**LAW IN THE SERVICE OF POLITICS: MOVING NEO-LIBERAL
INSTITUTIONALISM
FROM METAPHOR TO THEORY BY USING THE INTERNATIONAL TREATY
PROCESS TO DEFINE
"ITERATION"**

John K. Setear [*]

For most scholars, international law is to law as California is to the United States: sprawling and exotic enough to generate some inherent interest, growing too fast to ignore, and a little too flaky for comfort. International law, after all, has for hundreds of years possessed essentially no legislature, no international army or police, few lawsuits, and virtually no courts. How can it be "law" at all? This general unease is shared, in my somewhat more specific experience, by both legal academics and political scientists. International law, lacking a rich tradition of textual interpretation subject eventually to authoritative resolutions and to centralized enforcement and reform efforts, leaves the typical legal academic feeling a bit at sea. Lacking the coercive authority of a centralized decision-maker, international law leaves the typical political scientist schooled in international relations--who is almost certain to have been trained to think in terms of power, not persuasion--somewhat hesitant to think that there is a "there" there in international law.

This Article examines one of the few intersections between international law and a theory of international relations (IR)--the theory known as "neo-liberal Institutionalism"[1]--to argue that the

---

[1] The seminal Institutionalist work is Robert Keohane, After Hegemony:
Cooperation and Discord In The World Political Economy (1984). For some
other overviews of neo-liberal Institutionalism by adherents, see Robert
Keohane & Lisa Martin, The Promise of Institutionalist Theory, 20 Int'l Sec.
39 (1995); Charles Lipson, International Cooperation in Economic and Security
Affairs, 37 World Pol. 1 (1984). For some criticisms of neo-liberal
Institutionalism, see John Mearsheimer, The False Promise of International
Institutions, 19 Int'l Sec. 5 (1994); Joseph Grieco, Anarchy and the Limits

international legal rules governing the process of making and interpreting treaty obligations can provide some crucial definition to an IR "theory" that is otherwise doomed to the eventual irrelevance of mere metaphor. Neo-liberal Institutionalists focus upon the role of international "institutions" or "regimes" as the loci of cooperation in solving the dilemma of collective action. The Institutionalist view of international law falls between the unrelenting hostility of the dominant Realist school of IR theory and the ready enthusiasm of IR Liberals.[2] Interested in deploying many of the rational-choice tools of Realism without necessarily adopting its hard-edged conclusions, neo-liberal Institutionalists focus especially upon the "iterated Prisoner's Dilemma" (IPD), an important game-theoretical concept, as the key to understanding the evolution of international cooperation.[3]

Duncan Snidal, a political scientist who has employed a good deal of game theory in the course of his own IR theorizing, has argued that IR theorists tend to use game theory merely as "metaphor," lacking the rigor or negatability that comes from the specification of "models" and their associated "theories."[4] Part I of this Article sets forth Snidal's multi-part schema for the normative evaluation of social-science theorizing and then argues that neo-liberal Institutionalism occupies the lowest rung on Snidal's intellectual-evolutionary ladder.

The remainder of this Article focuses upon the particular game-theoretical concept of the "iteration." Iteration is perhaps the most crucial concept in the game-theoretical analysis of international politics underlying Institutionalist assertions about international cooperation and international institutions. Part II argues that neo-liberal Institutionalists have been lax in defining what they mean by an "iteration" in international relations. The laxity in this definition has disturbing implications. Without a rigorous definition of iteration as applied to international relations, the various associated game-theoretical concepts--payoffs, actions, strategies, and so forth--prove as difficult to define as the campaign promises of a skillful politician. The prospects for a "true theory" of neo-liberal Institutionalism, and with it a higher place for neo-liberal Institutionalism in Snidal's schema, seem bleak if such definitional flaccidity persists. One might even conclude that neo-liberal Institutionalism's casual employment of so critical a concept as iteration leaves neo-liberal Institutionalism's adherents vulnerable to the broader criticism that, like California cuisine, their theorizings are nothing more than trendy borrowings and stylish presentation.

Part III of this Article argues, however, that international law--and most particularly the treaty process--may be able to rescue neo-liberal Institutionalism from its predicament. The

---

of Cooperation: A Realist Critique of the Newest Liberal Institutionalism, 42 Int'l Org. 485 (1988).

