Barriers to Conflict Resolution

EDITED BY

Kenneth J. Arrow
Stanford University

Robert H. Mnookin
Harvard University

Lee Ross
Stanford University

Amos Tversky
Stanford University

Robert B. Wilson
Stanford University

W. W. NORTON & COMPANY
New York • London

Printed in the United States of America
First Edition

The text of this book is composed in Caslon,
with the display set in Futura Bold.
Composition and manufacturing by The Maple-Vail Book Manufacturing Group.
Book design by Beth Tondreau Design / Robin Bentz

Library of Congress Cataloging-in-Publication Data

Barriers to conflict resolution / edited by Kenneth J. Arrow . . . [et al.].
p. cm.
"A project of the Stanford Center on Conflict and Negotiation (SCCN) and the end product of a conference held at Stanford University in February 1991"—Pref.
Includes bibliographical references and index.
1. Conflict management—Congresses. 2. Negotiation—Congresses.
I. Arrow, Kenneth Joseph, 1921– .
HD42.B295 1995
658.8'48—dc20 95-16260

ISBN 0-393-03737-1

W. W. Norton & Company, Inc., 500 Fifth Avenue, New York, N. Y. 10110
W. W. Norton & Company Ltd., 10 Coptic Street, London WC1A 1PU

1 2 3 4 5 6 7 8 9 0

CONTRIBUTORS

Kenneth J. Arrow
Departments of Economics and Operations Research, Stanford University

Max H. Bazerman
Kellogg Graduate School of Management, Northwestern University

Robyn M. Dawes
Department of Social and Decision Sciences, Carnegie Mellon University

John T. Dunlop
Lamont University Professor, Emeritus, Harvard University

Jon Elster
Department of Political Science, University of Chicago

Ronald J. Gilson
School of Law, Stanford University; School of Law, Columbia University

Daniel Kahneman
Woodrow Wilson School for Public and International Affairs, Princeton University

Robert H. Mnookin
School of Law, Harvard University

Margaret A. Neale
Kellogg Graduate School of Management, Northwestern University

John M. Orbell
Department of Political Science, University of Oregon

Wolfgang Panofsky
Stanford Linear Accelerator Center, Stanford University

Edward A. Parson
John F. Kennedy School of Government, Harvard University

Howard Raiffa
Schools of Business and Government, Harvard University

Lee Ross
Department of Psychology, Stanford University

Ariel Rubinstein
Department of Economics, Tel-Aviv University

James K. Sebenius
Harvard Graduate School of Business Administration, Harvard University

Lawrence Susskind
Department of Urban Studies and Planning, Massachusetts Institute of Technology

Amos Tversky
Department of Psychology, Stanford University

Robert B. Wilson
Graduate School of Business, Stanford University

Richard Zeckhauser
John F. Kennedy School of Government, Harvard University

CHAPTER ELEVEN

Cooperation in the Unbalanced Commons

Edward A. Parson and Richard J. Zeckhauser

In June 1992, leaders of 178 nations met in Rio de Janeiro to conclude negotiations on a broad set of measures to protect the international environment, including treaties on climate change and biodiversity. Such multiparty negotiations are pervasive in modern international affairs, and many, like the Rio negotiations, involve the provision of collective goods, those from which all benefit but which states provide individually. Other examples include controlling weapons proliferation, supporting liberal trade policies, financing international organizations, and deterring aggression. Problems of similar structure also arise frequently in other contexts, for example, among individuals or among states within a nation.

In this chapter we consider how multiple negotiating parties pursue and reach agreement on distributing obligations to provide collective goods. Our analysis of the process of arguing and bargaining that brings parties to accept certain obligations in return for others' accepting reciprocal obligations is principally conceptual, and applies to most multiparty bargains. Our illustrations focus on international bargaining over the environment.

We argue that study of multiparty negotiations is too often stuck in an assumption of symmetry. In fact, environmental and other multiparty international negotiations typically involve substantial *asymmetry of interest:* negotiators differ sharply in the trade-off each perceives between the benefit of the collective good being provided and their costs in providing it themselves. Though nations may share a common purpose, such as containing climate change, such asymmetries can obstruct agreements by putting the nations at odds over which negotiated solutions are desirable.

There are three conceptually separable elements to the process of reaching agreement on a multiparty public-good bargain: who participates, what simplifying principles are used to define relative obligations, and what particular levels of provision are chosen within these principles. Asymmetries of interest complicate each of these three stages, principally by obscuring or eliminating the clear primacy that the simple focal points of unanimous participation and identical measures enjoy in the symmetric case. Instead, there can be one or more stable coalitions of contributors short of the set of all affected parties. There can also be multiple competing principles for allocating obligations, which impose different distributions of burdens but which all have plausible claims to fairness. Even after a simple principle is agreed on, asymmetric agents can differ strongly in their preferred level of stringency for enacting it.

We make three assumptions and exclusions in our discussion. First, we assume that negotiations are among unitary parties, with no supervening authority. Thus we exclude domestic politics and negotiations within a federal state, in which parties can be compelled to participate and a rich set of trade-offs across issues and over time is available (though some of our arguments apply here as well). Second, our focus on single-issue negotiations (with no side payments) excludes the ability to link issues, an important determinant of asymmetry in power. The pure public-goods character of a single pollution problem prevents even the largest agents from bringing targeted incentives to bear on others. Finally, we exclude consideration of negotiations over monitoring, enforcement, compliance, and institutional issues that would accompany any real collective-good negotiation, on the assumption that these functions could be provided cheaply enough not to obstruct otherwise acceptable agreements.

The chapter first discusses how the implicit assumption of symmetric interests makes the process of reaching agreement look too easy. We then present a simple formal model of asymmetric interests, and examine some of its implications in the two-agent case. Next, we discuss the elements involved in reaching agreement in multiparty negotiations. On the question of who participates, a ten-nation illustration shows how asymmetry can lead to stable coalitions of participants. Finally, on the question of what a particular coalition will do, we show how asymmetry complicates the process of bargaining over the simplifying principles that determine relative allocations of responsibility.

STANDARD APPROACHES ASSUME TWO PARTIES, OR SYMMETRY

Most formal analysis of international affairs employs two-person game models, even when the interactions studied involve multiple parties (Keohane 1984; Oye 1986; Snyder and Diesing 1977; Martin 1992). Some analyses, such as Axelrod's study of the repeated prisoner's dilemma (1984), justify this simplification by assuming that agents can discriminate in their choices toward particular others, thereby disaggregating an *N*-party interaction into a series of two-party ones. This disaggregation is not appropriate for multiparty collective-good negotiations, in which a party cannot limit the effects of its decisions to one other party.

