in its purest form, would have hoped to wash away.²⁹ Borrowing a phrase from Hume's precepts about causation more generally, the neoclassical paradigm in economics would have liked to be able to insist that there can be "no action at a distance" – here, at a temporal distance. It would have liked to be able to insist that complete descriptions of the present state of affairs, together with suitable covering-law style generalizations, are all we need to make reliable predictions about future states of affairs.

In a way, that is certainly true. Unless the past leaves a residue in the present, it is incapable of influencing either the present or the future through it.³⁰ But the past does leave such residues. Among those traces are the institutions created by past behavior and choices. Also among them is the impact of past choices and experiences in shaping our present preferences.³¹ In these and many other respects, what matters in our economic (and other) choices is not just where we sit at the moment but also how we got there. Most phenomena in social life manifest "path dependence" of just this sort. That influence of the past on the

²⁷The contract example is offered by Schelling (1960, p. 43) and elaborated by Hardin (1982b, pp. 260–62); the point is further generalized by Streeck (1992).

²⁸Knight 1992.

²⁹Except, perhaps, by reference to sunk costs in existing plant and lock-in to particular technologies at particular historical junctures (Arthur 1989). But for neoclassicists, while differential costs of technological retooling might help to explain the relatively greater competitiveness of some economies than others at any given moment, sunk costs are of course to be ignored in deciding what to do next. Their advice is always to look to the present and future, not to the past, in framing choices. By-gones are bygone forever.

³⁰Elster 1983, chap. 1.

³¹von Weiszäcker 1973; Stigler and Becker 1977.

present, although scorned by neoclassicists, is precisely what is central to institutionalism, both old and new, within economics.

Were we assigning key variables to particular social sciences, “choice” would have to be the one allocated to economics. Precisely because notions of individual choice are so central to economics, discussions of collective institutional constraints on those choices have always remained as a somewhat peripheral subtheme within the larger disciplinary discourse. Still, it is an important subtheme. Choices are always made in a context – and the context is not set by material scarcities and production technologies (or, to shift to another analytical frame that nonetheless shares the same myopia, by the modes and forces of production) alone.

Not only does economics need to appeal to some institutionalist story to set the context for individual choices. Through its emphasis upon individual choice and the ways in which those individual choices concatenate into collective ones, economics is also capable of providing a distinct (albeit, perhaps, distinctly partial) perspective upon the genesis and evolution of those institutions through the past choices of individuals and groups.

1.1.4 New Institutionalism in Political Science

Much of the new institutionalism in political science derives directly from cross-fertilization from this new institutionalism in economics through the subdiscipline of “public choice” which so successfully straddles that larger disciplinary divide.

Within public choice, the particular problem to which the new institutionalism is a solution is principally this. In both neoclassical economics and public choice models of politics built around them, the name of the game is to find an equilibrium. In that context, the Arrow-Condorcet problem – the specter of a perpetual disequilibrium, resulting from the intransitivity of majority voting – seems to be a (indeed, perhaps the) central problem in political life.³² Political institutions that constrain the possible range over which such voting might cycle provide one nifty solution to that central problem. Institutional devices such as committee structures in legislatures, or bicameral legislatures, or judicial norms or constitutional constraints on the majoritarian de-

³²Arrow (1951), especially as elaborated by Schofield (1976) and McKelvey (1979). Mathematically, these results are highly robust (Sen 1982, 1993). But, as I have noted in my discussion of economics above, their political and economic consequences are often greatly exaggerated.

cision-making can sometimes ensure a “structure-induced equilibrium” in a situation where none would otherwise exist.³³

Another crossover to political science from new institutionalism in economics takes the form of a public choice theory of the state bureaucracy, modeled upon the economic theory of the firm. The central problem in both cases is, within this tradition, conceived to be one of ensuring that the wishes of “principals” are actually carried out by those who are supposed to serve as their “agents.” How best to do this depends upon complex calculations of the comparative costs of monitoring in both directions, internally (keeping an eye on behavior of subordinates) as well as externally (keeping an eye on the quality of the goods and services delivered by outside providers). But in the public agency just as in the private firm, it sometimes makes most sense to minimize transaction costs by “institutionalizing” certain activities, internalizing some activities within a single organization rather than contracting privately with outside suppliers.³⁴

Focusing too tightly upon these important new institutionalist crossovers from economics, however, would lead us to overlook the distinctive perspectives on institutions that are to be gleaned from political science. After all, the new institutionalism in political science explicitly harks back to an honored old tradition within “Government” (as the discipline was then called) of studying the state in its institutional form: organization charts, agencies and bureaux, public administration quite broadly conceived, and what has subsequently come to be called “the state apparatus.” An exemplary text here might be Bryce’s masterly 1888 work on the *American Commonwealth*, which surveys the branches of national, state, and local governments, together with political parties and “social institutions” ranging from the bar and bench through Wall Street and the railroads.³⁵

The thrust of the “behavioral revolution” within political science – indeed, the revolution that gave birth to that more modern name for the discipline – was to deny that form mapped function, to deny that organization charts and institutional myths were accurate representa-

³³Shepsle 1979. Shepsle and Weingast 1981. Brennan and Hamlin 1992.

³⁴On the private sector, see Williamson (1975, 1981), building on Coase (1937). These calculations come out particularly clearly in Hitch and McKean’s (1960, chap. 12) discussion of internal markets and contracting in defense procurement. On public sector applications more generally, see Moe 1984, 1990, 1991; Niskanen 1971; Dunleavy 1991; McCubbins 1985; Calvert, McCubbins, and Weingast 1989, and, most especially, Weingast and Marshall 1988. For economic theories of rent-seeking, as applied to politics more generally, see Buchanan, Tollison, and Tullock 1980.

³⁵Bryce 1888.

tions of actual practice.³⁶ What matters is not what people are supposed to do, but what they actually do. Insofar as people internalize those role expectations, then form is a good guide to function and myth to practice; but insofar as they diverge (and this is often very far indeed), then it is function and practice, actual behavior rather than ruling myth, that students of real-world politics must study.

The truth of these central precepts of the behavioral revolution is undeniable. But the perfect generality of their applicability has been greatly exaggerated. For there are important respects in which institutions matter to behavior, and it is those to which the “new institutionalist” resurrection of that older institutionalist tradition within public administration points. The behavioralist focus usefully serves to fix attention upon agency, upon individuals and groupings of individuals whose behavior it is. But those individuals are shaped by, and in their collective enterprises act through, structures and organizations and institutions. What people want to do, and what they can do, depends importantly upon what organizational technology is available or can be made readily available to them for giving effect to their individual and collective volitions.³⁷

This presents itself to individual citizens as a constraint, to managers of the state apparatus as an opportunity. Governance – to use the new institutionalist catchphrase within public administration – is nothing less than the steering of society by officials in control of what are organizationally the “commanding heights” of society.³⁸

There are of course limits to the sorts of commands that might effectively be issued from those commanding heights. Managers of the state apparatus themselves face various constraints, both in what others will let them do and in what others will help them do. They are constrained both in their “relative autonomy” and in their “power to command” (i.e., to implement their decisions).³⁹ Old institutionalists might have been insufficiently sensitive to those constraints. But by the same token, behavioralist revolutionaries were insufficiently sensitive to those possibilities. For a moderately full account of political outcomes, we need

³⁶Within public administration, the most telling landmark was Nobel laureate Herbert Simon’s *Administrative Behavior* (1957); on the behavioral revolution more generally, see Ranny 1962.

³⁷This history is effectively traced, and these connections drawn, by March and Olsen (1984, 1989).

³⁸Indeed, *Governance* is the name of the unofficial journal of this movement in its Anglo-American form; in discussions on Continent, the phrase “steering” figures more prominently (Bovens 1990; Kaufmann, Majone, and Ostrom 1985).

³⁹Nordlinger 1981. Weaver and Rockman 1993.

to “bring the state back in” to supplement behavioralists’ essentially individual-level accounts of political dynamics.⁴⁰

The behaviorist account of individual action is essentially social-psychological in its inspiration. The other dominant account, within political science, builds more economic models of human action: upon notions of instrumentally rational, goal-seeking behavior. Within that tradition, too, the same belated recognition of the importance of institutionalized social conventions has recently emerged.

The principal analytic tool within rational-actor analysis of politics is game theory. In more popular representations of those techniques, the principal game serving as a template for all of social life is the infamous Prisoner’s Dilemma.⁴¹ The peculiarity of that game – what makes it such a good template for so many aspects of social life – is that it arguably captures the core of logic underlying the problem of collective (in)action.⁴² Within that game payoffs are structured in such a way that it is true, at one and the same time, that: (a) each player would be better off if all players pursued some cooperative course of action; (b) each player would be yet better off, whatever others did, defecting from that cooperative course; (c) the concatenation of such strictly dominant strategies for individual players in a one-shot playing of Prisoner’s Dilemma yields an outcome in which everyone defects, leaving all worse off than they would have been had everyone (including themselves) cooperated.

That turns out to be true, however, only in one-off playings of the game. When the game is iterated (played over and over again among the same players) the strategic structure of the situation changes dramatically. More cooperative forms of play are then rationally preferred, since each player will have a strategic interest in acquiring a reputation as someone who is prepared to cooperate with cooperative others.⁴³ Thus, cooperative norms evolve in this fashion over time even among the hyper-rational egoists of the most individualistic game theory.

⁴⁰By the same token, we have to “bring the state back in” to supplement excessively society-centered accounts that try to trace all outcomes to structural-functional imperatives and system dynamics, of one sort or another (see Evans et al. 1979, esp. Skocpol 1979; Krasner 1984).

⁴¹In any more refined applications, cooperative/coordination games loom equally large. See, e.g., Luce and Raiffa 1957; Snidal 1985; Ordeshook 1986.

⁴²Hardin 1982a.

⁴³See, e.g., Taylor 1987; Axelrod 1984, 1986; and Hardin 1982a,b; for an interesting application to cattle-trespass disputes and boundary fencing in the American West, see Ellickson 1991. Among “pure” strategies, tit-for-tat (doing whatever the other did last time) is the only exemplar, and it is the most famous such strategy in consequence. But there is in principle any number of “mixed strategies” which could yield better results for all players than endless mutual defection.

Through them conventions emerge and institutions arise, on even this most individualistic model of human behavior.⁴⁴

Thus there is convergence, within political science, from several directions upon the importance of institutions. Individualistic models – whether grounded in psychological models of individual behavioral propensities or in rational choice models of strategic calculation – cannot really explain individual choices, much less social outcomes, without some reference to the larger social institutional context of those individuals' actions. Behavioralists find they need to bring the state back in, game theorists find it emerging from within their models. Either way, institutions (political and otherwise) have once again come to the fore in political studies.

In addition to these intellectual currents within the discipline, there have of course also been developments within the external world which have served to remind students of politics of the importance of institutions. The breakdown of American hegemony toward the end of the post-World War II world led to an increased focus, among international relations scholars, on the various international regimes and institutions through which coordination was nonetheless achieved.⁴⁵ The breakdown of state socialism across central and eastern Europe with the end of the Cold War led to a flurry of political scientific interest in the sort of constitutions one might write, and the institutions one might try to impose, in a world in which the mediating institutions of civil society and the internal checks of civic virtue have so systematically been destroyed.⁴⁶ Finally, the breakdown of state sovereignty in western Europe, and the emergence of the European Community as an autonomous agent with serious power over its constituent members, has created renewed interest among political scientists there in questions concerning the proper design of federal institutions.⁴⁷

Were we assigning key variables among the social sciences, the key variable within political science would be "power."⁴⁸ The capacity for one person or group to control the actions and choices of others – or, better yet, to secure its desired outcomes without regard to anyone else's

⁴⁴Hardin 1982a,b. Elster 1989. For skeptical reflections on this account, see Bates 1988

⁴⁵Keohane 1984; 1986. Keohane and Nye 1977. Ruggie 1982.

