The Oslo-Potsdam Solution to Measuring Regime Effectiveness: Critique, Response, and the Road Ahead

*Jon Hovi, Detlef F. Sprinz, and Arild Underdal**

1. The Challenge of Measuring Regime Effectiveness

International organizations and regimes are established in order to perform a particular function or achieve a certain goal. One of the basic questions to be asked about these institutions is therefore how effective they are in delivering what they were established and designed to achieve. To answer this question we need to, first, determine precisely what we mean by "effective" in this particular context and, second, develop an operational procedure for measuring "effectiveness"—preferably in standardized terms that allow for systematic comparison across a wide range of cases. This article builds on previous efforts by the authors on the first question, yet concentrates primarily on the second problem.

Somewhat different solutions for measuring regime effectiveness have been proposed. The perhaps most explicit and rigorous formula is referred to as the "Oslo-Potsdam solution." Two attractive properties of this formula are that it combines two notions often used in assessing regimes, and that it yields comparable scores falling within a standardized interval between 0 and 1.[1] Yet, objections have been raised, most notably by Young.[2] This critique should be taken very seriously, since Young himself is a leading scholar in the field of regime analysis, actively engaged in developing tools for studying regime consequences. In this article, we review and respond to the critique, arguing that it is at least partly misplaced. In particular, we take issue with some of his objections to game theory. We nevertheless agree with several of Young's conclusions, albeit sometimes for other reasons than his own. We also acknowledge that there are potential problems in using game-theoretic constructs as a basis for empirical estimation of the various elements contained in the formula. However, we

* This research was supported by a joint grant from the German Academic Exchange Council (DAAD) and the Research Council of Norway (NFR). We gratefully acknowledge helpful comments by Oran R. Young, two other (anonymous) reviewers and Matthew Paterson. Anja Bauer and Lynn A. P. Nygaard provided excellent editorial assistance.

1. Underdal 1992; Sprinz and Helm 1999; and Helm and Sprinz 2000.
2. Young 2001.

Global Environmental Politics 3:3, August 2003

conclude that in measuring regime effectiveness empirically, we do not necessarily need to resolve all the more "advanced" conceptual issues raised by the critique.

The article is organized as follows. In Section 2, we briefly recapitulate the essence of the Oslo-Potsdam solution to measuring regime effectiveness, while Section 3 summarizes the critique that this solution has triggered in the literature on international regimes. Section 4 provides a response to this critique, explaining why it is at least partly misplaced. Finally, section 5 extends and develops further our own views on the Oslo-Potsdam solution.

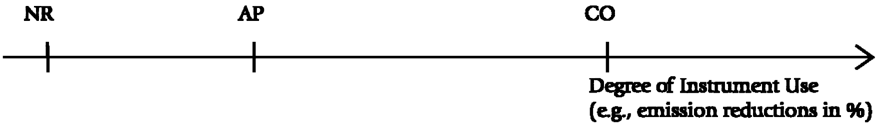
2. The Oslo-Potsdam Solution

The Oslo-Potsdam solution is an umbrella term that refers to two closely related options in empirical research that were designed to answer three distinct questions originally posed by Underdal:[3] "(i) What precisely constitutes the *object* to be evaluated? (ii) Against which *standard* is the object to be evaluated? (iii) How do we operationally go about comparing the object to our standard?" These questions are designed to target various aspects of regime effectiveness, and the two options in the Oslo-Potsdam solution use different scaling techniques to quantify a no-regime counterfactual (NR), a measure of actual performance (AP) of the international regime, and the determination of a collective optimum (CO). Because each of these scores refers to only one underlying dimension of evaluation, the effectiveness of international regimes is assessed either by relating the first two or the last two of these values to each other[4] or by combining all three.[5] We will briefly describe the two approaches in more detail.

Underdal[6] originally used the three elements to construct two different measures of effectiveness. One compares the actual performance (AP) obtained under the regime to the no-regime counterfactual (NR). The other compares the actual state of affairs (AP) to the best solution that could have been accomplished ("the collective optimum" or CO; see Figure 1). These measures were designed to address two different questions: The former was proposed as a tool for examining whether and to what extent "regimes matter," the latter was designed to answer the question of whether and to what extent a particular problem is in fact "solved" by the existing regime.[7] Helm and Sprinz[8] combine all three components in one comprehensive measure. They conceive of the distance CO-NR as the potential for improvement over the no-regime counterfactual, and the distance AP-NR as the realized degree of improvement by way

3. Underdal 1992, 228–229.
4. Miles et al. 2002.
5. Helm and Sprinz 2000; and Sprinz and Helm 1999.
6. Underdal 1992.
7. Note that both of these questions involve causal inference; they can be answered only if we are able to distinguish the change brought about by the regime itself from change caused by other factors.
8. Helm and Sprinz 2000.

Figure 1
Measuring Regime Effectiveness



Effectiveness Score $E = \frac{AP - NR}{CO - NR}$

Sensitivity of Effectiveness Score $S = \frac{AP + 1 - NR}{CO - NR} - E = \frac{1}{CO - NR} = \frac{dE}{dAP}$

Notes: NR = no-regime counterfactual
CO = collective optimum
AP = actual performance

Source: Helm and Sprinz 2000, 637.

of the international regime. By dividing the latter distance by the former, they arrive at a simple coefficient of regime effectiveness that falls into the standardized interval [0, 1] (see Figure 1 for illustration).[9]

Several different options exist to determine the no-regime counterfactual and the collective optimum. Helm and Sprinz[10] suggest a game-theoretic foundation by conceptually deriving the no-regime counterfactual as the Nash equilibrium of a game and a collective cost minimum for all countries involved in the regulatory domain (e.g., transboundary or global environmental problems).[11] In their empirical research, both Miles et al.[12] and Helm and Sprinz use a structured expert-based scoring mechanism to derive point estimates of the no-regime counterfactual (NR). In order to derive the collective optimum operationalized as a collective cost minimum, Helm and Sprinz combine abatement cost functions, indirectly derived (political) damage cost functions, simultaneous equation modeling, and calculations of environmental impact. In addition, the sensitivity of the effectiveness scores can be easily computed (see Figure 1 for an example). By contrast, Miles et al. established a hierarchical procedure for the derivation of data for the collective optimum, starting with the

9. The precise equations can be found in Helm and Sprinz 2000.
10. Helm and Sprinz 2000.
11. Sprinz and Helm 1999.
12. Miles et al. 2002.

best external sources available and ending with the case study specialist's best judgment. A closer comparison of the methods used by Miles et al. and Helm and Sprinz reveals similarities as well as differences.