[2] For a summary of the various schools of international relations, and of their evolution and inter-relationships, see John K. Setear, Responses to Breach of a Treaty and Rationalist International Relations Theory: The Rules of Release and Remediation in the Law of Treaties and the Law of State Responsibility, 83 Va. L. Rev. 1, 2-3 nn.2-5, 6-9 nn.8-9 (1997) [hereinafter Setear, Breach of Treaty].

treaty process is highly iterative and, more importantly for the purposes of making neo-liberal Institutionalism more rigorous, the iterations of the treaty process are well defined. The three phases of the treaty process--negotiation, signature, and entry into force--all have relatively clear beginnings and endings, and are thus well suited for objective definition as iterations. Of course, there are various complexities involved in actually translating these real-world stages of the treaty process into the iterations and actions of game theory. Part IV of this Article wrestles with those complexities and argues that the international legal process provides a surer footing for such translation than do non-legal areas of international relations.

This Article does not itself undertake an empirical study of the iterations involved in any particular treaty. Rather, the Article argues against the current laxity of Institutionalist thought in treating the concept of the iteration as a background assumption. It also urges consideration of the treaty process as a useful subject for studies by political scientists of an Institutionalist bent and for inquiries by legal academics interested in the interplay of politics and law in the international system. As an intersection of international law and international relations, therefore, the treaty process can help both to identify and to solve a problem of significant theoretical import for neo-liberal Institutionalists.

I. Neo-Liberal Institutionalism as Metaphor

This first part of the Article has three sections. Sub-part A briefly describes neo-liberal Institutionalism; sub-part B sets forth Snidal's multi-part typology on the uses of game theory in IR theory; and sub-part C argues that neo-liberal Institutionalism appears to operate as "metaphor," the least developed intellectual approach in Snidal's typology.

A. An Introduction to Neo-Liberal Institutionalism

The branches of the river of IR theory shift from year to year, but its main stream since World War II has been Realism.[5] Realists believe that international relations is a ceaseless competition for relative gains--especially respecting fundamental issues of national security--among essentially atomistic nations only lightly constrained by domestic politics or international law.[6] Opposition to Realism has at various times come from Idealism, from Liberalism, and from neo-liberal Institutionalism.[7] No matter its name, that opposition has always been less inclined than the Realists to view the world through blood-colored glasses, and more inclined to allow a role for international law.[8]

For the sake of simplicity, this Article uses the phrase "neo-liberal Institutionalism" to denominate the broad range of the partly distinct IR theories that, as opposed to Realism, are characterized by the belief that international institutions play an important role in coordinating international cooperation. The Institutionalist narrative of international relations begins with the

---

[6]

[7]

[8]

same assumptions employed by Realists, but the Institutionalist ending has a different twist. Neo-liberal Institutionalists treat international cooperation as a problem of collective action.[9] One may capture this problem with the game-theoretical concept of the Prisoner's Dilemma (PD), a situation that gives each player certain incentives to refuse to cooperate yet also bestows significant benefits if mutual cooperation occurs.[10]

The crucial distinction between the Realist and Institutionalist narratives lies in the Institutionalist emphasis on the idea that cooperation can evolve in an iterated Prisoner's Dilemma--that is, in a situation in which parties will play the PD repeatedly.[11] If parties can recognize one another at each encounter and remember their previous interactions--both of which seem reasonable assumptions in the context of international relations--then cooperation may evolve over time despite the short-run incentives to defect from the cooperative endeavor.[12] The computer-moderated tournaments of Robert Axelrod, himself an IR scholar, are especially well known in this regard.[13] In these tournaments, the "tit-for-tat" strategy--initial cooperation, followed by behavior in the current iteration corresponding to the opponent's behavior in the previous round--proved to be successful.[14] Indeed, the tit-for-tat approach to the IPD has been touted as a wise approach for the promotion of cooperation in a number of contexts.[15] Axelrod and Robert Keohane, the scholar most closely identified with neo-liberal Institutionalism, have co-authored an article discussing the possible virtues of the tit-for-tat strategy and examining the implications of Axelrod's abstract computer tournaments for the real world of international relations.[16]

In the Institutionalist view, the eponymous international institution can serve as a focus for international cooperation.[17] Just what constitutes an "institution" (also called an "organization" or "regime" in some contexts) has always been a bit mysterious, but two popular definitions give some flavor of the concept. An early definition characterized a regime as a set of "principles, norms, rules, and decision-making procedures around which actor expectations converge in a given issue-area."[18] More recently, Keohane has described institutions as "persistent and connected sets of rules (formal or informal) that prescribe behavioral roles, constrain activity, and shape expectations."[19] Other scholars offer their own definitions.[20]