⁴⁶Offe 1991; Offe and Preuss 1991; Elster, Offe and Preuss forthcoming. Similar issues arise, of course, with respect to the consolidation of "democratic transitions" elsewhere across the world, from Southern Europe and South Africa to Latin America and Asia (O'Donnell et al. 1986; Stepan and Skach 1993).

⁴⁷Sandholtz and Zysman 1989. Moravcsik 1991. Garrett 1992.

⁴⁸The nicest short statement of this position is still Dahl's 1963 *Modern Political Analysis*.

actions or choices – is what politics is all about. This larger disciplinary focus upon the causes and consequences of political power struggles imparts a peculiar cast to political science's institutionalism, both old and new. "Organization," in Schattschneider's famous phrase, "is the mobilization of bias."⁴⁹ The existence of institutions make certain things easier to do and other things harder to do. Holding positions within or control over those institutions gives some people greater capacity to work their will upon the world, at the expense of others lacking access to such institutionalized power resources. This was the focus of old-institutionalist analysts of *Politics, Parties, and Pressure Groups*, to borrow the title of V. O. Key's great book.⁵⁰ It remains central to the concerns of new-institutional analysts within political science quite generally.

1.1.5 New Institutionalism in Social Theory

All of this new institutionalism within each of the constituent social sciences speaks to larger themes within social theory more generally, ranging from the philosophy of social science to normative political philosophy. The issues raised by new institutionalists across the several disciplines have appeared and reappeared in those higher-order discourses in various guises over recent years. Inevitably, I will provide here only a stylized sketch of debates which often purport (occasionally rightly so) to query the very deepest truths of metaphysics or epistemology or ontology.

There has undeniably been a "return to grand theory in the human sciences."⁵¹ In part, that is just to say that large-scale, comprehensive explanatory projects are back on the agenda. The many such projects afoot basically bifurcate into two strands. One offers "structure" as the key explanatory variable, and supposes that a sufficient rich, elaborate explication of the ways in which structures shape social outcomes could in principle provide a completely comprehensive account. The other

⁴⁹Schattschneider 1960.

⁵⁰Key 1942.

⁵¹Skinner 1985. There is also of course now a retreat from "grand theory," under the poststructuralist, postmodernist banner. The emphasis there is upon contingency and uncertainty, of institutions as much as of anything else; in that version, however, poststructuralism lines up with the new institutionalism in history, which admits and indeed celebrates the historical contingency of the particular institutions we have but proceeds to use the fact of those institutions to explain what subsequently occurs. Others, sensitive to problems of contingency and uncertainty, see institutionalization as a solution to precisely that problem: that is the focus of, for example, the "new institutionalists" (although they do not themselves embrace the term) in the recent French social science debates (Wagner 1994).

strand sees “agency” as central, and supposes that a sufficiently rich, elaborate explication of the ways in which agents act and interact would in principle provide a completely comprehensive account of social outcomes.

This basic argument plays itself out in a myriad of ways, and a myriad of places. Sometimes it is portrayed as a dispute between disciplines, with sociologists cast as advocates of structural explanations and economists as advocates of individualistic, agency-based ones.⁵² Other times it takes the form of an intramural argument within some particular discipline or even within some particular theoretical tradition – most notably in recent years within Marxism, where we have seen the “rise and fall of structural Marxism” of an Althusserian bent, the emergence of a school of “rational choice” Marxism, and the attempt within that school to square essentially individualistic action premises with a broadly functionalist macrosociological framework of some recognizable Marxian form.⁵³

In the realm of social theory more generally, new institutionalism might be seen as nothing more (and nothing less) than the recognition of the need to blend both agency and structure in any plausibly comprehensive explanation of social outcomes.⁵⁴ Even the staunchest advocate of rational choice models as explanatory tools must concede that people’s preferences (which are the driving force in that model) have to come from somewhere outside the model; and one need not excavate very deeply to see that they come, ultimately, from structures of past experiences, prior socialization or social location.⁵⁵ And even the staunchest advocate of structural explanations cannot escape the fact that there have to be agents – albeit “socially constructed” ones – to serve as the carriers and enforcers of those structural imperatives, which those agents inevitably reshape in the process of reapplying and reinforcing them.⁵⁶

Proceeding in parallel to those discussions in empirical political and social theory is an allied dispute within normative political philosophy.

⁵²Barry 1970.

⁵³On the Althusserian moment, see Benton (1984). Rational choice Marxism reaches its fullest fruit in the work of Elster (1985a) and Roemer (1986). G. A. Cohen (1978) lays functionalist foundations he intends to be broadly consistent with that account, although the success of that particular aspect of his project is powerfully queried by Joshua Cohen (1982). Giddens’s 1984 work on “structuration,” previously mentioned, is perhaps the best of many attempts by Marx-inspired scholars unaffiliated with either camp to trace the interpenetration of agency and structure in a loosely Marxian framework, more generally.

⁵⁴Lukes 1973. Wendt 1987.

⁵⁵Satz and Ferejohn 1994. Gintis 1972.

⁵⁶Hindess 1988.

That is ordinarily posed as an argument between liberalism and communitarianism. In this highly stylized dispute, liberals are represented as championing Enlightenment models of human agency: individuals are rational, free-thinking, cosmopolitan, universalist, unencumbered. Communitarians point, in contrast, toward the ways in which individuals are inevitably embedded in social relations: when young, we all have to be taught something by somebody; and along the way we have all come to acquire various attachments to people and principles and projects growing out of our various social experiences. These, on the communitarian view, are the true "sources of the self."⁵⁷ In the real world, there simply is no completely independent, free-thinking, unencumbered self capable of performing the sorts of heroically universalistic calculations that figure so centrally in liberal Enlightenment just-so stories.

Here, too, there seems to be some prospect of a rapprochement between the two camps.⁵⁸ Enlightenment liberals can, and should, acknowledge the undeniable fact that everyone has been, and probably has to have been, raised in some particular culture or another, with its own distinctive values and concerns. Everyone has to start somewhere, and where you start and what baggage you bring with you makes it easier to move in some directions than in others. All those things can, and should, be readily conceded all round.⁵⁹ Even the staunchest liberal can easily agree to the importance of commitments (to other people, to principles, to causes) in people's lives: in Sen's wonderful phrase, rational maximization that fails to take due account of those sorts of values model the behavior not of "rational agents" as such but merely of "rational fools."⁶⁰

That we are all socially embedded in such ways does not necessarily imply that we can never transcend our original upbringing. That we start somewhere, and that we experience greater or lesser difficulties in overcoming the prejudices of that upbringing, does not privilege those prejudices. It does not mean that we cannot or should not try to achieve (or anyway approximate) the Archimedian point idealized by Enlightenment universalism. Liberals would staunchly insist that we can and

⁵⁷Sandel 1982; Taylor 1989.

⁵⁸By which I mean to say simply that liberal Enlightenment theorists have agreed that they should, and argued that they can, take on board much of what communitarians say as a matter of empirical sociology – without its having the devastating consequences for their larger theories that communitarians claim. Whether communitarian critics of liberalism will be satisfied with that concession is, of course, another matter.

⁵⁹Kymlicka 1989. Hardin 1995.

⁶⁰Sen 1977; see also Mansbridge 1990.

should try to overcome our particular prejudices and interests in judging what is right and good, both for our own societies and indeed for the world at large.⁶¹

Whether that residual liberal Enlightenment aspiration is in the end feasible, and if so how best to accomplish it, remain hotly debated questions in contemporary political philosophy.⁶² This is not the place to enter into that dispute. The important thing to note, in the context of the present discussion, is what can be agreed rather than what remains in dispute. The point is just that, in normative as well as empirical social theory, there has been a recent recognition of the importance of somehow blending accounts of social structure and human agency into some larger composite model of the human condition.

1.1.6 New Institutionalism: One from Many

New institutional themes thus emerge in a variety of forms in a variety of disciplinary contexts. Despite their differences, all those variations on new institutionalist themes are essentially, and importantly, complementary.

The new institutionalism is at root a reminder of the various contextual settings within which social action is set. Drawing together all those diverse disciplinary strands, a consolidated new institutionalism would serve to remind us, *inter alia*, of the following propositions.

1. Individual agents and groups pursue their respective projects in a context that is collectively constrained.

Among the many forms those constraints take, to some significant extent,

2. Those constraints take the form of institutions – organized patterns of socially constructed norms and roles, and socially prescribed behaviors expected of occupants of those roles, which are created and re-created over time.⁶³

Furthermore,

⁶¹Thereby achieving what Nagel (1986) regards as almost impossible: the “view from nowhere.”

⁶²The most notable recent contributions have been made by, and in response to, second-wave Rawlsian theorizing (Rawls 1993a,b).

⁶³Eisenstadt (1968, p. 409), notice, would distinguish between roles and institutions; in deference to his arguments there, I include non-role-specific normative expectations as well.

3. Constraining though they are, those constraints nonetheless are in various other respects advantageous to individuals and groups in the pursuit of their own more particular projects.⁶⁴

That is true in many ways, but in part it is because

4. The same contextual factors that constrain individual and group actions also shape the desires, preferences, and motives of those individual and group agents.

Elaborating on the nature of those constraints, we can say at least this much with confidence:

5. Those constraints characteristically have historical roots, as artificial residuals of past actions and choices;

and

6. Those constraints embody, preserve, and impart differential power resources with respect to different individuals and groups.

At the end of the day, however,

7. Individual and group action, contextually constrained and socially shaped though it may be, is the engine that drives social life.

Different new institutionalists from different disciplines would vary the emphasis across these various propositions. Indeed, some new institutionalists (some whole disciplines) may demur when it comes to any particular one of those propositions. Still, some such synthesis seems to capture the moving spirit of the new institutionalism as a whole. Indeed, it is the very breadth of that larger vision that makes the movement so compelling for so many from such diverse disciplinary backgrounds.

1.2 The Forms and Functions of Social Institutions

There is wide diversity within and across disciplines in what they construe as “institutions” and why. That diversity derives, in large measure, from the inclination within each tradition to look for definitions that are somehow “internal” to the practices they describe. The

⁶⁴Strikingly, this theme is central among institutionalists in economics, where the neoclassical paradigm would lead us least to expect it. John Commons (1931, p. 649) literally builds this into his definition of “an institution” as “collective action in control, liberation, and expansion of individual action.”

theoretical thrust is for an explanatory account that is at least in part hermeneutic: the aim is to provide an account of what institutions are, and why they arise, that would in some sense or another be recognized by participants themselves as describing their own behavior. When we are looking at contract law or congressional committees or whatever, what “the” institution is (and hence, to some extent, what “an” institution is in that setting) is traditionally tied up with what it does and why it is there. And since explanandums vary across disciplines, so do explanations – and hence so too do definitions of the very notion of an “institution” itself.

That definitional diversity derives, however, almost entirely from the inclination to opt for a discipline-based, theory-impregnated internalist-style definition of the term. Let us now shift our focus away from these internal accounts, and not worry so much about the role that particular institutions play in the lives of people in particular contexts. Let us instead adopt an “external” account of what institutions are and what they do. This is not a story that those engaged in the practice will necessarily recognize as their own: it will not describe their motives or goals or perceptions. Rather, it is a story told from the outside, by an observer not internal to the practice, about the effects of institutions. It is a story about “what they do,” not about “why they do it.”

Shifting to that externalist stance, it is much easier to come to an agreed definition of what an institution is. From this external point of view a social institution is, in its most general characterization, nothing more than a “stable, valued, recurring pattern of behavior.”⁶⁵ That characterization might be a little too general to be terribly helpful. We may, for example, want to further stipulate that an institution is necessarily a social phenomenon. Individuals are not themselves institutions, however “stable, valued and recurring” their patterns of behavior.⁶⁶ Neither do idiosyncratic habits of private individuals count as “institutions” for our purposes, however “stable, valued and recurring” the behaviors issuing from them.⁶⁷

Still, a relatively general characterization is precisely what is needed

⁶⁵Huntington 1968, p. 12. See similarly Eisenstadt 1968, p. 409.