Formal models of collective goods that do consider multiple agents most commonly assume that agents' interests are identical, typically by formulating the decision problem of a "representative agent" (Dasgupta 1982, 19–24; Schelling 1978; R. Hardin 1982; Cornes and Sandler 1983; Weitzman 1974). While the main result of these models—that without imposed controls or binding cooperation, public goods will be underprovided and commons overused—is robust to relaxation of the symmetry assumption, assuming symmetry makes the optimal point so simple that the process of negotiating to reach it looks too easy. Identical agents all pollute at equal, excessive levels in the absence of agreement, and at equal, lower levels in the optimal negotiated solution. These classic models do not address the negotiating process of agreeing to move from one point to the other, but it seems clear that identical agents would quickly focus their negotiations on equal reductions by all, and would then agree unanimously on the optimal level. Any reasonable principle used to argue for the required reductions would give the same, optimal result: equal absolute or proportional emission reductions, movement to equal levels, or equal measures per capita, per dollar of GNP, per square kilometer of land area, or relative to historical emissions.

Each agent would of course prefer an asymmetrical solution in which she reduces emissions less than others do. But any argument to support such a solution must be based on some unique characteristic of her situation. If all are identical and all know it, then any argument that justifies a lesser burden for one does so for all. Nor could any agent realistically hope to gain an advantage simply by convincing others of the firmness of her resolve; if one can be that stubborn, so can all. Since no one can expect to prevail decisively in either a principled argument or a contest of wills against a large number of identical copies of herself, the only plausible outcome in bargaining among truly identical agents is unanimous agreement on optimal equal reductions.[1]

Since agents would prefer to cheat if they could get away with it, even a

[1]One qualification is necessary. As Schelling (1960) points out, the availability of an external commitment mechanism may enable one agent (or some agents) decisively to secure an advantage, thereby creating asymmetry where none existed. But if all are identical in interest, resolve, and skill, the outcome of a race to commit can only be simultaneous incompatible commitments, a tie (to be broken by whatever rule applies), or a random process determining which lucky agent (or agents) succeeds in commiting first.

group of identical agents must monitor and enforce compliance. But in a symmetrical world, all other things being equal, even enforcement is likely to be easier. When any cheating represents a departure from otherwise completely uniform behavior, it is likely to be conspicuous and easy to detect. And if cheating is observed—even if no penalty can be applied—both cheater and noncheater will view it as threatening to bring down the agreement, as others imitate and seek equally favorable treatment.

A SIMPLE MODEL WITH ASYMMETRY

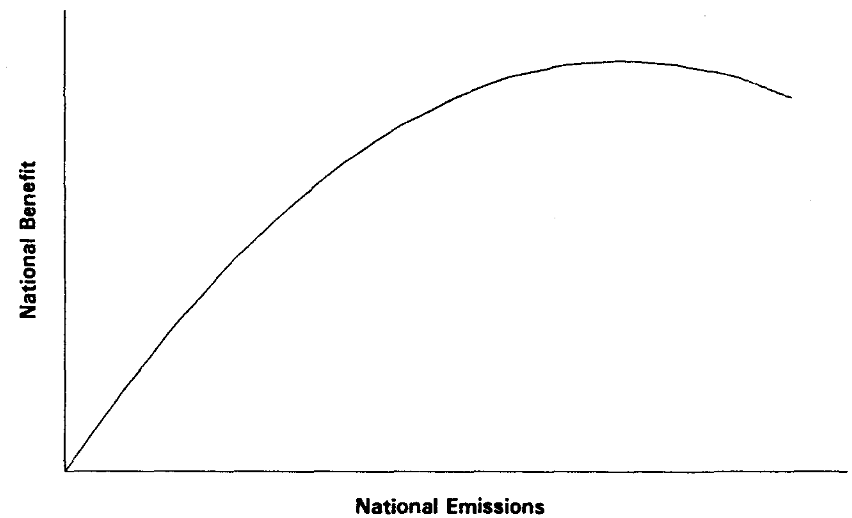
While symmetric models may be unhelpful because they make the process of reaching agreement look too easy, asymmetry of interest is difficult to model because it can be so diverse. In the environmental arena, for example, asymmetries of interest arise from several sources. Some nations may value the environmental good more highly than others, because they are wealthier, more vulnerable, larger, or have "greener" political values. Alternatively, some may find it costlier to reduce their emissions than others, due to differences in economic structure, capital stock, or technological capabilities. On some issues the harm may be asymmetrically borne, say, if one country's emissions blow or flow principally into another. Measures to control an environmental harm may also impose asymmetries, by limiting some activities more strictly than others or defining a new set of property rights. These forms of asymmetry can admit arbitrary variation in the functions that define nations' costs from providing, and benefits from consuming, an environmental good.

How significant are such asymmetries of interest on real issues? While it is not possible to measure comprehensive national environmental interests directly, a few measures of contributions, costs, and political attitudes on climate change illustrate how extreme asymmetries of interest might be. Carbon emissions per capita vary by more than a factor of 100 among countries worldwide, and by a factor of 8 even among the relatively similar countries of the European Community (EC) (World Bank 1992; World Resources Institute 1992; Subak and Clark 1990). One study of limiting EC carbon emissions found that marginal costs of a 10 percent cut would range from a few dollars to several hundred dollars per ton (Barrett 1991b). Estimated costs of protecting coastlines against a one-meter rise in sea level range from $40 to $1,800 per capita in various world regions (IPCC 1991). In a recent comparison of environmental attitudes across twenty-two countries, those considering climate change a "very serious" problem ranged from 33 percent to 73 percent (Dunlap et al. 1992).

We present a simple approach to modeling asymmetric interests that assumes that nations' interests vary only along a single dimension. In Figure

FIGURE 1

Benefit from National Emissions



1, each nation derives a private economic benefit from its own emissions with diminishing and ultimately negative marginal benefits.