What they have in common is that the basic components of the analysis are conceptually identical, namely measures of the no-regime counterfactual, actual performance, and the collective optimum. Both streams of research also agree that specific treaty obligations are unlikely to be an appropriate choice due to endogeneity problems.[13] A main difference is that Miles et al. use ordinal metrics, whereas Helm and Sprinz use interval scales—which give rise to the use of a sensitivity coefficient for the effectiveness score (see Figure 1). In addition, Miles et al. pursue a structured empirical approach, whereas Helm and Sprinz combine an analytical derivation of core quantities of interest (no-regime counterfactual and collective optimum) with structured empirical data generation. Third, Miles et al. generate measurement results at the level of the (aggregate) regime and its particular phases, whereas Helm and Sprinz derive country-level regime effectiveness scores as well as aggregate regime effectiveness scores. Finally, and perhaps most important, there is a difference in strategy of analysis. Whereas Helm and Sprinz develop a single measure, Miles et al. derive two independent measures that are not directly related to each other (see above). As a consequence, while analysis of the variance of regime effectiveness using the former measure necessitates one analysis, the latter measure requires two separate analyses.

The advantage of the Miles et al. measurement method is clearly its relative ease of application as compared to the technical requirements needed for the Helm and Sprinz procedure. In addition, there are differences in the conceptualization of the collective optimum. While Miles et al. suggest a "functional (technical) optimum," Helm and Sprinz consider a range of options—and choose a collective cost minimum without side-payments as their empirical solution. We discuss these and other alternative options later in the article, having first considered the critique of the Oslo-Potsdam solution.

3. Summary of the Critique

In a recent critique of the Oslo-Potsdam solution, Oran Young[14] raises a wide range of questions about and objections to the above formulas. Many of his remarks pertain as much to most other attempts at measuring effectiveness that we know of, but some focus on properties that distinguish the Oslo-Potsdam formula from other approaches. We will concentrate mainly on the latter category. Following Young himself, we distinguish between objections that deal with conceptual or analytical issues, on the one hand, and those that focus on problems encountered in empirical applications, on the other.

13. See Helm and Sprinz 2000, 635; and Bernauer 1995. See also Section 5.
14. Young 2001.

3.1 Conceptual Challenges

Young raises important conceptual questions regarding two of the three core concepts. With regard to Helm and Sprinz'[15] conceptualization of the no-regime counterfactual, Young points out that some games have more than one Nash (non-cooperative) equilibrium, meaning that the proposed solution will often fail to identify one unique outcome.[16] In addition, Young argues that the Nash equilibrium constitutes an extreme case in the sense that it leaves "no room for any kind of coordination or cooperation among the players."[17] By implication, it leads us into a worst case analysis that "tends to exaggerate any contributions regimes make to solving environmental problems."[18]

Another and somewhat related problem is, according to Young, that conceptualizing the no-regime outcome in terms of interactive decision-making among self-interested states leads us to focus only on policy games themselves and to neglect a number of other driving forces—demographic, economic and technological—that interact to produce important (environmental) impacts:

> Although large-scale environmental problems certainly do engender interactive decisionmaking among identifiable actors, the resultant processes constitute only one among a number of driving forces—including a range of demographic, economic, political, and technological forces—that interact with one another in complex ways, producing far-reaching impacts on problems quite apart from the dynamics of interactive decisionmaking. Unless we make the (to my mind far-fetched) assumption that the outcomes represented in a game-theoretic payoff space and the values the players assign to them somehow take all these forces (including interactions among them over time) into account, the Nash equilibrium—or, for that matter, any other game-theoretic construct—has little to recommend it as a way of representing the likely outcome in the event that the parties fail to reach agreement on a cooperative solution.[19]

Young also criticizes the notion of a "collective optimum," pointing out that it is "an essentially contested concept."[20] He seems to have two main concerns. One is that it is hard to find a common basis for making judgments of what is "optimal." Should we follow the utilitarian mode of reasoning applied in economics and assume that the optimal level of pollution abatement lies at the point where the marginal social benefits associated with the last unit of abatement just equals the marginal social cost of achieving that unit? Would some notion of critical thresholds be more appropriate?[21] His other concern is that the concept of "collective optimum," as defined in the Oslo-Potsdam solu-

15. Helm and Sprinz 2000.
16. Young 2001, 110–112.
17. Young 2001, 111.
18. Young 2001, 111.
19. Young 2001, 111.
20. Young 2001, 112–113.
21. Young 2001, 113.

tion, captures only a narrow domain of regime consequences. "It leads us to neglect consequences that are defined in other currencies, such as the distributive effects of regimes . . . or the side effects of regimes."[22] Unless we find some way of including such broader consequences into the analysis, we might end up granting a high score of effectiveness to a regime whose net social benefit is negative. Young finds this prospect "disturbing."

3.2 Empirical Issues

Young also raises several questions pertaining to the operational procedures for translating the formula into valid empirical scores. As described in section 2, the two research teams have applied somewhat different procedures for assigning scores, so some of his remarks will be more relevant to one of the approaches than to the other.

First, with regard to the no-regime counterfactual, he reminds us that counterfactual analysis is fraught with difficulties and pitfalls. Can we really rely on experts or technical databases to provide useful estimates of marginal damage and abatement costs? Are we not being led into a static mode of reasoning, leaving out—or forcing us to make arbitrary assumptions about—important factors such as improved scientific understanding of cause-effect relationships, technological change, and the like? This all implies, he says, that "specific procedures for measuring the no-regime outcome seem just as vulnerable on empirical grounds as they do analytically."[23] Second, with regard to the measurement of actual regime performance, he points out that the approach adopted tends to focus only on manifest behavior (such as emission reductions), leaving out important effects such as social learning and improved understanding of the nature of the problem.[24] He argues that by relying on these procedures the Oslo-Potsdam solution runs a serious risk of *undervaluing* the overall effects of regimes. Finally, our idea of what qualifies as the collective optimum will depend on our understanding of the nature of the problem and of the options available to us.

4. Responding to the Critique

We now turn to our response to the critique that was outlined in the previous section. Again, we find it convenient to consider separately the conceptual and empirical issues. In evaluating the critique, we consider two questions. The first is whether or not a particular objection is valid. The second is what conclusions should be drawn, given that a certain objection is deemed valid. There are four main options. First, one might give up entirely on the proposed venture, i.e.,

22. Young 2001, 113–114.
23. Young 2001, 112.
24. This argument seems to pertain only to such cognitive effects that do *not* lead to a change in the manifest behavior observed.

conclude that the goal of constructing an operational measure of regime effectiveness simply cannot be achieved. Second, one might see the goal itself as important and yet conclude that an entirely different approach is needed to accomplish it. Third, one might continue on the same basic course with some revision. Finally, one might conclude that even though the approach has significant shortcomings, it is—at least for some purposes—the best option currently available.