⁶⁶Even if personal identity is inconstant over time (as Parfit [1984] suggests) or unstable at any moment in time (as the “multiple self” literature [Elster 1985] suggests), we would still be reluctant to describe the “negotiated settlement” that constitutes any given individual’s ongoing identity as being “institutionalized,” except in some highly metaphorical sense.

⁶⁷This is to recall Hart’s (1961, p. 54) distinction between habit and rule: the difference between our (valued) habit of going to the cinema every Saturday night and our collective rule of men taking off their hats in church.

to capture the diverse range of social activities that we would want to deem to be institutions and to theorize alongside one another as such. At a minimum, we would want to include institutions in:

- “the sphere of institutions of *family and kinship*, which focuses on the regulation of the procreative and biological relations between individuals in a society and on the initial socialization of new members of each generation”;
- “the sphere of *education*,” which extending the former “deals with the socialization of the young into adults and the differential transmission of the cultural heritage of a society from generation to generation”;
- “the sphere of *economics*,” which “regulates the production, distribution and consumption of goods and services within any society”;
- “the sphere of *politics*,” which “deals with the control of the use of force within a society and the maintenance of internal and external peace of the boundaries of the society, as well as control of the mobilization of resources for the implementation of various goals and the articulation and setting up of certain goals for the collectivity”;
- “the sphere of *cultural institutions*,” which “deals with the provision of conditions which facilitate the creation and conservation of cultural (religious, scientific, artistic) artifacts and with their differential distribution among the various groups of a society”;
- “the sphere of *stratification*, which regulates the differential distribution of positions, rewards, and resources and the access to them by the various individuals and groups within a society.”⁶⁸

A central defining feature of “institutionalization” across all these spheres is the stable, recurring, repetitive, patterned nature of the behavior that occurs within institutions, and because of them. “Institutionalism” has been characterized as “the process by which organizations and procedures acquire value and stability.”⁶⁹ In an institutionalized setting, behavior is more stable and predictable. Furthermore, that is not an incidental by-product of institutionalization – not merely the consequence of “coming to value a certain organization or procedure” for some independent reasons. Instead, that very stability and predictability is, to a very large extent, precisely why we value institutionalized patterns and what it is we value in them.⁷⁰

⁶⁸Eisenstadt 1968, p. 410.

⁶⁹Huntington 1968, p. 12. See similarly Eisenstadt (1964; 1968, p. 410, 414–18).

⁷⁰Soskice, Bates, and Epstein 1992.

This fact is absolutely central to new-institutionalist analyses within both political science and economics, in particular. The role of institutions, economically, is seen as reducing costs associated with uncertainty across time. The crucial mechanism by which that is done is through a system of “nested rules,” with rules at each successive level in the hierarchy being increasingly costly to change.⁷¹ An only slightly exaggerated way of putting that point is to say that there can be a market in anything only if there is not a market in everything. Property rights and contracts for exchanges of property would be worthless if it were cheaper to buy and sell policemen and judges than it is to buy and sell parcels of property. It is precisely the stability and reliability of the more deeply nested rules governing the judicial system that makes feasible the flux of the market. Similarly, much of ordinary political bargaining and exchange is possible only against the backdrop of the stability provided by more deeply nested, institutionalized rules, ranging from informal norms of congressional behavior and committee structures to the Constitution itself.⁷²

Think for a moment about the nature of constitutions. They are supposed by their nature to be enduring, stable, hard to change; and for that very reason, they typically contain within them requirements for very large majorities and extraordinary procedures to be followed for their amendment and change. But, upon reflection, surely it should be something of a mystery why successive generations ever feel bound by those rules. The Founders were not superhuman demigods. What they did was simply scrap one set of institutional arrangements and start afresh. Why should successive generations feel bound to live by their rules for amending the Constitution, rather than feeling free to do as the Founders did in their day and start afresh themselves?⁷³

The answer lies, of course, in the value that we all derive from having our activities constrained in precisely the ways that enduring constitutions do. Being able to embody certain fundamental agreements in presumptively unchangeable rules allows us to make commitments to one another that are credible, in a way that they would not be were they embodied merely in ordinary legislation that was subject to amendment or repeal by any successive annual assembly. Just as in individual affairs contract law (the right to put ourselves in a position to be sued) allows us to make commitments for deferred delivery to people that might not otherwise be credible, so too in collective affairs does constitutional law (the right to put ourselves in a position to be judi-

⁷¹North 1990.

⁷²Buchanan and Brennan 1985. Hardin 1989. Ordeshook 1992.

⁷³Ackerman 1991.

cially overruled) allow us to make commitments to minority groups and sectoral interests that might not otherwise be trusted.

1.3 The Emergence and Change of Institutions

Institutions, though relatively stable, are not however eternal and immutable. In this section I shall consider, briefly, larger issues concerning the emergence and change of institutions. In the first instance, the issues are primarily empirical: where do institutions come from, and according to what forces do they change? The appropriate scope, or lack thereof, for theories of institutional design is often thought to turn upon a peculiarly intentionalist answer to those questions: in Section 1.3.2 I argue against any such easy presumption, having first (in Section 1.3.1) surveyed different generic styles of explanation of social and institutional change. I proceed, in Section 1.3.3, to survey design theories to try to adduce what “good design” might mean in these contexts. These design theories uncomfortably straddle both empirical and normative realms in ways that I try to sort out in Section 1.3.4.

1.3.1 Models of Social Change: Accident, Evolution, and Intention

There are, roughly speaking, three basic ways in which social institutions (or human societies more generally) might arise and change over time.⁷⁴ First, social change might occur by accident. There are, on this account, no forces of natural or social necessity at work, no larger causal mechanisms driving things. What happens just happens. It is – in a characteristically postmodernist turn of phrase that is also effectively captures the highly contextualized spirit of many microhistorical explanations – purely a matter of contingency.⁷⁵

Second, social change might be a matter of evolution. As in biological analogues from which these models borrow, the initial variation might have occurred utterly at random, as a matter of pure accident and happenstance. But there are, on these accounts, some selection mechanisms at work, usually competitive in nature, which pick out some variants

⁷⁴Though different in detail, this partitioning owes its inspiration and in crucial places its elaboration to Elster (1983); Knight (1992) usefully elaborates upon the evolutionary and intentional themes in rather different ways than those developed here.

⁷⁵Rorty 1989. There is in comparative politics a parallel, if less insistent, emphasis upon particular historical conjunctions as essential facilitating conditions for social change.

for survival.⁷⁶ Those variants which do survive over a protracted period might therefore be said to be somehow “better fitted” to their environment than those that did not.⁷⁷

Third, social change might be a product of intentional intervention. That is to say, the change might be the product of the deliberate interventions of purposive, goal-seeking agents. Those agents might be either isolated individuals or organized groups.⁷⁸ The changes that ensue from their intentional interventions may or may not be exactly what was intended by any one (or by any subset) of them. The changes may benefit some or all or none of the intentional agents.⁷⁹ Even where the outcome was intended by no one, however, the basic explanatory logic is still basically intentional in form, insofar as the story is still to be told essentially in terms of intentions and interactions among those intentions.

Any actual instance of social or institutional change is almost certain to involve a combination of all three of these elements. The problems that groups face, the solutions they concoct, and the way that they implement those solutions are all subject to accident and error. But the accidents and errors are rarely purely stochastic; and even when they are, they nonetheless typically arise in the backwash of intentionality, through the oversights and miscalculations of purposive agents engaged in projects of their own. Thus, what intentional agents intentionally do (or, more typically, fail to do) is important even in modeling social and institutional change as essentially accidental.

Intentionality has an even more central role to play in evolutionary stories about social and institutional change, for the “selection” mechanism that winnows out some variations in favor of others is in the social setting often essentially intentional in form. Agents, individually

⁷⁶This sort of analysis underlies everything from the Whig theory of history to neoclassical models of capitalist competition and one of the most recent theories of the development of the firm (Nelson and Winter 1982; Winter and Williamson 1991). Recent work on the emergence of cooperation in other social settings builds explicitly upon evolutionary models (see, e.g., Axelrod 1984, 1986; Ellickson 1991).

⁷⁷That may be faint praise, depending upon what you think of the competitive environment as a selection mechanism; just as bad art and music might coarsen people’s aesthetic sensitivities and thereby drive out good art and music, so too might some competitive environments select for the worst or the most common rather than the finest social forms.

⁷⁸The very fact that they have some internal decision mechanisms for settling upon joint action plans qualifying them as “intentional agents” for these purposes (French 1979).

⁷⁹“Benefiting all” recalls models which (after the fashion of Hume or Smith) analyze institutions in terms of conventions. “Benefiting some” recalls models which are more sensitive to power asymmetries and bargaining relations (Knight 1992).

or more often collectively, sometimes find themselves literally asked to decide which sort of social arrangements they would prefer to retain and reproduce. More often, they find themselves “voting with their feet,” deciding which among several alternative arrangements they would themselves prefer to participate in, with the institutions with relatively more participants being at a comparative advantage (economically, militarily, or whatever) vis-à-vis ones with fewer. In both these cases, the “selection mechanism” central to evolutionary models involves the intentional actions of purposive agents, either directly or indirectly.

Alternatively, that selection mechanism may take a more Hegelian form. This would ordinarily be described in terms of a “central animating idea” underlying any particular institution, with institutional change over time being analyzed in terms of that institution “working itself pure” in respect of that central idea.⁸⁰ One common example is the obvious tension, present at the founding of the American republic, between the “inalienable rights” ascribed to all men and the institution of slavery; on a Hegelian reading, which of course many would resist, the next half century of American history just amounted to this tension working itself out. For a possibly less contentious example, consider the history of the expansion of suffrage: once universal manhood suffrage was granted, it was very hard to provide any good, principled grounds for denying the suffrage to women or blacks as well. Many would tell a similar tale about expanding notions of citizenship rights more generally: once fundamental civil rights have been granted to all, there is an inexorable slide toward granting people basic political rights; and once those have been granted, there is another inexorable slide toward granting everyone certain basic welfare rights.⁸¹ Or, again, there might be a similar story to tell about the expansion of the European Union: the narrow initial idea of a free market in goods and labor necessarily entails a breakdown of barriers to free movement of labor across borders, which necessarily entails that pensions be made portable and social security entitlements uniform or transferable, so in the end much more than a common market for goods and services “naturally” evolves.

Whether or not any of those particular stories is compelling, the broader style of analysis that they represent certainly merits inclusion in this larger explanatory menu. What is going on, in each case, is most naturally described in terms of the “animating idea” of the system

⁸⁰I am grateful to John Ferejohn for impressing upon me the significance of this alternative in the present connection.

⁸¹Goodin 1992, chaps. 1,5. Marshall 1949. Another example might be the breakdown of master-servant law in the United States, as traced in Orren (1991).

somehow “working itself out.” It is still, in basic form, an evolutionary explanation, perhaps. But what is distinctive about this form of evolution is that the “selection” involved is done in terms of the “animating idea”: that, rather than in terms of pressures from a harshly competitive environment, explains the greater survival prospects of some institutional variants over others.

Thus this Hegelian story might be seen as basically evolutionary in form. But it is also at least in part intentional, as well. The selection may be being done *in terms of* the animating idea of the institution in question, but the selection is done *by* intentional agents working within and internalizing the animating ideas of that institution. In the case of slavery, it is the difficulty that the likes of Washington and Jefferson had reconciling their public and private lives – in good conscience denying their own slaves the treatment that their professed principles said was due to all men – that created the tension in question. In the case of expanding rights of suffrage or of citizenship, it is the difficulty of those resisting such expansions in finding any principled grounds for stopping anywhere along the slippery slope that creates the tension. And so on.