Regardless of terminological or definitional niceties, the basic idea is that the rules and

9

10

11

12

13

14

15

16

17

18

19

20

expectations promoted by institutions ease the difficulties of coordination.[21] The existence of stable institutions facilitates the sort of repeated interaction that assists the evolution of cooperation. Rules, combined with a concern for reputation, can discourage defections from the cooperative scheme once that scheme is established.[22] Institutions can brake non-cooperative tendencies in the international system by providing an environment for the resolution of various disputes.[23] At least one scholar of international law and international relations has observed that all of this talk about institutions sounds much like certain conceptions of international law advanced by international lawyers working roughly contemporaneously, or perhaps even before, the relevant political scientists.[24] For our purposes, however, we may avoid the issue of cross-disciplinary credit and simply proceed from this summary of neo-liberal Institutionalism (and the game theory embedded within it) to Snidal's discussion of the role that game theory could or does play in IR theory.

B. *Snidal's Typology on the Uses of Game Theory: Metaphor, Analogy, and True Theory*

In exploring the utility of game theory to IR theory, Duncan Snidal has advanced a four-part scheme describing various modes of analysis that IR theorists might employ in the study of their subject matter.[25] In Snidal's schema, a "metaphor" is the least rigorous use of a concept; an "analogy" is of greater rigor; and a "model," combined with an underlying "theory," is the highest mode of intellectual endeavor.[26]

Snidal does not expressly define metaphor, but he implicitly views metaphor as a simple assertion that one thing is like another. In criticizing the use of game theory as metaphor, for example, Snidal states:

> We do not improve a metaphor simply by translating it into a game matrix. Glib ssertions that "Issue *X* is Prisoner's Dilemma," or that the "Cuban missile crisis was a game of Chicken," efficiently convey a metaphor, but do not make the metaphor more plausible or take much advantage of the power of game theory. Typically, such statements simply restate what we already know--perhaps embellished by means of a particular reconstruction of historical events.[27]

Analogy, in Snidal's scheme, is a more elaborate comparison than a metaphor. Rather than the single correspondence asserted in a metaphor, an analogy asserts a whole series of correspondences between two analogized phenomena.[28] The person drawing the analogy then asserts that some known characteristic of one phenomenon will translate into an equivalent--and

---

22

23

24

25

26

27

28

previously undescribed--characteristic of the other phenomenon.[29] Snidal's example is an analogy between firms in an oligopolistic market and the international political system.[30] The firm is analogous to the nation-state; the maximization of profits is analogous to the maximization of the prospects for national survival; the existence of oligopolists is analogous to the existence of great powers; price wars are analogous to military conflicts; the assumption of rational behavior by firms is analogous to the assumption of rational behavior by states; and so on.[31] One may then hazard a specific analogy from a well-known conclusion about one phenomenon to a conclusion on a controversial topic in the other phenomenon: if oligopolistic markets have fewer price wars than more competitive markets, then international systems with a small number of great powers should have fewer military conflicts than international systems with a more even ("competitive") distribution of power.[32]

Analogies are more elaborate than metaphors, but they share a similar logical structure. Both are comparative. Both are sufficiently loose-jointed to allow for significant discrepancies between the two subjects under comparison. Such discrepancies need not lead to the abandonment, or even the reformulation, of the metaphor or analogy. Rather, one simply calls the comparison "imperfect."[33]

A model, in contrast, operates on much finer tolerances. The model involves an internal, not a comparative, logic.[34] From its assumptions come its predictions. Take a simple mathematical model: $v = kt$. This model states that some dependent variable, v, is a function of a product of two independent variables, k and t. When one compares the model to the phenomenon under examination, either the predictions are true or the model must be reformulated.[35] If v proves to vary with the cube of t or proves under some circumstances to vary with some other variable, m, then one must present a different or modified model before one can confidently begin making predictions based on the theoretical construct.

A model is simply the specification of a set of logical relationships among abstract variables. In order to translate a model into specific statements about the real world, one needs what Snidal calls a theory: "a deductive structure plus an interpretation of fundamental assumptions and theoretical constructs."[36] For example, in the model described above, a theory might state that v represented the velocity of an object falling through a fluid, k represented a constant that one could deduce from observation, t represented the time since the object had been released, and m, if incorporated into the model (perhaps to account for the effects of drag), represented the mass of the object. To distinguish Snidal's meaning of "theory" from the more common and casual usage of this word as any guess at explaining a particular phenomenon, this

---

29

30

31

32

33

34

35

36

Article uses the phrase "true theory" to denote Snidal's fourth category of uses for intellectual constructs.