4.1 Conceptual Challenges

To recapitulate, Young raised two basic conceptual issues, one pertaining to the no-regime counterfactual, the other to the collective optimum. We begin with the former.

A cornerstone in Young's critique of the Oslo-Potsdam solution is his skepticism to the Nash equilibrium as a basis for estimating the no-regime counterfactual. This skepticism is based on three objections. We argue that the first objection is valid, the second is overly general, while the third is valid yet offers little guidance for further research. On the other hand, we believe that at least two more (valid) objections may be added to Young's list, and these will be discussed in Section 5.

Young's first objection is that in some cases there are multiple Nash equilibria.[25] There is no question that this is a valid objection. In fact, there are sometimes infinitely many Nash equilibria. Obviously, this raises the question of which equilibrium is likely to be played. The issue goes back to classics like Luce and Raiffa[26] and Schelling,[27] who draw a distinction between games that have a "solution in the strict sense" and those that do not. While originally the distinction was made with reference to static games, in the literature on dynamic games the problem of multiple equilibria has motivated a number of refinements of the Nash equilibrium. Among the most important of these refinements are the subgame perfect equilibrium and the renegotiation proof equilibrium.[28] In some cases, such refinements solve the problem. For example, in Rubinstein's model of sequential bargaining there is a continuum of Nash equilibria, and yet there is a unique subgame perfect equilibrium. In other cases, however, there exists a large number of subgame perfect or renegotiation proof equilibria as well. Hence, Young's point that sometimes there are multiple

25. Readers that would like to update themselves on the basic concepts of game theory might consult Gates and Humes 1997; Hovi 1998; or Morrow 1994.
26. Luce and Raiffa 1957.
27. Schelling 1960.
28. A Nash equilibrium is a set of strategies (one for each player) that are best responses to each other. A subgame perfect equilibrium is a set of strategies that are best responses to each other *for all subgames* (i.e., for all parts of the game). Subgame perfection rules out Nash equilibria that are sustained by an empty threat (and therefore implausible). A (subgame perfect) equilibrium is renegotiation proof if it is rational for all players to resist an invitation to renegotiate—say, after an agreement has been violated. Renegotiation proofness rules out equilibria that are not collectively rational to implement.

equilibria, and that game theory is not always able to tell us which equilibrium will be played, remains valid.

The second objection is that the Nash equilibrium "assumes that no coordination takes place" between the players. Thus (according to Young) it uses a worst-case analysis to identify the no-regime counterfactual, which implies an overestimation of the effect of international regimes. At first glance, the latter point of view might seem to be supported by the experimental literature on the Prisoner's Dilemma and related games. This literature generally reports levels of cooperation that are considerably higher than predicted by game theory for situations of this kind.[29]

However, these predictions are generally derived under the presupposition that the subjects are exclusively motivated by material self-interest, usually (small) monetary payoffs. Experimental findings of more cooperation than predicted therefore do not prove that the subjects are not playing a Nash equilibrium. An alternative explanation is that at least some subjects are not exclusively concerned with material self-interest. If this is the case, then (true) Nash play would involve more cooperation than suggested by the Nash equilibrium in the experimenter's model. In short, this part of Young's critique is valid only if we assume that decision makers are motivated exclusively by material self-interest. In fact, game theory doesn't need to assume this.

But even if we were to accept Young's view of game theory's assumptions, his claim that the Nash equilibrium assumes no coordination would still be an exaggeration. In the absence of a particular regime there may be multiple equilibria—some of which are likely to involve more cooperative behavior than others. It is far from clear that, as a rule, the most non-cooperative of these equilibria is the one that is most likely to be played, and thus the prominent candidate for the no-regime counterfactual.[30] Also, Young's claim seems to imply that Nash equilibria will not be played in the presence of coordination. By contrast, in the game-theoretic literature, the general opinion is now almost the exact opposite: In general, one cannot expect rational players to reach a Nash equilibrium *without* some kind of coordination, such as an announced proposal for play.[31] Furthermore, in some models (such as models of sequential bargaining), the actions available to the players are explicit proposals and counterproposals.

29. To illustrate, consider the following public goods experiment. N subjects are given a sum of money (e.g., $5). The money can either be kept, or some or all of the money can be invested in a public good. Money invested is multiplied by some factor k (larger than 1, but smaller than N). The money invested, plus returns, is divided equally among all subjects. Since $k>1$, the overall sum of money is increased by the investment. However, since $k<N$, each subject's share of the returns is smaller than the amount invested. Game theory therefore predicts zero cooperation both in the one-shot and in the finitely repeated version of this game. Still, experiments indicate that, on average, subjects invest 40–60 per cent of the money in one-shot trials. In multiple trials, contributions average around 50 percent on the first trial, and then drop sharply, typically reaching 15–20 percent after 5 trials. Note, however, that average contributions never even come close to zero (Dawes and Thaler 1988, 188–190).
30. We are indebted to one of the anonymous reviewers for helping us to see this point more clearly.
31. See, for example, Weibull 2002.

In a model of this kind it is not even clear what "no coordination" means, since every possible outcome—including any equilibrium—involves communication (and hence coordination) between the players. At best, therefore, Young's claim is overly general.

Third, Young argues forcefully that game-theoretic modeling falls well short of incorporating all factors relevant to large-scale environmental problems. We believe there is much to be said in favor of this point of view. Ideally, to estimate the no-regime counterfactual, we would want a causal model that includes all relevant factors. Clearly, no game-theoretic construct is able to accomplish this. That said, one should recognize that this criticism can be extended to virtually all theories in political science. Any theory or model in political science must necessarily be partial in the sense that it directs attention to particular aspects of a real-world phenomenon, while leaving other aspects out. Thus, it is true—but hardly surprising—that no game-theoretic model is able to conform to the ideal that all relevant factors should be incorporated in the payoffs. However, as Jon Elster[32] reminds us, "you can't beat something with nothing." Because no other theory meets Young's standards either, his critique does not necessarily rule out the Nash equilibrium as a way of conceptualizing the no-regime counterfactual. The most constructive criticism would be to come up with an alternative that is demonstrably superior. As this is something that Young does not do in his article, we can only invite such contributions in the future.