Animating principles are themselves inanimate. They are incapable of “working themselves out,” in any literal sense at all. What animating principles “animate” is intentional agents who internalize them, and what “working themselves out” amounts to is those intentional agents implementing them consistently across the whole range of their appropriate application. Thus, even in this Hegelian version of the evolutionary story of social and institutional change, intentional agents are still central players. In that, as apparently in all other cases, we will need an explanatory account that draws on notions of intentionality and perhaps accident as well as pure evolutionary pressures.

1.3.2 Intentionality and Design

Much more can and should be said under each of these headings, of course. Much more can and should be said about the various ways in which all three of these sorts of theories might be combined into a hybrid theory, more credible than any one of these standing alone.⁸² For present purposes, however, that very stark and sketchy taxonomy will suffice. The aim here, recall, is merely to try to situate theories of institutional design in relation to some larger explanatory menu.

It is often thought that theories which talk in terms of institutional

⁸²For case studies see, e.g., Binder et al. (1971) and Almond et al. (1973).

design are necessarily tied to intention-based theories of social change.⁸³ The source of the thought is easy enough to discern. After all, design and redesign are essentially intentional activities; so it is only natural to suppose that talking in terms of institutional design must be appropriate basically just where the institutions in view have intentionality at their core.

By the same token, it is only natural to suppose that theories conjuring with notions of “design” are out of place absent intentional design or designers. Sometimes institutions just emerge accidentally, in unintended ways, in response to some historical accident or another; sometimes they just evolve naturally, in unintended ways, according to some deeper logic of their own. Insofar as institutions are like that, products of accident or evolution rather than intention, and insofar as theories of institutional design presuppose intention, then theories of institutional design (whether empirical or normative) have a very limited role to play. Or so it is standardly supposed.

That way of thinking is, however, in error across several fronts. First of all, it construes the scope of intentional explanation too narrowly. Institutions are often the product of intentional activities gone wrong – unintended by-products, the products of various intentional actions cutting across one another, misdirected intentions, or just plain mistakes.⁸⁴ To explain how those outcomes came about, we must refer essentially to intentions and the interactions among intentions. The explanation is still intentional in form, even if the outcome is not intended. An institution can thus be the product of intentional action, without its having been literally the intentional product of anyone’s action.

The Myth of the Intentional Designer (still less the Myth of the Intentional Design) is greatly to be avoided in theories of institutional design. Typically, there is no single design or designer. There are just lots of localized attempts at partial design cutting across one another, and any sensible scheme for institutional design has to take account of that fact. Thus, even within the realm of our intentional interventions, what we should be aiming at is not the design of institutions directly. Rather, we should be aiming at designing schemes for designing institutions – schemes which will pay due regard to the multiplicity of designers and to the inevitably cross-cutting nature of their intentional interventions in the design process.⁸⁵

⁸³This is at the core of Oakeshott’s (1962/1991) critique of “rationalism in politics” and parallel complaints against social planning in Hayek (1973), for just two famous examples.

⁸⁴Merton 1936.

⁸⁵Dryzek’s (1990, pp. 40–50) proposals for “discursive designs” for “model institutions” and their real-world approximations are very much in this spirit.

Intentional design of institutions is only part of the story, though. To some extent, it is undeniably true that institutions do come about essentially by accident or that they evolve according to a logic of their own, in ways altogether impervious to intentional intervention and direct human design. Even where direct design is impossible, however, indirect design is often nonetheless feasible. Accidents happen: but the frequency and direction of accidents can be significantly shaped by intentional interventions of social planners.⁸⁶ Evolution in the strictest sense may proceed of its own accord, largely independently of the organisms subject to it. But especially in social settings, the criteria of success in the struggle for survival – the things that the analogue of “natural selection” selects for in a social setting, the things that contribute to the longevity of any given institution and the chances for its successors surviving into the distant future – those are all things that can and should be subject to conscious social choice.

Thus, while theories of institutional design undeniably find their primary application in corners of the social world subject to intentional intervention and control, it is simply not true that they are altogether without application to other corners of the social world. Insofar as the social world is accident-prone, we might want to design around the risk of accidents, seeking robust institutions that can withstand the various shocks that will inevitably befall them. Insofar as the social world is subject to evolutionary pressures, we might want to apply design principles to reshape the selection criteria and social reward structures according to which some innovations succeed and others fail.

What theories of social change as accident or evolution are telling us is that social outcomes themselves are not (or not to any great extent) directly subject to intentional change, design, or redesign. Such theories strive to limit the scope for intentionality in descriptive, or hence prescriptive, models of social life. Whatever their aims, however, what these other theories are actually pointing to are possibilities of design and redesign at one level up. Outcomes may be the product of accident, but accident rates might be intentionally altered. Outcomes may be the product of evolutionary forces, but the selection mechanisms that guide that evolution might be intentionally altered. Design and redesign might still have some scope, even in those less intentional social worlds.

When intentional agents cannot work their will directly, they start looking for ways to do so indirectly. In the ways just sketched, they may well succeed in their quest for indirect mechanisms even where direct intentions will inevitably be thwarted. Ironically enough, the less directly intentional the social world in which we find ourselves, the more the appropriate sphere for design principles shifts away from the inten-

⁸⁶Perrow 1984.

tional shaping of level of policies and outcomes, and the more it moves toward the intentional shaping of institutions and practices. Hence, the claim that the social world allows little scope for intentional change and direct design actually expands rather than contracts the scope of theories of indirect – that is, properly *institutional* – design.

Just as theories of institutional design are sometimes thought to imply an intentionalist pattern of social explanation, so too are those theories sometimes thought to imply a “creationist” focus. Talk of “design,” this critique complains, seems necessarily to amount to talking of “design de novo,” with insufficient appreciation for the ways in which social engineers always work with materials inherited from and to some extent unalterably shaped by the past.

Sometimes, perhaps, institutional designers do indeed come perilously close to such pitfalls.⁸⁷ By and large, however, their focus on institutions safeguards theorists of institutional design against any creationist fallacies to which a focus on design might otherwise render them prey. Much of the point of studying institutions, after all, is to explore precisely those ways in which the past leaves traces in the present and constrains our present actions and future options. So designers of institutions, of all people, should be particularly sensitive to the ways in which past inheritances will inevitably constrain them in their own design activities.

It serves as a useful reminder of this point to refer, from time to time, of design and *redesign*, shaping and *reshaping*. Those phrases are clumsy, and their frequent incantation would be unduly ponderous. But the larger point for which they serve as markers should be constantly before us: in designing social institutions, we are always doing so against the backdrop of a set of past practices, which brings with it its own peculiar constraints and possibilities.

1.3.3 Theories of Design

So far, this chapter has concentrated heavily on one key word, “institution,” at the expense of the other, “design,” in the title of the book and of the series which it introduces. The reason is, quite simply, that the same is true of all large literatures across all the several disciplines which this chapter attempts to track.

The phrase “institutional design” is often dropped into these discus-

⁸⁷Coleman (1993, p. 2), for example, talks of the legal creation of the “fictional person” of the corporation as a “social invention” – as if it were an utterly new social form, utterly without prior social precedents, when in fact there were precursors aplenty in Roman and medieval canon law (Berman 1983, pp. 215–21).

sions, even into the titles of articles and books on these topics. But the phrase is characteristically dropped relatively casually, with little analysis as to what it means for institutions (or anything else, for that matter) to be “designed” and still less analysis of what principles might properly guide such design attempts. To find any explicit discussions of “design” problems, we must stretch well beyond any narrowly institutionalist focus in search of principles and propositions which we can bring back to bear in that sphere.

Definitions are easily enough come by. Perhaps the most useful states, simply, that “design is the creation of an actionable form to promote valued outcomes in a particular context.”⁸⁸ But for further discussions, literatures on public policy and political institutions often refer us very far afield indeed – to texts in aesthetics or engineering or architecture or product design or land-use planning.⁸⁹ There may well be something to be learned for the study of institutional design from these distant disciplines. But, to say the least, the points of analogy and disanalogy will have to be traced fully and carefully: the objects of design are so very different that there can be no serious thought of wholesale borrowing of the tricks of those very different trades and applying them unreflectively to the design of social institutions. To date, however, those glib analogies to design notions in distant disciplines have remained just that.

Thus, we find ourselves in an awkward position. The paucity of literature specifically on design issues in the study of social institutions forces us to look further afield for guidance – but not too much further afield or we will quickly come to the point where lessons learned might well be utterly nontransferable. Happily, however, we can find much of what we need relatively near to home. There have recently emerged highly useful discussions of “design” problems on at least three distinct levels within cognate social science literatures. In order of increasing generality, these discussions concern the design of policies, of mechanisms, and of whole systems.

There is a modest literature rooted in political science on “policy design.” What counts as design in that context varies somewhat. Sometimes it amounts to little more than the generation of policy options: the crafting of new solutions, through a creative combination of recollection and innovation and a serious engagement with both values

⁸⁸Bobrow and Dryzek 1987, p. 201. For fuller elaboration on essentially the same ideas, see Alexander (1964).

⁸⁹See, most strikingly, Alexander (1982); but even in texts more sensitive to political context the references to specifically design notions always seem to dredge these same distant waters (see, e.g., Bobrow and Dryzek 1987, p. 200; Dryzek 1990, p. 41; Linder and Peters 1987; Goodsell 1992).

and contexts.⁹⁰ Often the central issue in policy design is “feasibility” in some sense or another: the implementability of policy choices, taking due account of the resources and incentives facing various agents and agencies who will have necessarily to be involved in giving effect to policymakers’ intentions.⁹¹

The general purport of the large literature on “mechanism design,” a literature rooted in economics, is to prescribe mechanisms for resource allocation quite generally.⁹² Such discussions go to the very core of modern economics, amounting to nothing less than the quest for credible models and possible mechanisms to underwrite economic equilibria, the absence of the fictitious auctioneer postulated by Walras.

Within that economics literature on mechanism design, “the major unsolved problem” lies in “the proper integration of the information and incentive aspects.”⁹³ Mechanisms are apparently doomed to fail in their attempts to allocate resources in Pareto-optimal ways if they attempt to respect, at the same time, constraints of “informational decentralization” and “incentive compatibility”: producers or consumers or both will have an incentive to deviate from the formal rules of the allocation mechanism (a failure of “incentive compatibility”); and they are able to do so by misrepresenting facts about which (thanks to “informational decentralization”) they have unique, privileged information (producers about their production functions, consumers about their preferences).⁹⁴ That is the basic shape of the “mechanism design” problem as it appears within economics, and in that form it has given rise to a large literature on the optimal designs of various resource-allocation mechanisms: auctions and bidding systems; contract with imperfect information; and so on.⁹⁵

Parallel mechanism-design problems appear outside these narrowly economic contexts, however. Within politics, there is an analogous theorem about voting schemes. In any voting mechanism that strives to achieve Pareto-optimal outcomes, some voter will always have a strategic incentive to manipulate the outcome by misrepresenting his or her true preferences.⁹⁶ Similarly, in the public finance context, there is

⁹⁰Alexander 1982. See also Wildavsky 1979 and Bobrow and Dryzek 1987 for less single-minded elaboration on these issues.

⁹¹Wildavsky 1979. Linder and Peters 1987. Schneider and Ingram 1988. Ingram and Schneider 1988.

⁹²Hurwicz 1977.

⁹³Hurwicz 1977, p. 32.

⁹⁴Hurwicz 1977, pp. 28–9.

⁹⁵For samplers, see Arrow and Hurwicz 1977 and Hurwicz et al. 1985; for a good survey, see McAfee and McMillan 1987.