Snidal clearly believes that, although metaphor and analogy have their place, it is the combination of model and true theory that should be the goal of the IR theorist.

> Metaphor, analogy, model, and theory are complementary in social scientific research; they are each appropriate at various stages. It is often useful to go back and forth among them. As research *advances*, however, metaphor and analogy are of increasingly limited usefulness. The greater rigor and deductive power of the model, together with the interpretive richness and open-endedness of its corresponding theoretical framework, make that combination *ultimately* more productive.[37]

Snidal also clearly believes that the use by IR theorists of game theory is typically mired at the lower levels of analysis, despite the potential of game theory to assist in the formulation of a true theory of international relations:

> Too many "applications" of game theory have merely been in the spirit of sorting out whether the Cuban missile crisis was really Chicken or Prisoner's Dilemma. Such usage may be helpful for reconstructing and interpreting particular events, but it misinterprets the primary value of game theory as that of redescribing the world, and is therefore limited as a test of game theory. It would be a more appropriate test of a deductive theory to investigate the empirical correctness of its analytical predictions. This requires giving empirical content to it through its assumptions (e.g., about preferences and payoffs) rather than just adapting the theory (or one of its many models) to fit some historical or current event.[38]

Snidal was criticizing both Realism and neo-liberal Institutionalism for their casual use of the potentially powerful tool of game theory. IR theorists have since begun to employ more methodologically sophisticated applications of game theory, although whether the result is a model accompanied by a true theory is unclear.[39] I now turn more specifically to game theory and neo-liberal Institutionalism. Sub-part C argues that neo-liberal Institutionalism uses game theory merely as a "metaphor" under Snidal's scheme. Part II then focuses on the particular importance within Institutionalist thought of "iteration," a crucial game-theoretical concept.

C. *Game Theory as Metaphor*

The use of game theory in neo-liberal Institutionalism in its current form is, in Snidal's schema, a metaphor. The Institutionalist argument does go beyond asserting a mere correspondence between a particular situation and a game-theoretical concept (for example, describing the Cuban missile crisis as a game of Chicken). Neo-liberal Institutionalists make a

---

37

38

39

much broader comparison, one between all (or at least some important part) of international relations and the IPD.[40] Despite the breadth of this comparison, however, the shallowness of the correspondence still renders the use of game theory merely metaphorical.[41] In his seminal Institutionalist work, *After Hegemony*, Robert Keohane discusses game-theoretical work in at least a modicum of depth.[42] When it comes time to tie such work to international relations, however, the relationship between game theory and IR theory is asserted only as a vague correspondence:

> In developing a functional theory of international regimes, I will rely in part on the logic of Prisoner's Dilemma and theories of collective action . . . .[43]

> The literatures on collective action [and] Prisoner's Dilemma . . . all suggest the plausibility of a functional explanation for the development of institutions. Institutions, according to this argument, are formed as ways to overcome the deficiencies that make it impossible to consummate even mutually beneficial arrangements.[44]

> All of these incentives for compliance [in a world of multiple regimes] rest on the prospects of retaliatory linkage: as in Axelrod's (1981) simulation of Prisoner's Dilemma, 'tit for tat' is a more effective strategy to induce cooperation than submissiveness. We have seen that GATT contains provisions for retaliation; and the Bretton Woods Agreement of 1944 furnishes another relevant example.[45]

Does this treatment rise from metaphor to the level of analogy in Snidal's schema? Apart from the overall correspondence asserted between the context for international cooperation and the Prisoner's Dilemma (PD), Keohane explores only the correspondence between the nation-state in international relations and the player in game theory, and explores it only briefly at that.[46] There is no explicit discussion of the real-world equivalents of the game-theoretical notions of action, payoff, iteration, strategy,[47] discount rate, or probability of future interaction. The PD is, in Institutionalist thought, simply a metaphor.