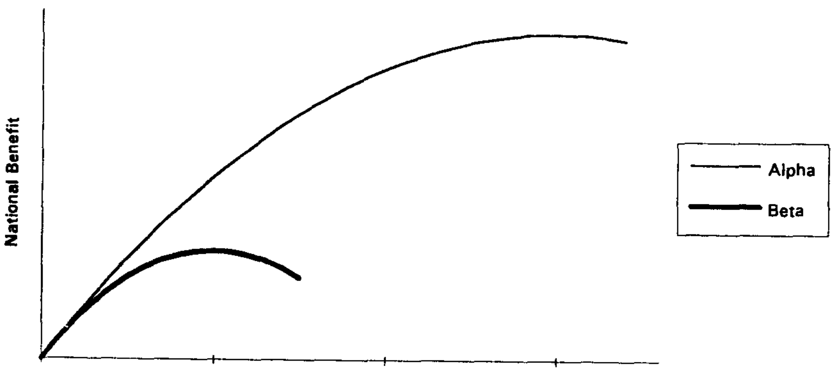
To compare benefits across nations, we assume first that nations comprise different numbers of individuals, but that all individuals in all nations are identical. All nations distribute their private product in the same way and use an additive social welfare function (or alternatively, they distribute their product equally among their citizens). With these assumptions, nations that emit at the same level per capita receive the same benefits per capita; one nation twice as populous as another will derive double the total benefits of the smaller one by emitting double the pollution (whereas the smaller nation could not double its own benefits by doubling pollution, due to diminishing returns). Benefit functions for different nations are thus scaled both horizontally and vertically in proportion to the size of the nation. Figure 2 shows this relationship for two nations, Alpha and Beta, where Alpha is three times larger than Beta.

Nations also suffer increasing marginal cost of environmental harm from the total of their own and others' emissions, as shown in Figure 3.[2] Harm begins with the first unit of world emissions; there is no threshold.

[2] A more general treatment would use a transport matrix, separately specifying the contribution of each agent to each other's harm. Our approach assumes full global mixing, equivalent to a unit transport matrix.

FIGURE 2

Emission Benefits for Two Nations (Alpha three times larger than Beta)



To compare harm across nations we assume that pollution is a pure public bad within each nation, so each nation's harm from a given global pollution level is proportional to its population. A nation twice as populous as another consequently suffers twice the harm from the same level of global pollution. Damage functions are scaled vertically, in proportion to national populations. Figure 4 illustrates this relationship of harms for Alpha and Beta (Alpha is three times the size of Beta).[3]

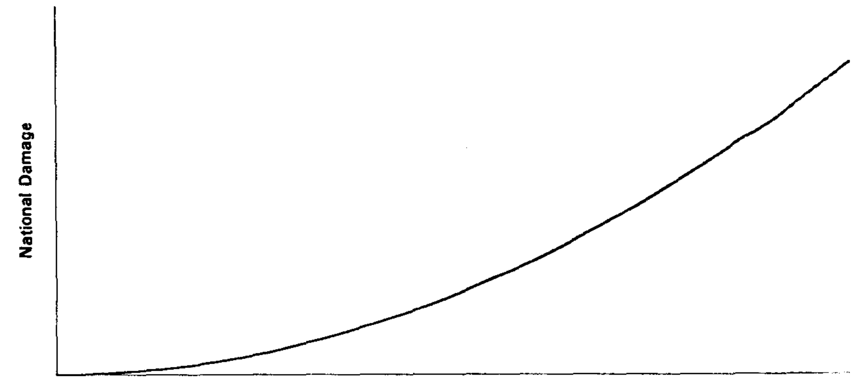
This structure yields three significant insights. First, it suggests a dual focus on the physical agreements undertaken, and the consequences in terms of each agent's benefits. Second, it identifies size with environmental concern. Because the largest nations subsume within their borders the largest fraction of the global harm emissions cause, they are willing to incur a proportionally higher cost to control emissions, at any global emissions level.

Third, it illustrates a three-stage structure of decision making that we contend realistically represents environmental issues: unconcerned, uncooperative, and cooperative decisions. Nations normally only learn of an activity's environmental harm after they have practiced it for some time.

[3]Vertical scaling of damage functions alone can also be used, representing differences in environmental sensitivity or concern (including, with a reversal of sign, nations who gain from pollution, as it has been suggested some may from climate change). Related approaches are found in Hoel 1990, Barrett 1991b, and Parson 1992, chap. 4.

FIGURE 3

Environmental Damage from World Emissions

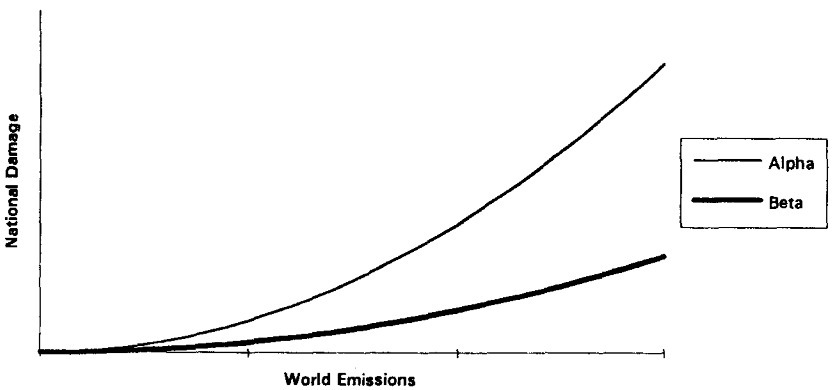


This was the case with ozone depletion, acid rain, and climate change. The delay may reflect advances in scientific knowledge or monitoring technology, or the increasing scale of an activity reaching previously unrecognized environmental constraints. The consequence of the delay is that agents initially optimize without concern for environmental effects, considering only the benefit functions of Figures 1 and 2, and not the cost functions of Figures 3 and 4. Agents in this "unconcerned" stage would emit at the top of their benefit functions, giving equal per capita emissions for all nations.

When agents recognize environmental harm and include it in their decisions they will cut back emissions unilaterally, but only to the point that equalizes their own marginal benefit and damage from their emissions. They thus achieve the uncooperative Nash equilibrium, where each takes the others' emission levels as given. Reductions at this "uncooperative" stage are suboptimal, since each neglects the harm its emissions cause others. Negotiations to pursue further emission reductions begin from this point of noncooperation. In this third, "cooperative" stage, each agent seeks advantageous conditional agreements to reduce emissions further, contingent on others' reducing theirs. These three stages, though conceptually separable, often overlap temporally in real negotiations.