⁹⁶Gibbard 1973. Satterwaite 1975.

an analogous problem in surmising “true” demand for public goods. It is no use just asking people how much they would be willing to pay for such goods, and levying a charge upon them accordingly: because the goods are public, people can use them freely once they have been provided by others’ payments; given that, people will have a strategic incentive to understate their “true” demand for public goods, if the supply of such goods will be substantially independent of what they say yet what they are charged strictly a function of what they say.⁹⁷ In both these more politicized contexts, there have been elaborate proposals for “optimal mechanisms” to circumvent such difficulties.⁹⁸

Finally, there is an even larger literature on “system design.” This literature is harder to place, and still harder to contain. For its initial inspiration, it drew heavily on operations research, computer modeling, and artificial intelligence.⁹⁹ In the first instance, the primary social scientific applications of those techniques were to problems in military resource allocation during World War II and, from the immediate post-war period, to cognate problems in economics more generally. From this tradition come techniques such as input-output analysis, econometric models (typically, Treasury models) of national economies, and schemes using quasi-markets and shadow prices in more systematic schemes of project planning and appraisal.¹⁰⁰ Latterly, similar techniques have been advocated for “systematic thinking for social action” across a wide range of social programs as well as in the economic and defense applications where such techniques have historically most naturally found their homes.¹⁰¹

The sorts of design considerations to which such systematic thinking points us pertain, principally, to issues of comprehensiveness in several dimensions. They invite us to reflect upon larger contexts; to be sensitive to all the various forces in play, and to all the complex interactions among them; to interrogate thoroughly our own values, and to assess carefully the way in which all these interactions might impact upon whatever it is we value and disvalue in social outcomes.

Indeed, something very much like that seems to be what references to “design” considerations point to across this whole range of social-scientific design literature.¹⁰² The reference seems always essentially to be to a notion of “goodness of fit” between the designed object (policy,

⁹⁷Samuelson 1954.

⁹⁸Of, e.g., “approval voting” as a solution to the former problem (Brams and Fishburn 1978) and the “Clarke tax” as a solution for the latter (Tideman 1977).

⁹⁹Simon 1969/1981.

¹⁰⁰Hitch 1958. Hitch and McKean 1960. Leontieff 1982. Little and Mirrlees 1974.

¹⁰¹Rivlin 1971.

¹⁰²And beyond: see particularly Alexander 1964.

mechanism, system) and the larger context in which it is set. In the case of a policy, a well-designed policy is one which fits well with the other policies, and the larger political/economic/social systems in which it is set. In the case of a mechanism, a well-designed mechanism is one that works well alongside other features of the social environment in which it is set, including other mechanisms in play there. Insofar as the mechanism works by manipulating incentives to individuals, a well-designed mechanism is one whose internal requirements are “compatible” with other incentives that individuals face, rather than offering people incentives for undercutting the goals (characterized as Pareto-optimality, or whatever) which we were trying to achieve by using the mechanism in the first place. In the case of a whole system, being well designed means that all the pieces fit together well in a harmonious whole: being well integrated, being in equilibrium (and perhaps robustly so, whether homostatically or otherwise).¹⁰³

1.3.4 The Normative/Empirical Interface

Such theories of optimal design – whether of policy, mechanism, system, or institution – are at the same time both empirical and normative. That basic fact is often acknowledged. The details of that connection are rarely explored adequately, however. We are typically left with the general impression that there is and should be some necessary, direct connection between the empirical and normative sides of the design project. But we are rarely given even the sketchiest indication of what makes the connection a necessary, much less a direct, one.¹⁰⁴

One way the connection might work is this. Let us suppose that those theories of optimal design are essentially normative in their fundamental motivation. They amount to a quest for some ideal state of the world. They are theories about what a good (indeed, perfect: optimal) arrangement would be. But then comes the thought that the ideal may be a pretty good guide to the real. After all (or so the thought goes) it is only to be expected, in the descriptive sense as well as the prescriptive one, that the ideal should be realized. What requires explanation is not doing the right thing, but rather departures from that ideal.

¹⁰³On these matters see especially Lockwood’s 1964 classic paper, “Social integration and system integration.”

¹⁰⁴That is true even of Arrow and Hahn’s 1971 proof of the fundamental theorems of welfare economics: it explains that idealized markets are Pareto-optimal, and thus gives us normative reasons of a kind for desiring them; but absent some other premises, that Pareto-optimality of ideal markets does not in itself explain how markets might empirically have come about.

Greater or lesser departures are common enough, of course, and explanations aplenty are required (and are indeed proffered) by analysts pursuing this basic line of attack. But what they explain is notable: not good outcomes, but bad; not perfectly attaining our goals, but failures to do so. The point is not just that, from certain other perspectives, the explanatory priority (what is “only to be expected” and what demands an explanation) may well be otherwise. That is true too, of course. But, more importantly, we should note the curiosity of taking anything at all for granted in this regard.

How we manage to achieve optimal arrangements requires explanation every bit as much as does our failure to achieve them. It may well be that optimality is self-enforcing, in some way or another. But the ways in which that is true need to be set out specifically. We need to know the explanatory mechanisms at work. They might be intentional rewards and punishments meted out by an angry God or all-powerful Lawgiver; they might be hidden-hand mechanisms guiding us toward a competitive equilibrium. But we need a story. Logically, it is as unsatisfactory to take optimizing for granted as it is to take failures to optimize for granted (saying, perhaps, “it’s only to be expected from politics” – or, equivalently for a certain stripe of economist, “rent-seeking is only to be expected in most social affairs”).

A second basic way of explaining some connection between the prescriptive and descriptive, the normative and the empirical, in optimal design theories is just this. There is no reason to suppose that agents in the real world will intentionally pursue optimal designs – or even that they will recognize them when they see them (or even that they will ever stumble across them, given how rococco some of those optimal designs are). The prescriptions have explanatory force, not because people internalize and intentionally act upon those prescriptions. Rather, those prescriptions are a good guide to what we will actually find in the real world because the same thing that makes us prescribe them as optimal designs (the “goodness of their fit to the larger context”) also makes them well suited to survive in their larger environment.

The reason that optimal designs predominate, on this account, has nothing to do with the frequency with which they are chosen and everything to do with the relative frequency of relatively persistent institutions. Longevity implies frequency, other things being equal. The best bet, at any moment in time, is that most institutions will be the ones that have been around for a while; or equivalently, the best bet a propos any given institution is that it is one which has been around for a while. Institutions which have been around for a while are most likely to be ones that are more nearly optimally designed to fit their environ-

ment (or, in older terminology, ones which are “functionally well adapted”). So normative criteria of optimal design can also ground predictions of empirical frequency.¹⁰⁵

Even that more modest connection between descriptive and prescriptive uses of design theory cannot be sustained without further presuppositions, however. The crucial presupposition, for present purposes, is that there is some mechanism which weeds out ill-fitting institutions over time, either refining and replacing them with ones that fit the environment better or else just disposing of them altogether.¹⁰⁶ There must, in other words, be some selection procedure at work to underwrite the assumption that over time the accumulation of institutions will tend to favor good-fitting ones over ill-fitting ones.

This seems to be a tough sort of claim to sustain. Government organizations, at least, display enormous longevity and persist well after their original reasons for existing have passed away.¹⁰⁷ Insofar as other social institutions are like that, then it seems implausible to postulate any tough competitive environment that weeds out ill-fitting institutions on anything like a systematic basis.¹⁰⁸

In the end, the best analysis of any necessary connection between descriptive and prescriptive aspects of optimal design theories lodges it squarely in the intentions of social agents. What theories of optimal design try to do is to give social agents good reasons for shaping institutions in some ways rather than others. Insofar as they are convinced of those arguments and moved by those reasons, those social agents will try to act upon those design prescriptions. Insofar as they succeed,

¹⁰⁵“Ground predictions” is an intentional fudge, stopping well short of stronger claims that they “provide explanations”; reasons for this wariness are found in Elster (1983, chap. 2), critiquing early functionalists (notably Merton 1957) and even sophisticated later ones (e.g., Cohen 1978, chap. 9).

¹⁰⁶Another presupposition, equally crucial to that larger explanatory scheme but less interesting for present purposes, concerns the rates at which new institutions are generated and at which old ill-fitting ones are killed off. If new institutions (some well-fitting, many not) arise at a great rate, and especially if the selection mechanisms work relatively slowly to weed out the ill-fitting among them, then most institutions at any given time might well be ill-fitting newer ones upon which the selection mechanism has not yet had a chance to work.

¹⁰⁷Kaufman 1976.

¹⁰⁸Or at least it does so absent some special explanation of the persistence over time of institutions, in this way. Such special explanations might be available: the reason for institutions persisting may be that they have come to acquire new functions to replace their old, and in that way to fit the larger environment well despite their changed focus; or it may be that they have shaped the environment to fit themselves. Some such claim is what must be sustained to save this larger style of explanation of the connection between empirical and normative aspect of optimal design theory.

institutions shaped by their actions will end up bearing something of the mark of those theories of optimal design.

Thus, the connection is there. But the connection comes through effects of the prescriptions on the intentions of agents, and through the effects of those agents' intentions on the social world. To say that is to claim (or ask) a lot. But any more facile claim – that optimal design theories are unreflectively internalized or automatically enforced through competition in a hostile environment – seems far less tenable. It seems far better to admit forthrightly that the point of moralizing (which is after all what we are doing in prescribing optimal social arrangements) is to shape people's values and preferences and, through them, their actions.

1.4 New Institutions for Old, Good Institutions for Bad

In this final section, I turn to larger questions of what sorts of design principles might usefully guide the shaping and reshaping of social institutions. The first issue to address (in Section 1.4.1) is in what sense design criteria track morality: to what extent, or in what sense, "good design" is actually "good" at all. I close (in Section 1.4.2) with a sampler of some design principles that might actually be commended both externally and internally.

1.4.1 Design Criteria and Moral Desiderata

The fundamental notion of design, as explicated in Section 1.3.3, relates to "goodness of fit" of the designed object to its environment. That definition, in turn, provides an obvious "internal" criterion of what it is for a design to be a "good design." A well-designed object is one that fits its environment well. A well-designed institution, in particular, would be one that is both internally consistent and externally in harmony with the rest of the social order in which it is set.

That internal criterion of good design built into the notion of design itself is misleadingly obvious, though. The larger question remains unasked. How well does that "internal" criterion fit with external standards of moral worth? From a larger moral point of view, it is an open question whether goodness of fit is necessarily good at all.

It is a familiar point in moral philosophy that any given individual can have both internal and external reasons for action, and that those two sets of reasons for action might well point in very different direc-

tions.¹⁰⁹ Internal reasons, in the individual's case, are reasons derived from that agent's own motivational structure: reasons internal to his existing beliefs and desires, principles and prejudices. External reasons, in the individual's case, are reasons derived from some larger moral truths unconnected with the individual's own extant beliefs and desires: truths about what is right and good and worthy. The latter sorts of reasons offer an external stance for the critique – for reevaluation and reformulation – of internal ones.

The same is true in the case of institutional designs. The internal criterion of good design, central though it is to the very notion of design itself, must be supplemented and held up to judgment against larger external evaluative criteria. From a collective and institutional point of view as well as from an individual moral point of view, there can be external as well as internal reasons for action. There can be good reasons for seeking institutions that fit ill, not well, with the rest of their environment.