The neo-liberal Institutionalists' treatment of game theory would be too vague in Snidal's view:

> Applying game theory to a substantive body of knowledge such as international relations raises a host of difficult empirical questions. For example: Who are the relevant actors? What are the rules of the game? What are the choices available to each actor? What are

---

40

41

42

43

44

45

46

47

the payoffs in the game? Is the issue best characterized as single-play or repeated play?[48]

A game-theoretic perspective requires analyzing states' motivations and how their preferences map into payoffs within a game model. Establishing this correspondence between an issue area and its game model is the toughest problem confronting successful empirical application of game theory.[49]

As discussed in Part II.A below, the question of "single-play or repeated play"--that is, the absence or presence of multiple iterations--is especially important. Without iteration, game theory predicts no cooperation.[50] Without an explanation for international cooperation, neo-liberal Institutionalism is hard pressed to distinguish itself from, and thereby offer an alternative to, Realism. I turn now to the importance of iteration within the game theory of the PD and to the treatment of iteration by neo-liberal Institutionalists.

II. Iteration and Neo-Liberal Institutionalism

This Article now turns to the Institutionalist treatment of iteration and suggests that the concept of iteration requires significantly more elucidation and rigor than has been evident in recent Institutionalist work. The Article then demonstrates the inadequacy of two common (implied) definitions of iteration and the importance of properly defining iteration in Institutionalist thought.

A. *The Importance of Iteration in Institutionalist Theory*

As mentioned above, Institutionalist theory makes pivotal use of the iterated Prisoner's Dilemma (IPD) and of various computer-moderated tournaments, conducted by Robert Axelrod, that involve the IPD.[51] By arguing that international cooperation can be explained on the basis of an IPD, neo-liberal Institutionalists can argue that cooperation may evolve in the international political arena. Neo-liberal Institutionalists offer international institutions as the mechanism by which such cooperation may evolve.[52] Realists are much more pessimistic about international cooperation and institutions because they view the international political arena as an *un*iterated PD.[53] To explain the crucial difference between an IPD (the neo-liberal Institutionalists' foundational metaphor for the evolution of international cooperation) and an uniterated PD (the metaphor preferred by Realists) requires an exegesis about game theory itself.

A retelling of the traditional "story" behind the PD sets the stage for an assessment of its crucial features. Imagine two criminals caught by the police following the commission of a

---

48

49

50

51

52

53

crime. The police are holding the criminals in separate cells[54] and are hoping that one or both of the criminals will betray the other, thus giving the police additional evidence. Each criminal has the option to remain silent or to betray her co-conspirator. If both criminals remain silent, both will receive a plea bargain (because of a lack of evidence) and thus be imprisoned only for a short time. If both attempt to "rat" on each other, both will be convicted on the original charge and both will go to jail for a long time. If one "rats" and one remains silent, the prisoner who talked will be rewarded by the police with immunity, while the silent criminal will face a very long sentence.[55]

The defining feature of a PD is the rank-ordering among the payoffs to be awarded to a player for a given outcome of the game. Let us use *C* ("cooperate" with the co-conspirator by remaining silent) and *D* ("defect," or betraying the co-conspirator and implicating her in the crime) to designate the two actions available to each player. Let us denote the outcome of the simultaneous choice by both players of their actions by stringing the two letters together; for example, *CC* indicates the outcome when both players choose the *C* action (to cooperate by remaining silent). If we place a player's own action first in the string--for example, *DC* indicates the outcome when a player chooses the D action and her opponent chooses the C action--then one may define the PD as a rank-ordering among the payoffs to each player, as follows: *DC* > *CC* > *DD* > *CD*, where ">" is read as "is preferred to" from the standpoint of the player whose action is listed first. This rank- ordering is true in a PD for each player viewing the situation from his or her own perspective. Figure I below shows a hypothetical payoff matrix for one version of the PD.

FIGURE I: THE PRISONER'S DILEMMA[56]

		PLAYER B	
		COOPERATE	DEFECT
PLAYER A	COOPERATE	(3,3)	(0,5)
	DEFECT	(5,0)	(1,1)

The PD gives scholars a ready metaphor for problems of cooperation ranging from arms control to water-rights allocation.[57] If both players cooperate with one another, then the payoff for each player will be higher than the payoff for each player when both players defect. If one player defects while the other blithely cooperates, however, then the defecting player reaps the highest possible payoff (and the cooperating player is left with the lowest possible payoff). If the players want reliable gains and can trust each other, then each should choose cooperate as her

---

54

55

56

57

action. The temptation to defect in hopes of unilateral gain complicates the situation, however.