Given Young's critique, one might ask how game-theoretic constructs can be a useful device for estimating the no-regime counterfactual. To answer this question, it is helpful to make a distinction between *point predictions* on the one hand and *directional predictions* on the other. A point prediction aims to determine the exact value of a certain (dependent) variable, given particular values on a set of other (independent) variables. By contrast, a directional prediction merely suggests the direction of change in a dependent variable that is likely to be caused by a particular change in some independent variable. As Fiorina[33] reminds us, most social science predictions (rational choice or otherwise) are made with a *ceteris paribus* clause. Only if all relevant (economic, political, institutional, demographic, technological, normative, cultural, and psychological) factors were fully incorporated into the model, could we hope to make an accurate point prediction. However, "since we can never know all of the relevant variables, let alone measure all of them accurately, the empirical predictions of all models in political science—not just R[ational] C[hoice]—are about relative differences and comparative statics."[34] What all this means is that it is considerably easier to provide (correct) directional predictions than it is to generate (correct) point predictions.

32. Elster 1986, 27.
33. Fiorina 1996.
34. Fiorina 1996, 88.

Consider a change of regime rules. In order to study the impact of such a change on regime effectiveness, we need estimates of effectiveness before and after the change. If relevant variables are left out, both estimates are likely to be biased. However, if the errors are likely to be approximately the same in both cases, we should still be able to predict correctly the *direction* of change.

Two simple examples may help to clarify the point. First, the noncooperative bargaining model proposed by Ariel Rubinstein in the early 1980s quickly became very popular in economics. More recently this success has spread to parts of political science as well. The popularity of the model defies the fact that experiments show that subjects typically settle for far more symmetrical agreements than predicted by the model, especially when the time available for bargaining is short. The popularity may, at least in part, be explained by the fact that the model has proved useful in generating interesting directional predictions about the impact of different institutional environments on bargaining outcomes. Examples include the effects of outside options,[35] domestic constraints due to ratification requirements,[36] and a need for decentralized enforcement that is characteristic of many international treaties.[37]

Second, the perhaps most widely used model from game theory is the Prisoners' Dilemma. As already mentioned, experiments consistently show that game theory's (point) predictions of behavior in this and related games substantially underrate the observed amount of cooperation. Again, however, this has not had much of an impact on the popularity of the model. The typically small payoffs of a game-theoretic experiment cannot be expected to override every conceivable other determinant of the relevant subjects' behavior. Therefore, the above results are neither very surprising nor particularly disturbing from the point of view of game theory. Nor should it trouble game theorists that other experiments find payoffs that game theory does not single out as relevant to be important determinants of the outcome of some games.[38] What really *would* be disturbing is if the level of cooperation or the outcome of bargaining processes could be shown to remain stable—or even to move in the wrong direction—as a result of change in a variable that is pointed out by game theory as highly relevant. Examples relating to cooperation include changing the time horizon of a public goods game from infinite to finite, changing the information structure of a finitely repeated game from complete information to the (right sort of) incomplete information, or allowing the subjects to make binding pre-play commitments. Examples relating to bargaining include changing the players' discount factors, changing the number of periods available for bargaining, or modifying institutional aspects of the kind mentioned above.

In summary, we agree with Young that it is often very hard to make accurate estimates of what would happen in the absence of a regime. However, the

35. Binmore, Shaked, and Sutton 1989.
36. Tarar 2001.
37. Hovi 2001.
38. Goeree and Holt 2001.

Figure 2
Summary of Conceptual Critique and Assessment

Reference Point in Oslo-Potsdam Solution	Critique	Assessment
No-regime counterfactual	Multiple Nash equilibria in some contexts	Valid
	Nash equilibrium assumes no coordination; implies worst-case analysis	Overly general
	Not all relevant factors incorporated	Valid, but would also be true of any manageable alternative
Collective optimum	Contested concept	True, yet alternative choices can also be implemented by way of the Oslo-Potsdam solution

ultimate aim of determining the no regime counterfactual will usually be to explain variance in regime effectiveness—either over time and/or across regimes. The good news from the above discussion is that imprecise point estimates of no-regime counterfactuals do not necessarily distort longitudinal or cross-sectional analyses of regime effectiveness.

We now turn to Young's second concern, namely that the collective optimum is an "essentially contested concept." This is true, on at least three levels. First, regime effectiveness is a normative concept, and there are clearly other normative standards by which regime performance may be evaluated. For example, Young mentions distribution, equity, economic efficiency and compliance.[39] Second, even when regimes are evaluated in terms of problem-solving effectiveness, other notions, including *ultimate* targets formulated in relevant international conventions, may be used instead of the collective optimum. Finally, the notion of collective optimum can itself be defined in different ways. Among the candidates are the solution that maximizes net long-term economic gain, the best solution that is politically feasible, and the solution that maximizes sustainable biological yield.

Any evaluation of regime performance builds on some normative concept. We are not in the business of authorizing one of these concepts as *the* standard for the field of regime studies. Although formulated with reference to the concept of effectiveness, the basic logic of the Oslo-Potsdam solution can also be applied to the other normative standards introduced above.

39. Young 1999, 5–9.

Figure 2 summarizes the discussion in this section, by listing in condensed form our response to the various points in Young's conceptual critique.

4.2 Empirical Challenges

Having responded to the part of Young's critique that addresses conceptual issues, we now turn to empirical challenges to the Oslo-Potsdam solution. It may be noted that some of these challenges flow from the theoretical objections. As we shall see, however, empirical measurement does not require us to solve all theoretical problems beforehand.

With respect to the no-regime counterfactual, Young asserts that no prior level of cooperation among countries is taken into account and that the procedure used by Helm and Sprinz[40] tends to exaggerate the contribution of regimes. While we have discussed the implications for using the Nash-based no-regime counterfactual above, and make further suggestions below, we would like to point out that the critique overlooks at least two points. First, the empirical solution to the problem rests on expert advice that delivered no- regime counterfactual emissions. This approach used an identical target year, and the experts had knowledge of what the level of inter-state collaboration was before the regime was formed. The experts were asked to make reasonable assumptions about the evolution in the period since the inception of the international regime. Therefore, it is not true that the empirical solution assumes no prior collaboration among countries. Second, focusing solely on deriving NR cannot permit inferences regarding potentially biased results of the overall measure of regime effect as all three components are necessary to make such inferences. This applies, more generally, to potential bias induced by any of the three quantities used in computing the Oslo-Potsdam solution.

The effect of changes in the numerical values of the three quantities can easily be shown using sensitivity scores. In Figure 1, we show how a one-unit increase in the estimate of the regime's actual performance will affect the magnitude of the effectiveness score.[41] Similar scores can be calculated for the other two elements in the formula, and should provide scholars and policy-makers with a good sense of how robust effectiveness scores are.