The most dramatic instance, of course, is an institution set in the context of an evil social order which morally we ought to destroy. Undoubtedly design theories could be applied to concoct an optimal mechanism for overseeing the labor of slaves. But optimality in that context – goodness of fit to *that* environment – constitutes criticism rather than praise from any credible external point of view.¹¹⁰

Less dramatic cases are in many ways even more interesting. A natural reading of the "goodness of fit" criterion of good design equates it with "harmony," with "promoting the smooth functioning" of the designed object and the larger system within which it is set. But sometimes disharmony is far from disadvantageous. In designing mechanisms for group decision-making, we are often well advised to designate someone formally to serve as "devil's advocate," challenging our shared presumptions and telling us things we do not want to hear, as a way of improving the quality of the overall decision that we reach.¹¹¹ We are often well advised to design institutions so as to encourage disharmony and hence dynamics, to force us to reconsider and perhaps to change

¹⁰⁹Williams 1981, chap. 8.

¹¹⁰Of course, if the optimal design enjoins decent treatment of slaves (to protect the slaveowner's capital investment, or to increase morale and hence output) then we might coincidentally concur in the recommendation, at least as an interim step. And if the preferences of slaves themselves were included among criteria to be considered in assessing the "optimality" of such schemes, we might have a further reason for respecting the recommendations. But even then, we might be perfectly prepared to recommend utterly non-Paretian moves in that context – we may well suppose that, in trying to improve the plight of slaves, there is no reason at all to limit ourselves to schemes that would not reduce the welfare of slaveowners.

¹¹¹George 1972.

the way we are doing things from time to time. (Competition, both in economics and politics, is commended on precisely those grounds.) We are often well advised to design institutions in such a way that they allow us to “take one step backward so we may take two steps forward.”¹¹²

Perhaps we might be tempted to say, in all those less dramatic cases, that what optimal designs actually prescribe is the deliberate creation of institutional irritants. In a way that is certainly true. But in so doing these prescriptions appeal to a notion of “optimality” that transcends any narrow reference to the internal criteria of design. The appeal of these larger notions of optimal design is not to smooth functioning, well-ordered internal relations, goodness of fit to the existing local environment, or whatever. The appeal is instead to some notions of “optimal design” that point to the goodness of fit of the institution to some larger objectives than those narrow ones embodied in the internal goals of the institution and its immediate environment.

The appeal is, ultimately, to some larger moral code. There is a sense in which an institution might be said to be well designed if it is internally consistent and externally harmonious with its larger social environment. But that is still an essentially internalist definition of optimal design which must eventually give way to larger external critiques, rooted in normative principles that are at the end of the day themselves independently defensible.

1.4.2 Some Desirable Principles of Institutional Design

What sorts of principles, then, should guide our institutional designs? Clearly, from what has just been said, they should be principles with deeper moral resonance. Good institutional design is not just a matter of pragmatics. It is not just a matter of aesthetic or functional “goodness of fit.” Equally clearly, though, we ought not have to suspend our quest for better social institutions until we have reached agreement on all the deeper truths of ethics and metaphysics.

In discussions of institutional design what we often can do, and all we usually ought be trying to do, is seeking principles that trade on “theories of the middle range” in both empirical and normative realms.¹¹³ We can hope to discover, and to embody institutionally, a raft of generalizations of reliable validity, at least within a certain (perhaps tightly circumscribed) sphere.

In the remainder of this section, I shall be sketching a few illustrative

¹¹²Elster 1979, chap. 1.

¹¹³Merton 1957.

principles of this sort. None are worked out fully: that is the task for other chapters, and indeed other books. None can be applied universally or commended without qualification: all have a two-edged aspect to them, somehow. Still, it is useful in closing this introduction to offer examples of the sorts of principles that we might be aspiring to adduce, even if this is of necessity a very preliminary cut at that project.

One middle-range social generalization that institutional designers might want somehow to embody in their principles, for example, concerns the twin and connected facts that humans are fallible and that societies change. If we are likely to err about matters either of fact or of value – or if the facts or values upon which our actions are predicated are themselves to some extent specific to situations which might well change – then it is a mistake to set those possible errors in concrete. It is far better to design our institutions in such a way as to be flexible in these regards, to admit of “learning by doing” and to evolve over time.¹¹⁴ Thus, we might say *revisability* is one important principle of institutional design.

Of course, there is another side to that story, too. We want to have the capacity, sometimes, to bind ourselves to a certain course of action and to ensure that we (or our successors) resist any temptations to deviate from it. Personally, we want to make commitments and stick to them: we want to keep our promises, honor our contracts, respect others’ trust and confidences. Politically, we want to make commitments without renegeing on them: we want to keep our political promises, to honor our treaties, and so on.¹¹⁵ Revisability is an important principle of institutional design, then, but it is also one which must somehow be kept within limits.

Just as we want our institutions to be capable of changing in response to relevant changes in the factual or evaluative universe, so too do we want them to be responsive only to relevant changes there. While we certainly do want our institutions to be open to alteration where appropriate, by the same token we want them to be resistant to sheer buffeting by changes in social circumstances that have no bearing upon the assumptions upon which those institutions were predicated.

That points to a further principle of institutional design which might best be termed *robustness*. They should be capable of adapting to new

¹¹⁴Wildavsky 1979. March and Olsen 1984, pp. 745–77.

¹¹⁵In public finance, this appears as the “capital levy” problem: narrowly calculating governments have an interest in luring investors into a jurisdiction with a promise of low tax rates fixed for a specific period, but having induced relatively immobile capital to relocate there then to raise the taxes before the specified period is up. But the same governments, of course, also have an interest in somehow being able to make credible commitments that they will not do any such thing.

situations: not brittle and easily destroyed by them. But they should adapt to new situations only in ways that are appropriate to the relevant respects in which the situations are new – changing fundamentally only where there has been some fundamental change in the factual or evaluative universe, and making only surface accommodations to changed circumstance where there has not.

Of course, what counts as “appropriate” and “relevant” here is again going to be a matter of contention and political controversy. Simple reference to the internal logic of the institutions themselves is insufficient to decide the matter for the same reason that internal standards of “good design” are themselves unsatisfactory. So, again, robustness is a principle of institutional design that is valuable, but only within limits the scope of which cannot in principle be specified.

Another empirical generalization which institutions must respect is the admixture of motives that moves most people, at least in most societies relevantly similar to our own. Within most social actors, self-seeking impulses exist alongside principled and even altruistic motives. In designing our institutions we ought at least take account of that fact, which might (as a further principle) be described as *sensitivity to motivational complexity*.

How exactly institutions might best accommodate that fact of motivational complexity is an open question, turning upon premises that are in part empirical and in part normative. Classic models of separation of powers – checks and balances between branches of government, grants of rights to individuals against the government, and pluralist institutions to ensure centers of countervailing power across society more generally – constitute one style of reaction.¹¹⁶ But by “designing institutions for knaves” such mechanical solutions risk making knaves of potentially more honorable actors. Depending upon exactly what structure we think people’s moral codes take, and exactly what opportunity structure they face, a more trusting model embodying a more direct appeal to moral principles might actually do a better job of evoking high-minded motives for action and of suppressing low-minded ones.¹¹⁷

One way to do that might, for example, be through explicit appeal to a Kantian-inspired *publicity* principle, requiring as a test of all institutions and institutional actions that they be (at least in principle) pub-

¹¹⁶Grofman and Wittman 1989. Moe 1990. Elster 1993. In the same spirit, Schultze (1977) suggests structuring incentives in more policy-specific spheres so as to make public and private interest coincide.

¹¹⁷Goodin 1982, chap. 6; 1992. Brennan and Lomasky’s 1993 account of voting in a large electorate is like that, as is Ackerman’s (1991; cf. Goodin 1992, chap. 6) account of the “constitutional moment.”

licly defensible.¹¹⁸ The thought underlying this proposal in part turns upon hypotheses in moral psychology: suppositions that people recognize themselves as having higher and lower motives, that they would be ashamed to admit openly to the latter, and that if all action has to be defended publicly only the higher motives would be appealed to as reasons for action.¹¹⁹ A more frankly political thought paralleling this one is simply this: insofar as we are trying to scare up political support for a project, insofar as we have to do so through open appeals to people's reason, we had better give them reasons that they themselves can embrace; it would clearly be counterproductive, in such situations, to appeal merely to our own narrowly selfish reasons for advocating a project, which of course are reasons that others cannot be expected to share.

But again, of course, publicity protects our public deliberations against only a certain range of contaminating factors. It may block the play of crass self-interest, but it will not do much to block the principled sacrifice of some possibly large segments of the community to the common good or to some "higher" cause. And whether our institutions should be designed with knaves or with potential angels primarily in view depends crucially upon our views as to the relative frequency of each in the population, and as to the relative damage that will be done (or good that will be missed) by making one assumption rather than another given those frequencies. This is a matter of fact, in some sense; but it is in practice usually an empirically undecidable question (we cannot, or cannot afford to, undertake the crucial experiments). Thus, again, whether the publicity principle or the principle of designing institutions for knaves is a good premise for institutional design depends fundamentally upon deeply contentious issues that admit only, in the final analysis, of a political resolution.

Finally, insofar as we are counting on trial-and-error, learning-by-doing processes to perfect our institutional arrangements, we ought embrace as a central principle of design a desire for *variability* in our institutional arrangements. We ought encourage experimentation with different structures in different places; and we ought, furthermore, encourage reflection upon the lessons from elsewhere and a willingness to borrow those lessons where appropriate. Federalism is sometimes defended on precisely this ground, as a "social laboratory" in which different approaches are allowed to emerge in different jurisdictions.¹²⁰

¹¹⁸This is the subject of Chapter 6; see also Dryzek 1990 and, more skeptically, Goodin 1992, chap. 7.

¹¹⁹This is J. S. Mill's argument in *Representative Government* for open voting and against the secret ballot (see Goodin 1992, chap. 7, for a discussion).

¹²⁰Wildavsky 1979, chap. 6.

There are difficulties with all such schemes, though. One is that precisely the sort of “political protectionism” that is required to prevent premature homogenization of arrangements across jurisdictions also gets in the way of borrowing across jurisdictions, once better arrangements have clearly emerged elsewhere. Another more serious danger is that, instead of serving as a “social laboratory” in which other jurisdictions will borrow the best from elsewhere, there will instead be a “race to the bottom.” In a whole raft of policy areas, from tax to regulatory policy, we often see the worst practice rather than the best being adopted in neighboring jurisdictions. Whether federal institutions, or other variance-maximizing principles of institutional design, are good ideas thus depends once again upon a fundamentally political judgment as to which is the most likely consequence.

Thus, none of these principles of institutional design are unqualified or sacrosanct. Many more, equally qualified in character, will emerge in the course of subsequent chapters. This short initial list is intended merely as a sampler. These are just a few examples of the sorts of middle-range theories of both empirical and normative import to which we might appeal in designing and redesigning social institutions.

The enterprise is an important one, though. Qualified though our generalizations may be, our institutions are even more influential. If we are to understand how social life works, and how it might work better, fixing our focus firmly upon institutions and their reshaping is one crucial step.

References

- Ackerman, Bruce. 1991. *We, the People*. Cambridge, Mass.: Harvard University Press.
- Alexander, Christopher. 1964. *Notes on the Synthesis of Form*. Cambridge, Mass.: Harvard University Press.
- Alexander, Ernest R. 1982. Design in the decision-making process. *Policy Sciences*, 14, 279–92.
- Almond, Gabriel A., Scott C. Flanagan, and Robert J. Mundt, eds. 1973. *Crisis, Choice, and Change: Historical Studies of Political Development*. Boston: Little, Brown.
- Arrow, Kenneth J. 1951. *Social Choice and Individual Values*. New York: Wiley.
- Arrow, Kenneth J. 1974. *The Limits of Organization*. New York: Norton.
- Arrow, Kenneth J., and Frank Hahn. 1971. *General Competitive Analysis*. San Francisco: Holden-Day.
- Arrow, Kenneth J., and Leonid Hurwicz, eds. 1977. *Studies in Resource Allocation Processes*. Cambridge: Cambridge University Press.