The game theorist believes that a single play of the PD is in fact quite easy to analyze: any two rational players will always defect.[58] The game theorist's reasoning is straightforward: a player contemplating her choice will see that she is better off defecting no matter what action her opponent chooses. If her opponent chooses to defect, then she will receive a better payoff from defecting herself (*DD*) than by cooperating (*CD*). If her opponent chooses to cooperate, then she will again receive a better payoff by defecting (*DC*) than she would receive if she cooperated (*CC*). Regardless of her opponent's choice, therefore, she gains more by defecting.[59]

Of course, if both players adopt this reasoning, then the outcome will be *DD*, which is unfortunate not merely because it is the second-worst possible payoff but because the equally symmetrical *CC* would have yielded a higher payoff to both players. Nonetheless, defection is the "dominant" strategy--holding the other player's choice constant, defection yields a higher payoff than cooperation.[60] Defection is also the "Nash equilibrium," a strong and commonly used normative criterion for choosing the proper action in a game.[61] Game theorists therefore predict that both players will choose to defect in an uniterated PD, and there is some experimental evidence to suggest that this argument is correct.[62]

These results apply only to the "one-shot" or uniterated PD. If players instead face a whole series of PDs to be played sequentially, then rational players may find that cooperating is a desirable action.[63] The "shadow of the future" can make the short-run gain from defection unattractive because that defection risks the ire of one's opponent in future rounds.[64] Keynes may have been correct to say that we are all dead in the long run,[65] but that same long run may make us rationally cooperative in the time between cradle and grave.

An especially prominent analysis of the long-run interaction of players in an IPD is the work of Robert Axelrod.[66] Axelrod solicited numerous algorithmic strategies for playing an IPD from game theoreticians and then pitted those strategies against one another in a variety of computer-moderated tournaments.[67] The winner of the tournaments was the simple tit-for-tat strategy, and a variety of the other successful strategies also displayed a strong tendency towards cooperation--at least when the strategies they faced were also cooperatively inclined.[68] "Cooperation *can* evolve" (or perhaps "cooperation can *evolve*") has been the message taken

---

58

59

60

61

62

63

64

65

66

67

68

from these tournaments by neo-liberal Institutionalists (and many others).[69]

The importance of iteration to this analysis can hardly be overstated. An *un*iterated PD, as just discussed above, is a game-theoretical recipe for mutual defection.[70] An IPD, in contrast, presents the possibility that cooperation will occur.[71] In fact, if cooperation is to occur, game theory suggests that either there must be an infinite number of iterations or the players must be unaware of which iteration will be the final round.[72]

Axelrod's tournaments were densely iterative. First, they involved a large number of rounds of interaction.[73] Second, the payoff scheme did not involve the discounting of future payoffs by the players,[74] an assumption consistent only with a densely iterative environment.[75]

The presence of iteration is therefore a necessary condition for the evolution of cooperation, and the evolution of cooperation in the international political arena is at the heart of the Institutionalist narrative. Iteration is thus a crucial assumption in Institutionalist thought. This Article now turns to the way in which neo-liberal Institutionalists have actually treated this central concept.

B. *The Casual Way in Which Neo-Liberal Institutionalists Treat Iteration*

Despite the importance of iteration in the game theory relevant to Institutionalist assertions, I can find no explicit definition of iteration in the work of neo-liberal Institutionalists. Nevertheless, it is clear to the neo-liberal Institutionalists that the concept of an iteration is crucial to their explanation of the existence of international cooperation:

> [The evolution of] cooperation need not involve any negotiation at all, since mutual adjustment can take place without direct communication between the participants. In this book, however, we focus on coordination achieved through bargaining. Such bargaining typically occurs not only in one bargaining episode but in several, over a period of time. Negotiations on international monetary arrangements, trade, and energy take place continuously and are expected to continue indefinitely into the future. Furthermore, the fact that many closely related negotiations take place simultaneously increases the "multiple-play" rather than "single-play" character of the game. Usually, unlike the actors in Prisoner's Dilemma, governments can reverse decisions to cooperate if they discover that their partners are reneging on their own agreements. This possibility has an effect similar to that of iteration of the game, since it reduces the incentives to defect. Thus even insofar as international negotiations can be modeled in the simple form of

---

69

70

71

72

73

74

75

Prisoner's Dilemma--and we will see later that to do so requires a number of questionable simplifying assumptions--the pessimistic standard conclusion of [the] single-play Prisoner's Dilemma does not follow.[76]