FIGURE 4

Environmental Damages for Two Nations (Alpha three times larger than Beta)



AN ILLUSTRATION WITH TWO PARTIES

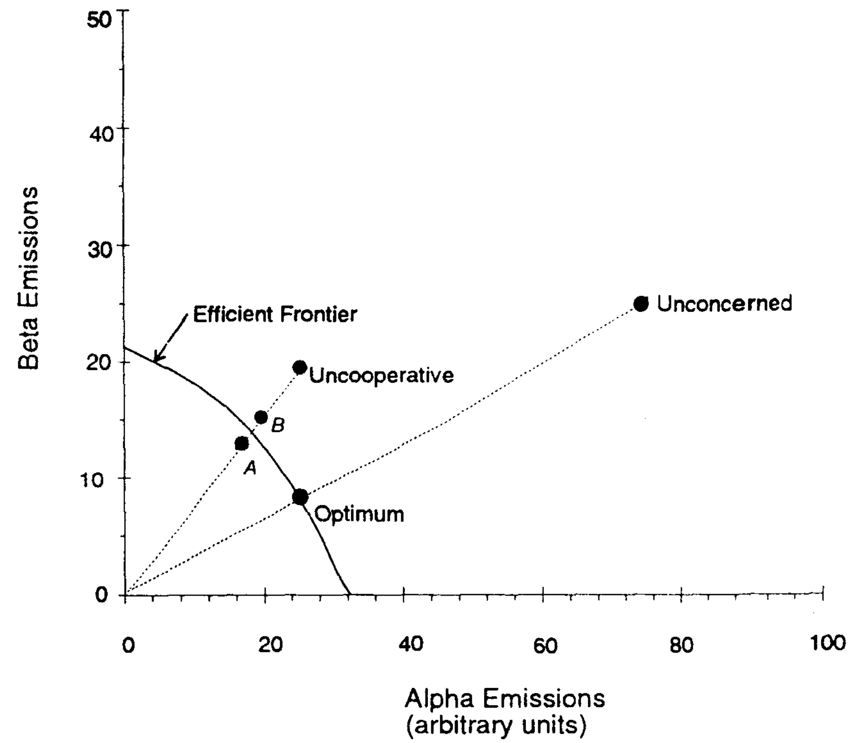
Figures 5 and 6 illustrate this three-stage model with a simple example of bargaining between two nations Alpha and Beta. Alpha is three times larger than Beta. These figures use quadratic cost and benefit functions, requiring one arbitrary parameter that defines the relationship of costs and benefits for Beta, and with proportionality therefore for Alpha. The qualitative results apply with any concave functions.

Figure 5 graphs various possible emission decisions of the two agents. The horizontal and vertical axes measure Alpha's and Beta's emissions respectively (in arbitrary units, denominated so that at the unconcerned point, Alpha emits seventy-five units and Beta twenty-five). At the unconcerned point per capita emissions are equal, with Alpha's total three times Beta's. At the optimal point, per capita emissions are also equal (and lower). The optimization weights each person's welfare equally, so it is unaffected by the asymmetric grouping of more people into Alpha than Beta.

In the initial movement from unconcerned to uncooperative, however, the asymmetric grouping matters in two ways. First, Alpha makes larger unilateral reductions because it bears proportionally more environmental harm within its borders. That the larger country bears a disproportionate burden is an instance of a well-known result for other international public goods, such as alliance burden sharing or output reductions in OPEC (Olson and Zeckhauser 1966). Second, at this stage asymmetry promotes emission reductions. If a fixed total world population is split in some proportion between two countries, and each sets its emissions uncooperatively, then

FIGURE 5

Two-Agent Bargaining: Emissions



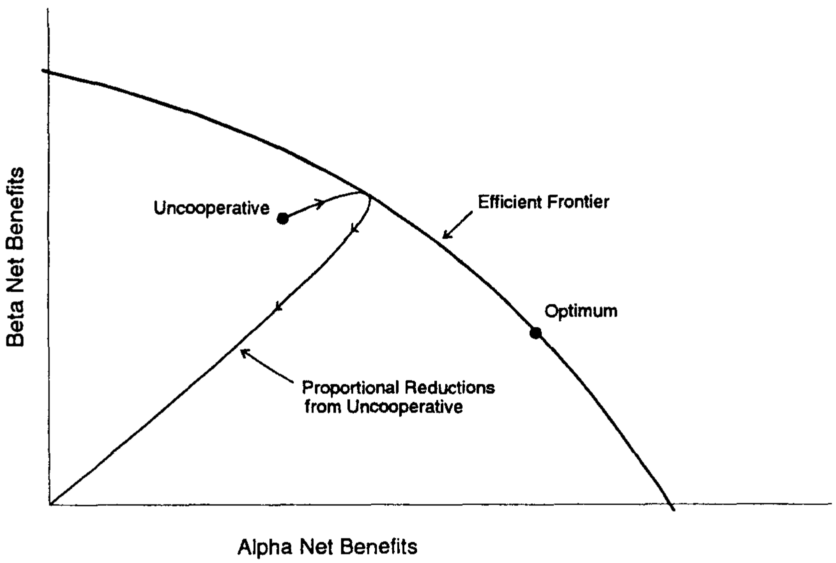
world emissions are largest when the two countries are equal in size, and decrease monotonically as they become more unequal. At the limit, with all people gathered in a single country, it makes the optimal level of reduction.[4]

In negotiations to move beyond the uncooperative point, however, asymmetry obstructs movement in two ways. First, the optimum is unlikely to be

[4]This simple consequence of convexity of cost and benefit functions is not the only case where asymmetry creates gains. For example, when some subset of a group is required for a task that cannot be shared—joining a small-town posse or a rescue party, for example—an obvious rank-ordering of ability to contribute would mitigate the problem of each waiting for others to join. In a quite different spirit, when negotiation concerns multiple issues (including issues of time and risk), it is asymmetric valuation that creates opportunities for joint gains (Raiffa 1982; Lax and Sebenius 1986).

FIGURE 6

Two-Agent Bargaining: Benefits



achieved without side payments, since reaching it can require much larger reductions from Beta than from Alpha (proportionally, and perhaps in absolute terms). Figure 6 illustrates the consequences to each nation of the same set of possible emission decisions; the horizontal and vertical axes are net benefits to Alpha and Beta respectively. Figure 6 shows that Beta can be better off at the uncooperative point than at the optimal point, perhaps substantially so. With certain parameter values it is even possible for Alpha's emissions to be *lower* at the uncooperative point than at the optimal solution, so reaching the optimum would require Alpha to *increase* emissions.

Second, instead of trying to reach the optimum point, Alpha and Beta might agree on the "fair" rule of making equal proportional reductions from their uncooperative emissions—seeking a point on the dashed line joining the uncooperative point to the origin in Figure 5, or on the curved path marked with arrows in Figure 6. In this case, agents will disagree on how far to reduce, since their benefits are maximized at different reduction levels, marked *A* and *B* in figure 5. Solution concepts such as the Nash bargaining solution or the constrained joint optimum lie between their two preferred points.

NEGOTIATION AMONG MULTIPLE ASYMMETRIC AGENTS

The two-party negotiation illustrated above shows several effects of asymmetry on potential distributions of obligations. But in these examples there is no agreement unless all (i.e., both) affected parties choose to participate; the only negotiation is over how much each will do. In multiparty negotiations, in contrast, a cooperative agreement can be negotiated with less than full participation.