- Arthur, W. B. 1989. Competing technologies and lock-in by historical events. *Economic Journal*, 99, 116–31.
- Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.
- Axelrod, Robert. 1986. An evolutionary approach to norms. *American Political Science Review*, 80, 1095–1112.
- Barrett, Michele, and Mary McIntosh. 1982. *The Anti-social Family*. London: Verso.
- Barry, Brian. 1970. *Socialists, Economists and Democracy*. London: Collier-Macmillan.
- Bates, Robert H. 1988. Contra contractarianism: some reflections on the new institutionalism. *Politics & Society*, 16, 187–401.
- Becker, Gary S. 1976. *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press.
- Becker, Gary S. 1981. *A Treatise on the Family*. Cambridge, Mass: Harvard University Press.
- Benton, Ted. 1984. *The Rise and Fall of Structural Marxism*. London: Macmillan.
- Berger, Brigitte, and Peter L. Berger. 1983. *The War over the Family*. Harmondsworth, England: Penguin.
- Berger, Peter L., and Hansfried Kellner. 1965. Arnold Gehlen and the theory of institutions. *Social Research*, 32, 110–5.
- Berger, Peter L., and Thomas Luckman. 1966. *The Social Construction of Reality*. Garden City, N.Y.: Doubleday.
- Berman, Harold J. 1983. *Law and Revolution: The Formation of the Western Legal Tradition*. Cambridge, Mass.: Harvard University Press.
- Binder, Leonard, et al., eds. 1971. *Crises and Sequences in Political Development*. Princeton, N.J.: Princeton University Press.
- Bobrow, Davis B., and John S. Dryzek. 1987. *Policy Analysis by Design*. Pittsburgh: University of Pittsburgh Press.
- Bovens, Mark A. P. 1990. The social steering of complex organizations. *British Journal of Political Science*, 20, 91–117.
- Brams, Steven J., and Peter C. Fishburn. 1978. Approval voting. *American Political Science Review*, 72, 831–47.
- Brennan, Geoffrey, and James M. Buchanan. 1985. *The Reason of Rules: Constitutional Political Economy*. Cambridge: Cambridge University Press.
- Brennan, Geoffrey, and Alan Hamlin 1992. Bicameralism and majoritarian equilibrium. *Public Choice*, 74, 169–79.
- Brennan, Geoffrey, and Loren Lomasky. 1993. *Democracy and Decision: The Pure Theory of Electoral Preference*. Cambridge: Cambridge University Press.
- Bryce, James. 1888. *The American Commonwealth*. London: Macmillan.
- Buchanan, James M., Robert D. Tollison, and Gordon Tullock, eds. 1980. *Toward a Theory of the Rent-seeking Society*. College Station: Texas A&M University Press.

- Calvert, Randall, Matthew D. McCubbins, and Barry R. Weingast. 1989. A theory of political control and agency discretion. *American Journal of Political Science*, 33, 588–61
- Coase, R. H. 1937. The nature of the firm. *Economica*, 4, 386–405.
- Cohen, G. A. 1978. *Karl Marx's Theory of History*. Oxford: Clarendon Press.
- Cohen, Jean L., and Andrew Arato. 1992. *Civil Society and Political Theory*. Cambridge, Mass.: MIT Press.
- Cohen, Joshua. 1982. Book review of G. A. Cohen, *Karl Marx's Theory of History*. *Journal of Philosophy*, 79, 253–73
- Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, Mass.: Harvard University Press.
- Coleman, James S. 1991. Constructed organization: first principles. *Journal of Law, Economics & Organization*, 7S (Special issue), S7–S23.
- Coleman, James S. 1993. The rational reconstruction of society. *American Sociological Review*, 58, 1–15.
- Commons, John R. 1931. Institutional economics. *American Economic Review*, 21, 648–57. Reprinted in Samuels 1988, vol. 1, pp. 18–27.
- Commons, John R. 1934. *Institutional Economics*. New York: Macmillan.
- Dahl, Robert A. 1963. *Modern Political Analysis*. Englewood Cliffs, N.J.: Prentice-Hall.
- Dahrendorf, Ralf. 1958. Out of utopia: toward a reorientation of sociological analysis. *American Journal of Sociology*, 64, 115–27.
- Dryzek, John S. 1990. *Discursive Democracy*. Cambridge: Cambridge University Press.
- Dunleavy, Patrick. 1991. *Democracy, Bureaucracy and Public Choice*. Hemel Hempstead: Harvester-Wheatsheaf.
- Eisenstadt, Shmuel N. 1964. Institutionalization and change. *American Sociological Review*, 29, 235–47.
- Eisenstadt, Shmuel N. 1968. Social institutions: the concept. Vol. 14, pp. 409–21 in David L. Sills, ed., *International Encyclopedia of the Social Sciences*. New York: Macmillan.
- Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes*. Cambridge, Mass.: Harvard University Press.
- Elster, Jon. 1979. *Ulysses and the Sirens*. Cambridge: Cambridge University Press.
- Elster, Jon. 1983. *Explaining Technical Change*. Cambridge: Cambridge University Press.
- Elster, Jon. 1985a. *Making Sense of Marx*. Cambridge: Cambridge University Press.
- Elster, Jon, ed. 1985b. *The Multiple Self*. Cambridge: Cambridge University Press.
- Elster, Jon. 1989. *The Cement of Society*. Cambridge: Cambridge University Press.

- Elster, Jon. 1993. Majority rule and individual rights. Pp. 175–216 in Stephen Shute and Susan Hurley, eds., *On Human Rights*. New York: Basic Books.
- Elster, Jon, Claus Offe, and Ulrich Preuss. Forthcoming. *Constitutional Politics and Economic Transformation in Post-communist Societies*. Cambridge: Cambridge University Press.
- Elsthain, Jean Bethke. 1981. *Public Man, Private Woman*. Princeton, N.J.: Princeton University Press.
- Esping-Andersen, Gösta. 1990. *The Three Worlds of Welfare Capitalism*. Oxford: Polity.
- Evans, Peter R., Dietrich R. Rueschemeyer, and Theda Skocpol, eds. 1979. *Bringing the State Back In*. Cambridge: Cambridge University Press.
- French, Peter A. 1979. The corporation as a moral person. *American Philosophical Quarterly*, 16, 207–15.
- Garrett, Geoffrey. 1992. International cooperation and institutional choice: the European Community's internal market. *International Organization*, 46, 532–60.
- George, Alexander. 1972. The case for multiple advocacy in making foreign policy. *American Political Science Review*, 66, 751–85.
- Gibbard, Allan. 1973. Manipulation of voting schemes: a general result. *Econometrica*, 41, 587–601.
- Giddens, Anthony. 1984. *The Constitution of Society*. Oxford: Polity.
- Gintis, Herbert. 1972. A radical analysis of welfare economics and individual development. *Quarterly Journal of Economics*, 68, 572–99.
- Goffman, Erving. 1970. *Strategic Interaction*. Oxford: Blackwell.
- Goodin, Robert E. 1982. *Political Theory and Public Policy*. Chicago: University of Chicago Press.
- Goodin, Robert E. 1992. *Motivating Political Morality*. Oxford: Blackwell.
- Goodsell, Charles T. 1992. The public administrator as artisan. *Public Administration Review*, 52, 246–53.
- Granovetter, Mark. 1985. Economic action and social structure: the problem of embeddedness. *American Journal of Sociology*, 91, 481–510.
- Granovetter, Mark. 1992. Economic institutions as social constructions: a framework for analysis. *Acta Sociologica*, 35, 3–11.
- Grofman, Bernard, and Donald Wittman, eds. 1989. *The Federalist Papers and the New Institutionalism*. New York: Agathon Press.
- Hardin, Russell. 1982a. *Collective Action*. Baltimore: Johns Hopkins University Press.
- Hardin, Russell. 1982b. Exchange theory on strategic bases. *Social Science Information*, 21, 251–72.
- Hardin, Russell. 1989. Why a constitution? Pp. 100–20 in Bernard Grofman and Donald Wittman, eds., *The Federalist Papers and the New Institutionalism*. New York: Agathon Press.
- Hardin, Russell. 1995. *One for All*. Princeton, N.J.: Princeton University Press.

- Hart, H. L. A. 1961. *The Concept of Law*. Oxford: Clarendon Press.
- Hayek, Friedrich A. 1973. *Law, Legislation and Liberty*. London: Routledge & Kegan Paul.
- Hindess, Barry. 1988. *Choice, Rationality and Social Theory*. London: Unwin-Hyman.
- Hindess, Barry, and Paul Hirst. 1977. *Mode of Production and Social Formation*. London: Macmillan.
- Hirst, Paul. 1994. *Associative Democracy*. Oxford: Polity.
- Hitch, Charles J. 1958. Economics and military operations research. *Review of Economics & Statistics*, 40, 119–209.
- Hitch, Charles J., and Roland N. McKean. 1960. *The Economics of Defense in the Nuclear Age*. Cambridge, Mass.: Harvard University Press.
- Homans, George C. 1964. Bringing men back in. *American Sociological Review*, 29, 809–18.
- Huntington, Samuel P. 1968. *Political Order in Changing Societies*. New Haven, Conn.: Yale University Press.
- Hurwicz, Leonid. 1977. The design of resource allocation mechanisms. Pp. 3–37 in Kenneth J. Arrow and Leonid Hurwicz, eds., *Studies in Resource Allocation Processes*. Cambridge: Cambridge University Press.
- Hurwicz, Leonid, David Schmeidler, and Hugo Sonnenschein, eds. 1985. *Social Goals and Social Organization*. Cambridge: Cambridge University Press.
- Ingram, Helen, and Anne Schneider. 1988. Improving implementation through policy design: framing smarter statutes. Paper presented to the American Political Science Association annual meetings, Washington, D.C.
- Jackman, Robert. 1972. *Politics and Social Equality*. New York: Wiley.
- Kaufman, Herbert. 1976. *Are Government Organizations Immortal?* Washington, D.C.: Brookings Institution.
- Kaufmann, Franz-Xavier, Giandomenico Majone, and Vincent Ostrom, eds. 1985. *Guidance, Control and Evaluation in the Public Sector*. Berlin: W. de Gruyter.
- Keohane, Robert O. 1984. *After Hegemony*. Princeton, N.J.: Princeton University Press.
- Keohane, Robert O., and Joseph Nye. 1977. *Power and Interdependence*. Boston: Little, Brown.
- Keohane, Robert O., ed. 1986. *Neorealism and Its Critics*. New York: Columbia University Press.
- Key, V. O. 1942. *Politics, Parties and Pressure Groups*. New York: Crowell.
- Knight, Jack. 1992. *Institutions and Social Conflict*. Cambridge: Cambridge University Press.
- Krasner, Stephen D. 1984. Approaches to the state: alternative conceptions and historical dynamics. *Comparative Politics*, 16, 223–46.
- Kymlicka, Will. 1989. *Liberalism, Community and Culture*. Oxford: Clarendon Press.