The critique of the choice of empirically deriving the actual performance (AP) and collective optimum (CO) values can be taken together. We have chosen to focus on (environmental) problem solving, following Lave and March's advice to keep things simple.[42] Judging from his writing, Young[43] clearly prefers a much wider conception that includes many other factors. This would either leave us with the undertaking of determining upper bounds across multiple dimensions, or the task of constructing an index that aggregates these dimensions into a single measure. While determining the collective optimum is certainly a

40. Helm and Sprinz 2000.
41. See Helm and Sprinz 2000, 642–643, for an empirical example.
42. Lave and March 1975.
43. Young 2001.

Figure 3
Summary of Empirical Critique and Assessment

Reference Point in Oslo-Potsdam Solution	Critique	Assessment
No-regime counterfactual	No cooperation among countries	Not valid for actual measurement procedure involving case experts
Actual Performance & Collective Optimum	Scope of factors included in actual performance and collective optimum is narrow, i.e., focused on (environmental) problem-solving	True, and advisable at early stage of research
All three quantities (NR, AP, CO)	Directional bias in effectiveness score	No directional bias known with respect to (environmental) problem-solving; sensitivity coefficients provide transparent information for the effect of changes of each of the three quantities

demanding task, neither of these two alternatives is likely to provide an easy way out. Lave and March's strategy guards against overly high complexity and affords critics much more transparent ways to replicate and evaluate the findings. We do, however, agree with Young that after refining our measurement tools further, including additional aspects of international regimes into the measurement procedure should be seriously considered.

5. Advancing Alternative Solutions: Our Perspective for Future Research

As we have seen, Young's article provides a number of critical comments on the Oslo-Potsdam solution. Many of these comments are well taken and considerably improve our understanding of the problems involved in measuring regime effectiveness. In this section, we use Young's work as a stepping stone for addressing some additional conceptual and empirical issues to further develop and refine the Oslo-Potsdam solution.

5.1 Conceptual Advances

Regarding the no-regime counterfactual, use of the Nash equilibrium raises at least two conceptual problems that are not mentioned by Young. The first is that the Nash equilibrium is not always robust against small changes in the assumptions about the model's parameters. In fact, even minor changes in the assumptions can sometimes make a substantial difference to the conclusions. In a one-shot game we may arrive at one conclusion if the parties are assumed to move simultaneously, a second conclusion if one player is allowed to move ahead of the other, and a third conclusion with the reverse sequence of moves.[44] In a repeated game of complete information, the set of equilibria may be largely independent of the number of iterations as long as the horizon is (known to be) finite, but change dramatically if the horizon is modified from finite (but very long) to infinite.[45] And introducing even a very small amount of (the right kind of) incomplete information can sometimes produce behavioral predictions that are very different from those obtained in a complete information version of the same model.[46] Even if we could somehow accomplish the impossible task of designing a model that included all relevant factors, therefore, our point prediction of the no-regime counterfactual might still be grossly mistaken unless the chosen parameter values were (exactly) correct. Clearly, this serves to further underline our previous point that it is usually advisable to concentrate on directional rather than point predictions.

The second point is whether the strong focus on the Nash equilibrium is well founded. By now there seems to be a near consensus among applied game theorists that the relevant non-cooperative equilibrium concepts for various types of games are as shown in Figure 4.[47]

If we accept the categorization shown in Figure 4, the notion that the no-regime counterfactual should be measured by way of the Nash equilibrium would seem to imply that international cooperation is (always) best modeled as a static game of complete information.[48] This does not seem to fit very well with the standard view of international interaction as repeated and often heavily infused with uncertainty. A more promising idea might be to model each case of cooperation on its own merits, and then proceed to use whichever equilibrium concept is most appropriate, given the specifics of the model at hand. One possibility would then be to compare the relevant equilibrium with some notion of the collective optimum.

We now turn to the collective optimum. How this optimum can be measured empirically is considered further below. Here, we further elaborate on

44. For example, this is true in a Chicken game (see, for example, Fink, Gates, and Humes 1998).
45. For example, this holds for the repeated Prisoners' Dilemma as well as for the repeated Chainstore game (see, for example, Fudenberg and Tirole 1991).
46. This is true in the finitely repeated Prisoners' Dilemma as well as in the finitely repeated Chainstore game (Kreps and Wilson 1982; Kreps et al. 1982; and Milgrom and Roberts 1982).
47. A Bayesian Nash equilibrium is a set of strategies that are best responses to each other, *given the players' beliefs*. A perfect Bayesian equilibrium satisfies two requirements. First, the players' strategies are best responses to each other for all subgames, given the players' beliefs. Second, during the course of play, beliefs are updated by using Bayes' rule (wherever possible) on the assumption of equilibrium play. The other equilibrium concepts listed in Figure 4 have been explained earlier. Note that the other types of equilibria mentioned in Figure 4 are also Nash equilibria.
48. Complete information means that the players' strategy sets and utility functions are common knowledge. A game is static if no player is able to observe and react to the action of another player (Fudenberg and Tirole 1991). Thus, a static game is a one-shot game where the players move simultaneously.

Figure 4
Types of Games with Corresponding Equilibrium Concepts

	Complete Information	**Incomplete Information**
Static Games	**Nash equilibrium**	**Bayesian Nash equilibrium**
Dynamic Games	**Sub-game perfect equilibrium/Renegotiation proof equilibrium**	**Perfect Bayesian equilibrium**

possible conceptualizations of such an optimum. The literature offers a number of different options. A first possibility is to use some technical notion. For example, in the literature on international fisheries management, the concept of maximum sustainable yield (in biological terms) has been popular, although economists tend to see the related concept of maximum economic yield as a more relevant target. Second, the collective optimum might be conceived of as a Pareto-optimal outcome. This is least problematic in cases where a unique Pareto optimum exists. For example, in the experimental public good game referred to above (see footnote 29), the unique Pareto optimal outcome is that all players invest the entire amount in the public good (assuming that side payments are allowed). However, in many cases we find several Pareto optimal outcomes. Additional criteria are then needed to single out a unique candidate for the collective optimum. One place to look for such criteria is in cooperative game theory. For example, the Nash bargaining solution reaches a unique cooperative outcome on the basis of four axioms, namely (i) Pareto optimality, (ii) invariance to positive linear transformations of the utility functions, (iii) symmetry, and (iv) independence of irrelevant alternatives.[49] Similarly, the Kalai-Smorodinsky[50] bargaining solution obtains uniqueness by applying a slightly different set of axioms.[51]

Using one of these two bargaining solutions implies comparing a non-cooperative equilibrium (Nash or otherwise) with a solution derived from cooperative game theory. A very different approach to measuring regime effectiveness would be to compare the non-cooperative equilibria of two different versions of the same model. For example, one might first compute the non-cooperative equilibrium without any regime at all (the no-regime counter-

49. Nash 1950.
50. Kalai and Smorodinsky 1975.
51. Specifically, the axiom of independence of irrelevant alternatives is replaced by one of monotonicity, which states that if the maximum utility level for a player is larger in bargaining problem I than in bargaining problem II, then the solution of bargaining problem I should not give this player a smaller utility than the solution of bargaining problem II.

factual). Then one might proceed to compute the non-cooperative equilibrium with a specific institutional structure—an "ideal regime"—in place (the collective optimum). The underlying idea of the latter approach would be that even a cooperative outcome must be sustained as a (non-cooperative) equilibrium.