Consequently, in multiparty negotiations there are two distinct issues to settle: Who is in, and how much will each do? Asymmetry of interests can affect both issues. Complicating matters further is the fact that the two issues interact. Changing what the participants in an agreement will do can change the set of agents willing to participate, and changing the set of participants can change the stringency of measures they are willing to undertake.

We examine the two issues separately, however, considering two ideal types of bargaining situations: one in which the meaning of "participating" is somehow fixed and agents negotiate over who will participate, and one in which a fixed set of participants must agree on the distribution of obligations among them.

WHO PARTICIPATES?

First, we assume that what it means to participate in an agreement is fixed, so *who participates* is the only item for negotiation. In a real environmental negotiation this "fixed decision" case can arise in several ways. The object of discussion could be an intrinsically binary decision, such as nations deciding whether to require separated ballast tanks in oil tankers, or to prohibit hunting whales, or small-town citizens deciding whether to vote for property tax reductions. Even on matters that are not intrinsically binary, one specific proposal may so dominate debate that nobody discusses other measures, but only whether they will join or not. For example, negotiations over European acid rain control for many years only considered whether countries would make a proposed 30 percent sulfur dioxide reduction (Levy 1993). In formal models of bargaining, joining is often made into a binary decision by assuming that any fixed set of participants will reach some particular bargaining outcome, such as the Nash bargaining solution (Nash 1950).

In this situation each agent faces a binary decision: whether to participate or not. Since negotiation begins from the uncooperative point, nobody is willing to participate alone; but because of the public-good character of the issue (and assuming the meaning of participation has been sensibly chosen), all prefer unanimous participation to unanimous nonparticipation.

Between zero and full participation, however, each agent's willingness to join is conditional on others' joining as well, and on how their joining in turn depends on that agent's joining. Such conditionality relationships may take two forms. The first are *rules,* based perhaps on equity principles or politics. For example, agents may require that particular others join if they do: Canada may be unwilling to join an agreement unless the United States does, or the United States unless Germany and Japan do; the European Community may either all join together or not at all; or all agents may perceive a natural rank-ordering of decisions, requiring the biggest, the richest, the greenest, or those who benefit the most to join before others will.[5]

Agents' conditions for joining can also be driven by a *benefits calculation.* An agent will only join a coalition if the benefit of joining exceeds the benefit of staying out. Agents may perceive that other potential coalition members will not join unless they do, effectively raising the cost of not joining. An agent's choice is between the benefit of joining the proposed coalition and the benefit of not joining the smaller coalition that would form without her, from which all who require her participation have withdrawn (and all others who require theirs, and so on).

These two forms of conditions on individual participation allow us to define conditions for stable coalitions. A coalition C is stable if and only if: (1) no member requires the participation of somebody not in C as a condition of its participation, and (2) the withdrawal of any member would result, directly or indirectly, in other withdrawals whose effect would be to leave the original agent worse off. (Note that this condition does not require that *all* others withdraw, but only enough to offset the first one's gain from withdrawing.)

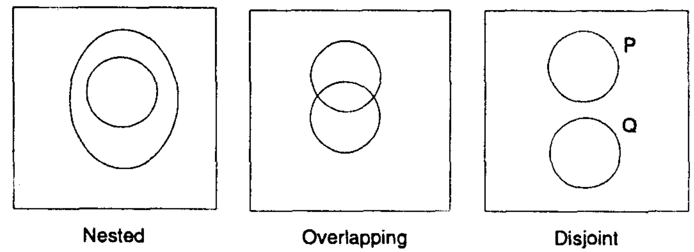
Stable coalitions comprise natural groups of cooperators that are likely to persist once established, for all members will recognize that if they withdraw they will end up worse off. In the case where agents act only on benefit calculations, not on rules, stable coalitions are the asymmetric equivalent of the "minimum viable coalitions" (MVCs) of identical-agent cooperation models. MVCs are participating subgroups of just sufficient size that their members are better off if all participate than if none do (Schelling 1978). Since anyone's withdrawal from an MVC leaves the rest worse off than if none participate, each member would reasonably expect all others' participation to depend on its own, and so stay in.

Whether stable coalitions are defined by benefit calculations, rules, or

[5]The experimental literature on individual decision-making shows clearly that such nonconsequential equity rules can be important. People *do* decline to make or accept advantageous offers that violate obvious fairness norms (Bazerman and Neale 1992). Of course, these empirical results are for individuals, not nations.

FIGURE 7

Stable Coalitions, Three Possible Configurations



both, many potential stable coalitions can exist within any group negotiating. Whether a stable coalition can be expected to emerge in the first place depends in part on the number of stable coalitions and their relationships to each other.

If only one stable coalition exists, and negotiators have full knowledge, we expect agents to perceive their interests correctly (and each to perceive the consequences of attempting to free ride), so the stable coalition will form. If multiple stable coalitions exist, relationships among them can be of three kinds, as shown in Figure 7. Pairs of stable coalitions can be nested, overlapping, or disjoint.

If there is a natural ordering of agents joining an agreement, by size or environmental concern or net benefits, then stable coalitions will be nested. In this case we expect the largest to form, since all within the largest coalition are better off within it than with any other stable outcome.

If stable coalitions are overlapping or disjoint, strategic interaction between them can obstruct the formation of any of them, so the outcome is indeterminate. Disjoint stable coalitions, such as groups P and Q on the right of Figure 7, are each stable relative to the uncooperative starting point, but each would benefit at no cost if the other formed. If each group would no longer benefit from forming once the other already has formed (which depends on details of the payoffs), then each has an interest in delaying and in finding ways to press the *other* group to form. With overlapping coalitions, this tendency to delay may be partly offset by those agents in the intersection pressing those in the remainder of either stable coalition to join them.

It is each agent's *perception* of whose participation depends on her own that determines agents' perceived consequences of joining, and so for the agents collectively determines what coalitions are stable. Forming a stable coalition thus requires the emergence of a consistent set of expectations of whose participation depends on whose. If all the agents in some group come to perceive that they all depend on each other, and that nobody else is likely to join, then they all are likely to join. Any group would wish to resist the formation of such perceptions if the alternative is that some other group will bear the burden, but it may not be able to. A sufficiently large subgroup possessing some salient common characteristic, even if it is irrelevant to their benefits from cooperating (perhaps in a group of individuals some are Stanford graduates, or have red hair), could come to be identified by its members and others as a "natural" coalition. If each member of such a group comes to expect that enough others will join if she does, then all will. Their shared characteristic puts them at a disadvantage in the subtle struggle to push the responsibility onto some other group.