- Leontief, Wassily. 1986. *Input–Output Economics*. 2nd ed. New York: Oxford University Press.
- Linder, Stephen H., and B. Guy Peters. 1985. From social theory to policy design. *Journal of Public Policy*, 4, 237–59.
- Linder, Stephen H., and B. Guy Peters. 1987. A design perspective on policy implementation: the fallacies of misplaced prescription. *Policy Studies Review*, 6, 459–76.
- Little, I. M. D., and J. A. Mirrlees. 1974. *Project Planning and Appraisal for Developing Countries*. London: Heinemann.
- Lockwood, David. 1964. Social integration and system integration. Pp. 244–57 in George K. Zollschan and Walter Hirsch, eds., *Explorations in Social Change*. London: Routledge.
- Luce, R. Duncan, and Howard Raiffa. 1957. *Games and Decisions*. New York: Wiley.
- Lukes, Steven. 1973. *Individualism*. Oxford: Blackwell.
- Lukes, Steven. 1974. *Power: A Radical View*. London: Macmillan.
- Manis, Jerome G., and Bernard N. Meltzer, eds. 1967. *Symbolic Interaction*. Boston: Allyn and Bacon.
- Mansbridge, Jane J., ed. 1990. *Beyond Self-interest*. Chicago: University of Chicago Press.
- March, James G., and Johan P. Olsen. 1984. The new institutionalism: organizational factors in political life. *American Political Science Review*, 78, 734–49.
- March, James G., and Johan P. Olsen. 1989. *Rediscovering Institutions*. New York: Free Press.
- Marshall, T. H. 1949. Citizenship and social class. Reprinted pp. 70–134 in Marshall, *Class, Citizenship and Social Development*. Chicago: University of Chicago Press, 1963.
- McAfee, R. Preston, and John McMillan. 1987. Auctions and bidding. *Journal of Economic Literature*, 25, 699–738.
- McCubbins, Matthew D. 1985. The legislative design of regulatory structure. *American Journal of Political Science*, 29, 721–48.
- McKelvey, Richard D. 1979. General conditions for global intransitivities in formal voting models. *Econometrica*, 47, 1085–1112.
- Merton, Robert K. 1936. The unintended consequences of purposive social action. *American Sociological Review*, 1, 894–904.
- Merton, Robert K. 1957. *Social Theory and Social Structure*. Glencoe, Ill.: Free Press.
- Milgrom, Paul R., Douglass C. North, and Barry R. Weingast. 1990. The role of institutions in the revival of trade: the law merchant, private judges and the Champagne fairs. *Economics & Politics*, 2, 1–23.
- Moe, Terry M. 1984. The new economics of organization. *American Journal of Political Science*, 28, 739–77.
- Moe, Terry M. 1990. Political institutions: the neglected side of the story.

- Journal of Law, Economics & Organization*, 6S (Supplement), S213–54.
- Moe, Terry M. 1991. Politics and the theory of organization. *Journal of Law, Economics & Organization*, 7S (Supplement), S106–29.
- Moravcsik, Andrew. 1991. Negotiating the Single European Act: national interests and conventional statecraft in the European Community. *International Organization*, 45, 19–56.
- Nagel, Thomas. 1986. *The View from Nowhere*. New York: Oxford University Press.
- Nelson, Richard R., and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, Mass: Harvard University Press.
- Niskanen, William A. 1971. *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton.
- Nordlinger, Eric A. 1981. *On the Autonomy of the Democratic State*. Cambridge, Mass.: Harvard University Press.
- North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
- O'Donnell, Guillermo, Philippe C. Schmitter, and Laurence Whitehead, eds. 1986. *Transitions from Authoritarian Rule*. Baltimore: Johns Hopkins University Press.
- Oakeshott, Michael. 1962/1991. *Rationalism in Politics*. New edition. Indianapolis, Ind.: Liberty Press.
- Offe, Claus. 1991. Capitalism by democratic design? Democratic theory facing the triple transition in east central Europe. *Social Research*, 58, 865–92.
- Offe, Claus, and Ulrich Preuss. 1991. Democratic institutions and moral resources. Pp. 143–71 in David Held, ed., *Political Theory Today*. Oxford: Polity.
- Okin, Susan Moller. 1989. *Justice, Gender and the Family*. New York: Basic Books.
- Ordeshook, Peter C. 1986. *Games Theory and Political Theory*. Cambridge: Cambridge University Press.
- Ordeshook, Peter C. 1992. Constitutional stability. *Constitutional Political Economy*, 3, 137–75.
- Orren, Karen. 1991. *Belated Feudalism: Labor, the Law and Liberal Development in the United States*. Cambridge: Cambridge University Press.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Parsons, Talcott. 1937. *The Structure of Social Action*. New York: McGraw-Hill.
- Parsons, Talcott. 1952. *The Social System*. London: Tavistock.
- Parsons, Talcott, and Edward A. Shils, eds. 1951. *Toward a General Theory of Action*. Cambridge, Mass: Harvard University Press.
- Pateman, Carole. 1983. Feminist critiques of the public/private dichotomy. Reprinted pp. 118–40 in Pateman, *The Disorder of Women*. Oxford: Polity, 1989.

- Perrow, Charles. 1984. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books.
- Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, N.J.: Princeton University Press.
- Ranny, Austin, ed. 1962. *Essays on the Behavioral Study of Politics*. Urbana: University of Illinois Press.
- Rawls, John. 1993a. *Political Liberalism*. New York: Columbia University Press.
- Rawls, John. 1993b. The law of peoples. Pp. 41–82 in Stephen Shute and Susan Hurley, eds., *On Human Rights*. New York: Basic Books.
- Riker, William H. 1980. Implications from the disequilibrium of majority rule for the study of institutions. *American Political Science Review*, 74, 432–46.
- Riker, William H. 1983. *Liberalism against Populism*. San Francisco: W. Freeman.
- Rivlin, Alice M. 1971. *Systematic Thinking for Social Action*. Washington, D.C.: Brookings Institution.
- Roemer, John, ed. 1986. *Analytical Marxism*. Cambridge: Cambridge University Press.
- Rorty, Richard. 1989. *Contingency, Irony and Solidarity*. Cambridge: Cambridge University Press.
- Ross, Edward Alsworth. 1901. *Social Control*. London: Macmillan.
- Ruggie, John G. 1982. International regimes, transactions and change: embedded liberalism in the post-war economic order. *International Organization*, 36, 379–415.
- Samuels, Warren J., ed. 1988. *Institutional Economics*. Aldershot: Elgar.
- Samuelson, Paul A. 1954. The pure theory of public expenditure. *Review of Economics Statistics*, 36, 387–89.
- Sandel, Michael J. 1982. *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press.
- Sandholtz, Wayne, and John Zysman. 1989. 1992: recasting the European bargain. *World Politics*, 42, 95–128.
- Satterthwaite, Mark A. 1975. Strategy-proofness and Arrow's conditions. *Journal of Economic Theory*, 10, 187–217.
- Satz, Debra, and John Ferejohn. 1994. Rational choice and social theory. *Journal of Philosophy*, 91, 71–87.
- Schattschneider, E. E. 1960. *The Semi-sovereign People*. New York: Holt, Rinehart and Winston.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.
- Schneider, Anne, and Helen Ingram. 1988. Systematically pinching ideas: a comparative approach to policy design. *Journal of Public Policy*, 8, 61–80.
- Schneider, Anne, and Helen Ingram. 1988a. Improving implementation

- through policy design: framing smarter statues. Paper presented to the American Political Science Association annual meetings, Washington, D.C.
- Schofield, Norman. 1976. Instability of simple dynamic games. *Review of Economic Studies*, 45, 575–94.
- Schultze, Charles L. 1977. *The Public Use of Private Interest*. Washington, D.C.: Brookings Institution.
- Sen, Amartya. 1977. Rational fools: a critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, 6, 317–44.
- Sen, Amartya. 1982. *Choice, Welfare and Measurement*. Oxford: Blackwell.
- Sen, Amartya. 1993. Internal consistency of choice. *Econometrica*, 61, 495–521.
- Shepsle, Kenneth A. 1979. Institutional arrangements and equilibrium in multidimensional voting models. *American Journal of Political Science*, 23, 27–59.
- Shepsle, Kenneth A., and Barry Weingast. 1981. Structure-induced equilibrium and legislative choice. *Public Choice*, 37, 503–19.
- Simon, Herbert A. 1957. *Administrative Behavior*. 2nd ed. New York: Free Press.
- Simon, Herbert A. 1969/1981. *The Sciences of the Artificial*. Cambridge, Mass.: MIT Press.
- Skinner, Quentin, ed. 1985. *The Return of Grand Theory in the Human Sciences*. Cambridge: Cambridge University Press.
- Skocpol, Theda. 1979. Bringing the state back in: strategies of analysis in current research. Pp. 3–37 in P. R. Evans, D. R. Rueschemeyer, and T. Skocpol, eds., *Bringing the State Back In*. New York: Cambridge University Press.
- Skocpol, Theda. 1992. *Protecting Soldiers and Mothers: The Political Origins of Social Policy in the United States*. Cambridge, Mass.: Harvard University Press.
- Smith, Rogers M. 1988. Political jurisprudence, the “new institutionalism” and the future of public law. *American Political Science Review*, 82, 89–108.
- Snidal, Duncan. 1985. Coordination versus Prisoner’s Dilemma: implications for international cooperation and regimes. *American Political Science Review* 79: 923–42.
- Soskice, David, Robert H. Bates, and David Epstein. 1992. Ambition and constraint: the stabilizing role of institutions. *Journal of Law, Economics and Organization*, 8, 547–60.
- Spencer, Herbert. 1879. *The Principles of Sociology*. London: Williams and Norgate.
- Steinmo, Sven, Kathleen Thelen, and Frank Longstreth, eds. 1992. *Structuring Politics: Historical Institutionalism in Comparative Analysis*. New York: Cambridge University Press.

- Stepan, Alfred, and Cindy Skach. 1993. Constitutional frameworks and democratic consolidation: parliamentarianism versus presidentialism. *World Politics*, 46, 1–22.
- Stigler, George J., and Gary S. Becker. 1977. De gustibus non est disputandum. *American Economic Review*, 67, 76–90.
- Streeck, Wolfgang. 1992. Beneficial constraints: on the economic limits of rational voluntarism. Pp. 5–34 in *Transformation Processes in Eastern Europe—Challenges for Socio-economic Theory*. Seminar papers, no. 16. Cracow: Cracow Academy of Economics.
- Taylor, Charles. 1989. *Sources of the Self*. Cambridge: Cambridge University Press.
- Taylor, Michael. 1987. *The Possibility of Cooperation*. Cambridge: Cambridge University Press.
- Tideman, T. Nicolaus, ed. 1977. Special supplement on the demand revealing process. *Public Choice*, vol. 29, no. 2 (supplement).
- Tilton, Timothy A. 1990. *The Political Theory of Swedish Social Democracy*. Oxford: Clarendon Press.
- Wagner, Peter. 1994. Survey article: dispute, uncertainty and institution in recent French debates. *Journal of Political Philosophy*, 2, 270–89.
- Weaver, R. Kent, and Bert A. Rockman, eds. 1993. *Do Institutions Matter? Government Capabilities in the United States and Abroad*. Washington, D.C.: Brookings Institution.
- Weingast, Barry R., and William Marshall. 1988. The industrial organization of Congress; or, why legislatures, like firms, are not organized as markets. *Journal of Political Economy*, 96, 132–63.
- Weintraub, E. Roy. 1979. *Microfoundations: The Compatibility of Microeconomics and Macroeconomics*. Cambridge: Cambridge University Press.
- Weizsäcker, C. C. von 1973. Notes on endogenous change of taste. *Journal of Economic Theory*, 3, 345–72.
- Wendt, Alexander E. 1987. The agent-structure problem in international relations theory. *International Organization*, 41, 335–70.
- Wildavsky, Aaron. 1979. *Speaking the Truth to Power: The Art and Craft of Policy Analysis*. Boston: Little, Brown.
- Williams, Bernard. 1981. *Moral Luck*. Cambridge: Cambridge University Press.
- Williamson, Oliver E. 1975. *Markets and Hierarchies*. New York: Free Press.
- Williamson, Oliver E. 1981. The modern corporation: origins, evolutions, attributes. *Journal of Economic Literature*, 19, 1537–68.
- Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.
- Williamson, Oliver E., and Sidney G. Winter, eds. 1991. *The Nature of the Firm: Origins, Evolution, and Development*. New York: Oxford University Press.

Wolf, Charles, Jr. 1990. *Markets or Governments*. Cambridge, Mass.: MIT Press.

Wrong, Dennis. 1961. The oversocialized conception of man in modern sociology. *American Sociological Review*, 26, 184–93.