To summarize, using game-theoretic constructs to model the no-regime counterfactual and the collective optimum raises a number of questions. Solving these issues is probably vital for deductive reasoning to have a dominant place in future research on regime effectiveness. However, we do not believe that it is necessary to solve all the conceptual issues before making empirical assessments. While the empirical solutions offered by Helm and Sprinz[52] and Miles et al.[53] draw on the conceptual solution, they could be undertaken even without completely resolving all conceptual issues. We will now consider some of the available options for the empirical stage.

5.2 Empirical Advances

As we have seen, estimating or determining the no-regime counterfactual is a difficult challenge. Ideally, what we would need is a comprehensive model incorporating all relevant factors. In the absence of that, we have to rely on some second-best solution. Among the options are the following:

- use of expert review teams;
- simulation;
- political cost-benefit analysis;
- use of the status quo ante; and
- statistical effect separation.

We will briefly introduce each of these options below.

The first option, using *expert review teams,* is the one that has been most often used to establish the no-regime counterfactual. Both Helm and Sprinz and Miles et al. used research teams to formulate explicit coding manuals that were distributed to specialists in the relevant issue area. The specialists were then instructed how to undertake the scoring on either an ordinal or an interval scale. This involved carefully phrased ex post questions for a no-regime counterfactual by way of a thought experiment. This thought experiment was undertaken by a range of issue specialists within a country, including distribution requirements regarding the profile of such specialists[54] or a structured flow diagram of which advice to follow, starting with the most reliable external assessment and potentially ending with the best assessment by the academic case expert.[55] If the scores of the values of the no-regime counterfactual differed, averaging techniques were used both within the country study as well as between the country

52. Helm and Sprinz 2000.
53. Miles et al. 2002.
54. Helm and Sprinz 2000.
55. Miles et al. 2002.

report and an independent cross-country coder[56] or were assisted by inter-case calibration techniques.[57]

A second option is *simulating* the counterfactual, either ex ante or ex post—congruent with different states of knowledge on the subject matter. For example, it is possible to use a political decision-making model[58] to derive political decisions within a polity without the benefit of an existing international regime. This procedure would necessitate running simulations with data inputs from the beginning of the time period under investigation in "ex ante" mode. Similar in spirit is *political cost-benefit analysis*—that is, deriving the equilibrium between marginal economic abatement and damage cost functions which are often used in environmental economics. If only national abatement and damage cost functions are used in the calculation, then the intersection of both functions would constitute an optimal (counterfactual) policy that does not take into account damages caused elsewhere by a nation's polluting or extracting activities. While it is comparatively easy to derive abatement cost functions, it is quite often difficult to obtain economic or political damage cost functions, although the latter have been indirectly derived by some researchers.[59] More research will certainly be necessary until we know more about the functional form of such damage costs functions.

Young himself has suggested a procedure using the *status quo ante* as a baseline for estimating regime effectiveness.[60] Assuming that we know the actual performance, we can move on to computing the difference between actual performance and the status quo ante. One can then examine the causal mechanisms through which the regime could have produced the estimated change and, on this basis, try to determine how much of the difference can plausibly be attributed to the regime. This approach is commonly adopted in empirical case studies,[61] which often rely on the kind of thought experiment characteristic of historical analysis. Although usually conducted in non-formal qualitative terms, this approach resembles simulation in its basic logic. Without a carefully crafted thought experiment, use of the status quo ante as a baseline clearly omits the introduction of a control group in quasi-experimental designs.[62] If changes occur,

56. Helm and Sprinz 2000.
57. Miles et al. 2002. Past research relied completely on point estimates for the no-regime counterfactual. Yet it might be more interesting to learn about the distributions of such estimates. Assuming that such distributions are unbiased, researchers would prefer efficient estimates (small rather than wide distributions) to establish the regime effectiveness score with high confidence. Using alternative values within a distribution (mean, median, modal value or the values for the 5th or 95th percentile) would inform us much more comprehensively than the present sensitivity score of regime effectiveness (see Figure 1). Using the inputs from expert review teams could be generalized to systematically generate distributions of responses to multiple phrasings of the no-regime counterfactual and subsequent decision whether to utilize the full or partial range of such values in subsequent scoring procedures and whether to capitalize on the moments of the distribution.
58. For example, Bueno de Mesquita and Stokman 1994; and Bueno de Mesquita 1994 and 2003.
59. Pastor and Wise 1994; and Helm and Sprinz 2000.
60. Young 2001, 112.
61. Bennett forthcoming; and Wettestad 1999, 39.
62. Cook and Campbell 1979.

the status quo ante does *not* give us the proper counterfactual because it fails to take into account what would have happened in the meantime.[63] Young himself is quite aware of this limitation. Yet its ready availability is unfortunately enticing to many social scientists and the public media.

Projections based on *statistical effect separation* may prove useful. The counterfactual value would here be based on a comprehensive explanatory statistical model of the positioning on the assessment scale. In essence, we would ask which factors explain the level of performance at the relevant point in time with all exogenous variables being selected by a theory-driven explanation. The statistical results would be generated by cross-sectional analysis or a pooled cross-sectional analysis, the latter including time-series measures of the variables included in the cross-sectional analysis. Using the average variable scores while setting the value of variables representing regime functions at zero, we would be able to derive the no-regime counterfactual statistically.[64] This approach is particularly appealing in cases of models with high explanatory power.

One particular route of statistically separating the effect of international regimes has been suggested by Mitchell.[65] He suggests replacing the systemic level of analysis with the analysis of country-level performance variables at discrete time steps, e.g., yearly data. In particular, Mitchell suggests deriving the dependent variable of country-specific "regime effort units" as the product of per annum percentage change of emissions multiplied by per unit effort (in $ or . .).. Subsequently, variation in regime effort units could be explained by variables representing regime properties (e.g., sanctioning systems and other variables) as well as other control variables (see also above). This procedure turns the regression *coefficients* associated with regime variables into measures of regime effectiveness with the dependent variable reflecting the per annum costs that a country is willing to shoulder. Given its focus on explaining cost outlays, this procedure would turn research on regime effectiveness away from its dominant focus on problem-solving, yet it is also obvious that the dependent variable chosen should induce problem-solving.