With strongly asymmetric groups, the process of identifying such characteristics will be difficult and contestable. When agents possess *many* varying characteristics, more or less conspicuous or relevant to their participation, each agent has an interest in maximizing others' perception that characteristics that *they* share (but she does not) are the "natural" bases for forming cooperative groups. With many overlapping stable coalitions, some characteristics will come to seem more salient and relevant than others, perhaps due to arbitrary artifacts of language, history, and context, as well as various agents' attempts to frame issues in ways that render their own participation inessential. Each agent's costs and benefits influence this struggle, affecting her ability to threaten credibly not to join, but do not by themselves determine the outcome.

AN ILLUSTRATION: COOPERATION IN A TEN-NATION WORLD

We illustrate the preceding argument with a specific model of ten-nation bargaining over carbon dioxide emissions that illustrates nested stable coalitions. The ten nations have quadratic cost and benefit functions as shown in Figures 1 through 4 above; their sizes are chosen so that their emissions at the "unconcerned" stage match those of the ten highest carbon-emitting nations of the 1980s, as shown in Table 1. Emissions are expressed in millions of metric tons of carbon, and are rounded.

The arbitrary calibration of costs and benefits is set so that movement from the unconcerned to the uncooperative (Nash equilibrium) point represents a reduction of somewhat less than 10 percent for the largest nations. Bargaining starts from this uncooperative point. From this point, no nation would undertake further reductions on its own, but nations will consider joint agreements to reduce together.

As in all bargaining models, restrictive assumptions are needed to generate unique outcomes. We invoke a particularly simple participation rule: Countries are arranged in order of size and environmental concern from biggest to smallest, and no country will consider joining an agreement to reduce emissions unless all larger countries, who also receive larger net benefits, have joined. The two largest consider joining first, then the third considers joining them, and so on. Beyond this rule, only a nation's net benefits determine whether it will join or not; if it will be better off by joining, it joins.

TABLE 1

Nations in the illustration

Nation Number	Corresponds to	Unconcerned Emissions (MT of carbon)
1	USA	1300
2	former USSR	900
3	China	450
4	Germany	275
5	Brazil	250
6	Japan	250
7	UK	150
8	Mexico	130
9	India	125
10	Colombia	120

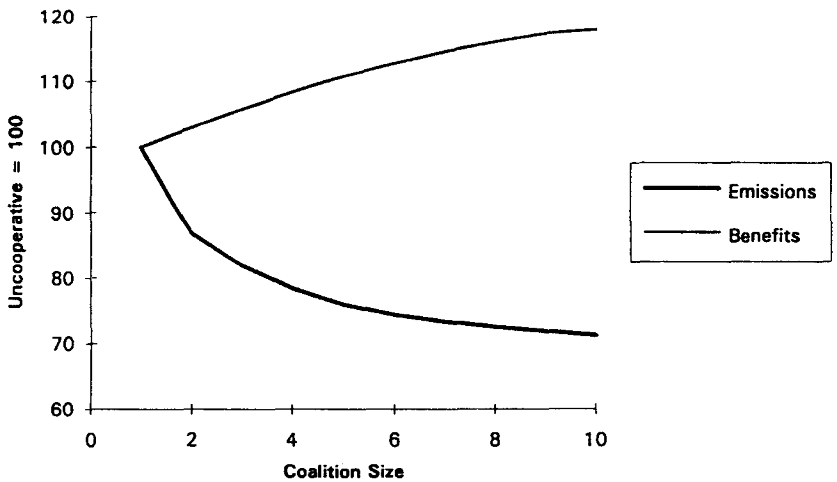
The meaning of "joining an agreement" is assumed to be fixed; all participants will make equal proportional reductions from their uncooperative emission levels, to a level given by the Nash bargaining solution. Nonparticipants continue to optimize uncooperatively, so as the cooperative coalition grows and cooperators decrease their emissions, noncooperators increase theirs slightly.

Figure 8 illustrates the result of this bargaining process. Cooperative coalitions make progressively larger emission reductions, and gain progressively larger benefits, as the coalition grows. Adding an additional cooperator makes each member of the original set of cooperators go further, given the Nash bargaining solution. Noncooperators increase their emissions by less than cooperators decrease theirs (a consequence of concave benefit functions), so total emissions decline as the cooperative group grows. In this figure, both emissions and benefits are scaled relative to their levels in the uncooperative equilibrium, which are assigned the value 100.

Table 2 shows each nation's net benefits when each size of coalition forms, illustrating the basis of nations' decisions to join or not join. Because all benefit figures are in arbitrary units, only comparisons between a nation's benefits under different coalitions are significant. The first column, headed "Uncoop", shows benefits to each nation at the uncooperative equilibrium, the last under a coalition of all ten nations, and the others under intermediate sizes of cooperative coalitions. With the assumptions and parameters used here, there are three points to note.

FIGURE 8

Cooperators' Emissions and Net Benefits as a Function of Coalition Size



First, the largest two nations join to reduce emissions, even if nobody else joins them. This two-nation coalition is thus a stable cartel, in that no nation would unilaterally wish either to join or leave it (Donsimoni et al. 1986). The third nation would not benefit from joining this group, so it does not. But note that if the third and fourth nations both join, they—as well as all others—enjoy higher benefits than under the two-nation coalition. Here then is a coalition that can form if nations three and four know the relevant benefits and can come to expect that they will both move together. With full knowledge and no obstacles to communication, they are likely to join. Once formed, this group will likely be stable, since its marginal member, the fourth, will expect that if it withdraws so will the third, and so will evaluate its benefits relative to those under the two-nation coalition. Both nations three and four prefer to remain participating. Similarly, the nine-member coalition is stable because nations five through nine are all better off under the nine-member coalition than under the four-member one.[6] The ten-nation coalition is not stable, since nation ten prefers to be outside the nine-member coalition rather than inside the ten-member one.

The particular configuration of stable coalitions in Table 2 merely illustrates the range of possibilities. Other configurations can be obtained by vary-

[6]The apparent ties for nations five through eight are due to rounding. They are in fact strictly better off under the nine-nation coalition.

TABLE 2

Net benefits to each nation as a function of coalition size

		Number of nations reducing below uncooperative level								
Nation	**Uncooperative**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	573.2	573.6	574.2	574.6	575.1	575.6	575.9	576.2	576.5	576.8
2	397.2	**397.8**	398.3	398.7	399.1	399.6	399.8	400.1	400.3	400.6
3	198.7	199.6	199.6	199.8	200.1	200.3	200.5	200.6	200.8	200.9
4	121.4	122.0	122.5	**122.2**	122.4	122.5	122.6	122.7	122.8	122.9
5	110.4	110.9	111.3	111.7	111.3	111.4	111.5	111.6	111.7	111.8
6	110.4	110.9	111.3	111.7	112.0	111.4	111.5	111.6	111.7	111.8
7	66.2	66.5	66.8	67.0	67.2	67.4	66.9	67.0	67.0	67.1
8	57.4	57.7	57.9	58.1	58.2	58.4	58.5	58.1	58.1	58.2
9	55.2	55.5	55.7	55.8	56.0	56.1	56.2	56.3	**55.9**	55.9
10	53.0	53.2	53.4	53.6	53.7	53.9	54.0	54.1	54.2	53.7
Total Benefits	1743	1748	1751	1753	1755	1757	1758	1758	1759	1760

ing the parameters of the common cost and benefit functions, and the relative sizes of participating nations. Under different configurations the smallest stable coalition can contain either two or three members, various intermediate sizes can be stable, and the unanimous coalition can either be stable or not. This example illustrates two results: that cooperative coalitions short of unanimity can form and be stable; and that when stable coalitions are nested, the largest can be expected to form.

WHAT DO THE PARTICIPANTS DO?

The preceding analysis addressed what sorts of coalitions will form when the distribution of requirements to be imposed on any particular coalition is specified. We now turn to the other ideal type of multiparty negotiations, and inquire what sorts of agreements a specified coalition will reach.

We suppose that a fixed set of N agents has agreed to cooperate and must negotiate the magnitude of emission reduction each will undertake. They must accept an agreement unanimously, since we assume that nobody in the provisional set of participants opts out. This means deciding on the values for N numbers, that is, picking one point in an N-dimensional space.

How will they choose such a point? As in the two-dimensional case, there are many potential distributions of obligations that all would prefer to the uncooperative status quo, and all know it. Unlike the two-dimensional case, though, it seems unlikely that they will come to agreement through unstructured haggling. Such a negotiation would be too complex and take too long. The negotiation has too many degrees of freedom, and moreover would grant each agent unlimited license to argue (whether sincerely or not) that its special situation calls for a lower burden than others will bear.

Instead, agents will normally first negotiate over devices that simplify and restrict the space of possible agreements by fixing relationships among the obligations they undertake. We call such devices *principles*. They serve two purposes: an *informational* one, promoting a manageable negotiation by reducing the amount of information that negotiators must process in seeking and reaching agreement; and a *moral or rhetorical* one, providing support for claims that particular allocations of obligations are fair or not, and that particular bargainers are carrying their due burdens or not.

A principle, once chosen, reduces the bargaining space but not to a single point. Hence, choosing a principle does not end the negotiation, but simplifies it, leaving a restricted set of items still to be negotiated. Many often-used principles, such as equal proportional emission reductions or equal contributions per unit GNP, are so constraining that only a single degree of freedom remains to be negotiated: how much total reduction or contribution. Other principles could leave more. For example, industrial countries could cut to one level and developing countries to another, or emission entitlements could be defined as linear functions (to be negotiated) of population, current emissions, GNP, and land area.

By far the most common principle used in environmental agreements is equal proportional emission reductions. For example, more than seventy nations have ratified the Montreal protocol to protect the ozone layer, a 1987 agreement to cut chlorofluorocarbons (CFCs) by 50 percent, amended in 1990 to eliminate them entirely. Under the Convention on Long-Range Transboundary Air Pollution (LRTAP), twenty-one nations have agreed to cut sulfur emissions by 30 percent, and twenty-three to freeze emissions of nitrogen oxides (while a "club" of twelve nitrogen activists separately agreed to 30 percent cuts). The eight North Sea nations agreed to reduce pollution by dioxins and heavy metals by 70 percent and other chemicals by 50 percent, while the six Baltic nations pledged to cut chemicals and nutrients flowing to that sea by 50 percent. In two nonbinding declarations since eclipsed by the weaker climate convention signed in Rio, about a dozen industrial nations pledged to cut carbon emissions by 20 percent by the year 2005 (UNEP 1987; 1990; Levy 1993; Haas 1993; Toronto Declaration 1988; Hague Declaration 1989).

The few exceptions to equal proportional reductions almost all fall into two categories: grouping countries into two classes (industrial and developing), with identical measures for all members of each class; or specially negotiated, seemingly universal measures that in fact create special exceptions for one or a few parties (and everybody knows it). For example, the original 1987 Montreal protocol included a provision drafted to accommodate one Soviet CFC plant, which included the output of a new plant in a country's baseline if it met four conditions; only the Soviet plant met the conditions (UNEP 1987, article 2 para. 6; Benedick 1991, 83). In each case, large groups of countries are still subjected to equal proportional reductions.[7]

Principles other than equal percentage reductions could be effective simplifying devices for complex negotiations, but not just any constraint will do. There are a number of desiderata if a principle is to simplify negotiations. It should be easy to articulate, recognize, and agree upon; and it should be "sticky," resistant to incremental chiseling or renegotiation once adopted. It must tie together the contributions of different agents in a supportable manner, so that all perceive it to be stable, expect others not to chisel, and expect to be hurt if they try to chisel.

For example, consider the following principle: "Emission reductions of countries A, B, C shall be in the ratio 1 to 1.32 to 1.73." Such a principle, stating a crystal-clear algebraic relationship among parties' emission reductions, could in theory serve the function of reducing bargaining degrees of freedom. It could never do so in practice, though, since it appears utterly arbitrary and would be subject to participants' constant efforts to improve their positions just a little. It would take forever to negotiate, and would not be resistant to chiseling and renegotiation.[8]

Rather, an effective negotiating principle should be a *focal point* in the sense of Schelling (1960), commanding attention by virtue of salience, uniqueness, or discreteness. These characteristics are defined cognitively, relative to the perceptions of the particular agents negotiating; as both Schelling and Kreps (1990b) point out, different resolutions may appear salient or unique to different groups. Some principles are likely to be focal points for almost all groups, however; for example, setting some quantity equal for all participants; setting something to zero or another round number; or maintaining something at its status quo value.

The 1990 amendments to the Montreal protocol on the ozone layer illustrate the usefulness and power for negotiations of strong focal points. When it became clear from scientific and technical reports that "zero" (i.e., a complete phaseout of a broad class of chemicals) was feasible and environmentally desirable, the ninety-odd negotiating nations agreed on full phaseouts essentially without difficulty. "Zero" is a uniquely salient and powerful

---

[7]We know of only one recent environmental negotiation that proceeded by unstructured haggling, and yielded highly asymmetric assignment of obligations: the European Community's Directive on Emissions from Large Combustion Plants. This agreement's uniqueness, and its five-year, twice-weekly negotiating history, support our claim that this approach is difficult (Grubb 1989).