In conclusion, counterfactuals are clearly not easy to derive. Tetlock and Belkin[66] suggest the following standards for the ideal counterfactual: (i) clarity; (ii) logical consistency or contestability; (iii) historical consistency (minimal-rewrite rule); (iv) theoretical consistency; (v) statistical consistency; and (vi) projectability. While it will be difficult to honor all six criteria simultaneously, we generally suggest emphasizing the first three criteria, especially the minimal-rewrite rule.

63. As the period of time to be assessed approximates zero, the use of the status quo ante as the no-regime counterfactual will become more appropriate. At the same time, we should not expect to find much regime effect during any infinitely small amount of time.
64. As mentioned above, the challenge of establishing a credible counterfactual rests on establishing the scope of variables being affected.
65. Mitchell 2002.
66. Tetlock and Belkin 1996, 18.

To establish the collective optimum, we initially have to decide which normative standard to apply (see section 4.1). Once that standard has been specified, one can then move on to determine the upper bound, i.e., the "best" solution. As in the case of the no-regime counterfactual, we are most often left with a menu of second-best options. Given that our standard is regime effectiveness, these options include:

- regime goals;
- (technical) functional optima;
- cost minima;
- environmental thresholds; and
- results from stakeholder dialogues.

International regimes often espouse specific *goals,* such as a symmetric or country-specific scale of emission reductions that an array of countries shall undertake to be in compliance with an international treaty. While using the numerical expression of the regime goal as a representation of a collective optimum may be attractive, it is fraught with serious methodological problems. In particular, using official regime goals as collective optima confuses regime compliance with regime effectiveness.[67] Furthermore, as (most) international environmental treaties are negotiated documents, countries do normally not accept obligations they are unable or unwilling to meet. Thus, negotiated treaties are often a classical example of selection bias[68] or an endogeneity problem where both the capacity and willingness to comply influences which obligations governments accept.[69] Furthermore, compliance approaches would omit countries without obligations (non-signatories) even if they contribute to the creation of the environmental problem and potentially undertake some efforts at ameliorating it.

As described in Section 2, Miles et al.[70] derive *functional (technical) collective optima,* in particular bio-physical optima derived from natural science expertise or expert advice. In contrast, Helm and Sprinz[71] use a collective (economic) abatement *cost minimum* for the set of all countries; in other words, to achieve the goal of environmental problem solving, an algorithm is used that minimizes the costs for all countries involved. This approach, however, overlooks any notion of equity because countries are unlikely to contribute to the regime's goal independent of other countries' contributions.[72] In particular, side-payments provide powerful incentives to actually use the cost-saving potentials in countries with low costs of abatement.

67. More precisely, this applies to shorter-term and intermediate goals. "Ultimate goals," which constitute a proper choice, would certainly be permissible.
68. For an overview of selection bias in international relations methods, see Sprinz and Wolinsky forthcoming, ch. 15.
69. Bernauer 1995; and Helm and Sprinz 2000.
70. Miles et al. 2002.
71. Helm and Sprinz 2000.
72. Ward, Grundig, and Zorick 2001.

Using *environmental thresholds* may offer the best representation of environmental problem-solving. In essence, environmental thresholds point to strong discontinuities in dose-response (or pollution-impact) relationships and thus represent the logical upper bound to avoid environmental harm. While this approach is theoretically attractive, there is no unified methodology to derive thresholds as most natural scientists rely on visual inspection techniques. Some hope may be vested in deriving statistical criteria for the diagnosis of such thresholds.[73]

As many of these options may pose their own challenges or force the researcher to make (normatively-based) decisions, it might be advisable to use a *stakeholder process* to derive collective optima. Stakeholders or organized interest groups in a particular issue domain normally hold positions that can be discussed in a roundtable-style setting. It is quite conceivable that such stakeholders (scientists, politicians, business and environmental NGOs, etc.) may come up with a commonly agreed upon upper performance benchmark—yet recommend completely different policies to be pursued in the foreseeable future. For example, some industry representatives might agree that 50–70% emission reductions of greenhouse gases is needed to avoid "dangerous interference with the climate system"—and yet see such emission reductions as not being worth the economic and political costs of restructuring the energy sector.

Given the absence of ideal solutions, we are left with a menu of second-best options. At this point in time, we have no basis for recommending one option as superior to all others. Moreover, it remains an open question to what extent the estimates produced by different methods coincide. In this situation, we recommend more systematic efforts to compare results produced by the various options. By holding the issue domains constant, researchers using different methodological orientations should be brought together to find numerical calibrations for the lower and upper benchmarks of regime performance. If their results agree reasonably well, we could infer that the choice of method does not strongly influence the findings. Conversely, if the numerical values deviate strongly from each other, we should explore systematically why they are different, whether the differences are methods-related, and, if so, what methods are most reliable.

6. Conclusions

This article has offered a review and a discussion of the recent critique that has been voiced against the Oslo-Potsdam solution to measuring regime effectiveness. Our response to the critique may be summarized as follows. We are aware that complicated problems remain unsolved. In particular, there are conceptual and analytical difficulties related to the identification of both the no-regime counterfactual and the collective optimum. Furthermore, we have taken a few

73. Sprinz and Churkina 1999.

additional steps to refine the Oslo-Potsdam solution. We recognize that other approaches can be developed, and we would welcome similar efforts to specify the conceptual framework and operational measurement procedures underlying these approaches. At this point in time, we think it is an open question which approach will prove most fruitful in the long run. However, we are optimistic that the Oslo-Potsdam solution has a sound conceptual foundation, and that it can be further refined and extended. We look forward to continue to do so in a productive dialogue with our critics.

References

Bennett, Andrew. Forthcoming. Case Study Methods: Design, Use, and Comparative Advantages. In *Models, Numbers, and Cases: Methods for Studying International Relations*, edited by Detlef F. Sprinz and Yael N. Wolinsky-Nahmias. Ann Arbor, MI: The University of Michigan Press.

Bernauer, Thomas. 1995. The Effect of International Environmental Institutions: How We Might Learn More. *International Organization* 49 (2): 351–77.

Binmore, Ken, Avner Shaked, and John Sutton. 1989. An Outside Option Experiment. *Quarterly Journal of Economics* 104: 753–770.