[8]Note, however, that a principle to cut back in proportion to GNP or to emissions excess over 1990 levels might be acceptable, and such a principle could yield contributions in the ratio 1 to 1.32 to 1.73.

negotiating outcome, but cases such as ozone where it is also the right (or a feasible) answer are rare. Consequently, this precedent is of limited use in current major international negotiations on issues such as climate change and species loss, where total elimination of the harmful activities is not possible.

How will a principle be chosen to serve this simplifying function? This is no problem at all with symmetric agents, for every reasonable principle is equivalent to equal reductions. When agents differ on many dimensions, however, the number of plausibly fair principles will be large (Young 1991). With many seemingly reasonable rules for reductions available, none can approach the salience that equal reductions have in the symmetric case.

Moreover, with asymmetric interests different principles will impose relatively lighter or heavier burdens on different nations. "Equal proportional reductions" favors countries with high emissions and low abatement costs, while "equal per capita emissions" favors populous countries. Different reasonable arguments may justify higher emissions for different countries: Japan may claim higher carbon emissions because its energy use to GNP ratio is already the lowest of the industrial countries, the United States because it defends the world, or Canada because it is so big and cold. Negotiators can be expected to favor, often sincerely, those "neutral" principles that advantage them (Messick and Sentis 1983).

But the same uniqueness that makes principles effective simplifying devices once adopted makes them hard to negotiate over. A compromise between two principles is likely to lack the sticky, discrete character that makes principles useful. When there are many candidate principles, different agents strongly favor different ones, and compromise is difficult, reaching agreement will be a challenge. Moreover, if failure to choose any principle means a breakdown or an impossibly cumbersome and drawn-out negotiation that all want to avoid, then the negotiation over *which* principle to use has the character of a chicken game, giving agents incentives to fabricate commitments to one principle or another.

How agreement will be reached in any real negotiation we cannot predict. As long as negotiators remain committed to incompatible principles, there will be no agreement. But the record of real negotiations suggests that parties usually do break this impasse, surprisingly often by agreeing on symmetric measures despite asymmetries of interest. Burtraw and Toman (1991) argue that a process through which perceptions of fairness change as negotiations proceed is likely to promote agreement. If the negotiating group includes some "neutrals" who initially lack strong preferences between the principles under contention, then persuading the neutrals may be a particularly effective way of forcing the group's expectations to converge on one principle rather than another.

Finally, the choice of a principle, while necessary, is but a first step toward concluding the negotiation. When, for example, parties have agreed on equal proportional emission reductions, they must still agree on the level. Moreover, the two negotiations we have treated as separable ideal cases—who participates and what they do—are in fact simultaneously determined. If those most concerned begin discussing an agreement among themselves, they may then need to change the measures they adopt to entice new members into their coalition. Alternatively, if more parties become interested in joining, the resulting gains may make it possible to adopt stronger measures.

The recent climate negotiations illustrate several of the phenomena we have described. A group of activist nations, led by the European Community, had pressed strongly for a convention with binding national emission limits. They perceived U.S. participation as essential to any successful agreement, however, so they accepted a much weaker treaty than they preferred in order to induce the U.S. to sign (rather than acting by themselves, or signing a convention in which some committed to limit emissions and some did not). The convention includes "principles," but they are so vague as to put scarcely any limits on possible future negotiations of specific emission limits. About the only clear implication is that developed countries will do the bulk of emissions limitation. The convention does include significant measures for the development of national plans, reporting of national emissions, and creation of institutions. Many hope that these hortatory and institutional measures alone will induce national actions sufficient to bring about whatever global emission reductions are required. We think it more likely that these measures and principles will simply serve as foundations for future negotiation of more concrete national emission obligations. If so, then most of the hard bargaining remains to be done (UN General Assembly 1992).

CONCLUSION

The usual analytic view of the world, which assumes symmetric interests, is not useful for most collective-good negotiations in the real world. In most important multiparty negotiations, the parties are highly asymmetric in their interests. This has several important consequences. First, there can be stable groups of participants in an agreement short of universal participation. Insistence on universal participation is rooted in symmetric thinking.

Second, most multiparty bargains will need to employ simple principles to distribute obligations among participants. With significant asymmetries of interest, however, there will exist multiple plausible candidate principles, each with its own distributional implications. Appealing concepts such as equal sacrifice in pursuit of a common purpose become ambiguous. Choosing a principle is likely to be contentious and difficult. Even with an agreed-upon principle, differences will remain, on questions such as how stringently it should be applied.

The principle of equal proportional emission reductions, widely employed in environmental agreements, can be highly inefficient and unequal in the distribution of burdens, given asymmetry of costs (Bohm 1990). But the requirements of salience, seeming fairness, and stickiness may mean that the choice is between a flawed agreement that can be reached and an efficient one that cannot. In such cases the political and informational advantages of a viable principle must outweigh its efficiency losses.

This trade-off may change as the stakes in a negotiation rise, particularly if the asymmetries among nations are considerable. On climate change, the most pessimistic projections of losses from high-cost emission abatement policies approach 10 percent of some nations' GNPs, while the most pessimistic projections of losses from climate change approach catastrophe (Broecker 1987; Manne and Richels 1990). If these projections have a chance to prove accurate, the potential environmental risks make an agreement to control emissions desirable, while the losses inherent in a simple agreement of equal proportional reductions would probably render such an agreement infeasible. Parties who would be willing to go along with symmetric deals if the stakes were low would likely be more sensitive to perceived unfairness in the burden they bear relative to the gains they reap when the sacrifices are significant. If nations are to achieve agreement on issues as consequential as climate change, significant creativity may be required in inventing and identifying focal points and principles of fair burden, enlarging the set of approaches to feasible agreement and thereby reducing the likelihood of sustained impasse.

Cooperation in the unbalanced commons will be more difficult than the consideration of symmetric models, or precedents of moderate-stakes agreements, would suggest. Such cooperation will be facilitated by understanding the conceptual lessons laid out here. Simple principles will play a salient role, in defining both the group of cooperators and the actions they take. Nations will be unlikely to participate if their peers, and those they perceive to have a greater responsibility, do not. The language of agreement will enshrine symmetric treatment, by some definition, within groups of fairly heterogeneous nations. This pursuit of symmetry, though sacrificing efficiency, will promote both the feasibility of agreements and their perceived fairness.