Bueno de Mesquita, Bruce. 1994. Political Forecasting: An Expected Utility Method. In *European Community Decision-Making: Models, Applications, and Comparisons*, edited by Bruce Bueno de Mesquita and Frans N. Stokman, 71–104. New Haven: Yale University Press.

_______. 2003. *Principles of International Politics.* Second Edition. Washington, D.C.: CQ Press.

Bueno de Mesquita, Bruce, and Frans N. Stokman, eds. 1994. *European Community Decision-Making: Models, Applications, and Comparisons.* New Haven: Yale University Press.

Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimentation: Design & Analysis Issues for Field Settings.* Boston, MA: Houghton Mifflin Company.

Dawes, Robyn. M., and Richard H. Thaler. 1988. Anomalies: Cooperation. *Journal of Economic Perspectives* 2 (3): 187–197.

Elster, Jon. 1986. Introduction. In *Rational Choice*, edited by J. Elster, 1–33. Oxford: Blackwell.

Fink, Evelyn C., Scott Gates, and Brian D. Humes. 1998. *Game Theory Topics: Incomplete Information, Repeated Games, and N-Player Games.* Thousand Oaks, California: Sage.

Fiorina, Morris. 1996. Rational Choice, Empirical Contributions, and the Scientific Enterprise. In *The Rational Choice Controversy*, edited by Jeffrey Friedman, 85–94. New Haven: Yale University Press.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory.* Cambridge, Massachusetts: MIT Press.

Gates, Scott, and Brian D. Humes. 1997. *Games, Information, and Politics. Applying Game Theoretic Models to Political Science.* Ann Arbor: University of Michigan Press.

Goeree, Jacob K., and Charles A. Holt. 2001. Ten Little Treasures of Game Theory and Ten Intuitive Contradictions. *American Economic Review* 91 (5): 1402–1422.

Helm, Carsten, and Detlef F. Sprinz. 2000. Measuring the Effectiveness of International Environmental Regimes. *Journal of Conflict Resolution* 45 (5): 630–652.

Hovi, Jon. 1998. *Games, Threats and Treaties. Understanding Commitments in International Relations.* London: Pinter.

_______. 2001. Decentralized Enforcement, Sequential Bargaining, and the Clean Development Mechanism. *Nordic Journal of Political Economy* 27 (2): 135–152.

Kalai, Ehud, and Rann Smorodinsky. 1975. Other Solutions to Nash's Bargaining Problems. *Econometrica* 43 (3): 513–518.

Kreps, David, and Robert Wilson. 1982. Reputation and Imperfect Information. *Journal of Economic Theory* 27 (2): 253–279.

Kreps, David, Paul Milgrom, John Roberts, and Robert Wilson. 1982. Rational Cooperation in the Finitely Repeated Prisoners' Dilemma. *Journal of Economic Theory* 27 (2): 245–252.

Lave, Charles A., and James G. March. 1975. *An Introduction to Models in the Social Sciences.* New York, N.Y.: Harper & Row.

Luce, Duncan, and Howard Raiffa. 1957. *Games and Decisions.* New York: Wiley.

Miles, Edward L., Arild Underdal, Steinar Andresen, Jørgen Wettestad, Jon Birger Skjærseth, and Elaine M. Carlin. 2002. *Environmental Regime Effectiveness: Confronting Theory with Evidence.* Cambridge, MA: The MIT Press.

Milgrom, Paul, and John Roberts. 1982. Predation, Reputation, and Entry Deterrence. *Journal of Economic Theory* 27 (2): 280–312.

Mitchell, Ronald B. 2002. A Quantitative Approach to Evaluating International Environmental Regimes. *Global Environmental Politics* 2 (4): 58–83.

Morrow, James D. 1994. *Game Theory for Political Scientists.* Princeton, NJ: Princeton University Press.

Nash, John. 1950. The Bargaining Problem. *Econometrica* 18 (2): 155–162.

Pastor, Manuel, and Carol Wise. 1994. The Origins and Sustainability of Mexico's Free Trade Policy. *International Organization* 48 (3): 459–489.

Schelling, Thomas. 1960. *The Strategy of Conflict.* Cambridge, MA: Harvard University Press.

Sprinz, Detlef F., and Carsten Helm. 1999. The Effect of Global Environmental Regimes: A Measurement Concept. *International Political Science Review* 20 (4): 359–369.

Sprinz, Detlef F., and Galina E. Churkina. 1999. *The Analysis of Environmental Thresholds.* NATO Advanced Research Workshop "Caspian Sea: A Quest for Environmental Security." Venice International University, Venice, Italy, mimeo.

Sprinz, Detlef F., and Yael N. Wolinsky-Nahmias, eds. Forthcoming. *Models, Numbers, and Cases: Methods for Studying International Relations.* Ann Arbor, MI: The University of Michigan Press.

Tarar, Ahmer. 2001. International Bargaining with Two-Sided Domestic Constraints. *Journal of Conflict Resolution* 45 (3): 320–340.

Tetlock, Philip E., and Aaron Belkin. 1996. Counterfactual Thought Experiments in World Politics—Logical, Methodological, and Psychological Perspectives. In *Counterfactual Thought Experiments in World Politics—Logical, Methodological, and Psychological Perspectives*, edited by Philip E. Tetlock and Aaron Belkin, 1–38. Princeton, N.J.: Princeton University Press.

Underdal, Arild. 1992. The Concept of Regime 'Effectiveness'. *Cooperation and Conflict* 27 (3): 227–240.

Ward, Hugh, Frank Grundig, and Ethan Zorick. 2001. *Formal Theory and Regime Effectiveness: Rational Players, Irrational Regimes.* Paper Presented to the 2001 Annual Meeting of the International Studies Association. The Chicago Hilton, 20–24 February, Chicago, IL.

Weibull, Jörgen. 2002. *Testing Game Theory.* Stockholm School of Economics: Working Paper 382. Available at http://netec.wustl.edu/WoPEc/data/Papers/hhshastef0382.html

Wettestad, Jørgen. 1999. *The 1999 Concerted Action Workshop on the Effectiveness of Global and Regional Environmental Agreements: EU Concerted Action Project on the Effectiveness of International Environmental Agreement.* Lysaker, Norway: The Fridtjof Nansen Institute.

Young, Oran R. 1999. *The Effectiveness of International Environmental Regimes: The Causal Connections and Behavioral Mechanisms.* Cambridge, Mass.: MIT Press.

_______. 2001. Inferences and Indices: Evaluating the Effectiveness of International Environmental Regimes. *Global Environmental Politics* 1 (1): 99–121.