On one level, this passage presents a perfectly sensible argument that, as a matter of metaphor, international relations more closely resembles an iterated than a one-shot PD. Bargaining occurs in "several [episodes] over a period of time," or even "continuously and . . . indefinitely."[77] The simultaneity of various negotiations "increases the 'multiple-play' . . . character of the game."[78] The ability of governments to reverse their decisions "has an effect similar to that of iteration."[79] Thus, because the game of international relations is iterated, the "pessimistic standard conclusion of [the] single-play Prisoner's Dilemma does not follow."[80]

On another level, however, this passage illustrates only too well the difference between metaphor and more sophisticated modes of analysis. Viewed as a prelude to modeling, the passage is imprecise and full of inconsistencies. Note, for example, that the third sentence discusses bargaining "episodes," while the fourth sentence mentions negotiations that take place "continuously." From a game-theoretical perspective, the examination of episodic ("discrete") games is quite different from that of continuous games. For example, Axelrod's tournaments, as well as most other work treating the PD, involved discrete games. One cannot rely upon the results of episodic games to model continuous phenomena. To move forward with neo-liberal Institutionalism, one would therefore have to choose between an episodic and a continuous representation of the phenomena at issue, and then proceed accordingly down either of the two resulting pathways.

One should also observe that, at the level of model rather than metaphor, Keohane's assertion that "many closely related negotiations which take place simultaneously" will increase the "multiple-play" nature of the game is precisely backwards. Greater simultaneity makes a situation *less* like a multiple-play game: at the limit, in a *single*-play game, *all* relevant events occur simultaneously.[81] One can, of course, conceive of myriad games in which a player chooses multiple actions simultaneously,[82] but the resulting complexities do not neatly correspond to the complexities which result from an increase in the number of iterations in a game.[83]

The first two sentences in the passage quoted above also reveal a somewhat loose approach to the underlying game theory at issue: "[The evolution of] cooperation need not involve any negotiation . . . . In this book, however, we focus on coordination achieved through

76

77

78

79

80

81

82

83

bargaining."[84] The evolution of cooperation not only "need not involve any negotiation," but *must* not involve any *enforceable* negotiation. The PD, as explored by Axelrod and others, is *by definition* part of a class of games in which negotiation, as opposed to action, is irrelevant.[84] The mere presence of negotiations does not preclude the existence of a PD, but the parties' actions must be the only credible indication of behavior, not the parties' words. If negotiations are relevant, then the PD as traditionally analyzed no longer applies, and neo- liberal Institutionalists must look elsewhere for their modeling insights.

Finally, one should note that the possibility that "governments can reverse decisions" is not "*similar* to that of iteration of the game."[86] The possibility of reversal *is itself* one additional iteration. A decision can only be "reversed" if there is a subsequent period in which to accomplish the reversal.[87]

Considered as metaphor, the statement quoted at length above may be quite adequate. As a (preliminary) statement about an actual model of international relations, however, the passage is unsatisfactory.[88] Moreover, one may not dismiss this passage as unrepresentative of Institutionalist theory. Not only is the passage taken from the seminal Institutionalist work, it is also the *most* extensive discussion of the concept of iteration that I have found in the Institutionalist literature.

Before underscoring the importance of a rigorous definition of iteration and attempting to develop such a definition from the law of treaties, this Article examines further two implicit notions of iteration that appear in the Institutionalist literature. Neither is satisfactory.

C. *Two Implicit Notions of Iteration in Neo-Liberal Institutionalism*

One can perceive two implicit notions of iteration in the Institutionalist literature. One such notion stems from the observation that international relations occur over time. The other stems from the assertion that nations retaliate against other, uncooperative nations.

Like most complex relationships, international relations play out over time, not all at once.[89] One might also note that nation-states typically persist over long periods of time. Nations therefore have the opportunity, over time, to interact with one another repeatedly. If they do in fact interact repeatedly, then one might analogize each repetition to an iteration.[90]

Unfortunately, this view tells us little about exactly what constitutes a repetition or an interaction. What exactly constitutes an iteration in the relations among nations? The passage of

---

84

84

86

87

88

89

90

a day? A month? A year? Each diplomatic cable? Each press conference called by a cabinet-level officer? Each summit meeting between heads of state? In addition, this view leaves one with little sense of the substantive breadth of an issue covered in an iteration. Does any meeting between any two representatives of the relevant nations constitute an iteration in the game of international relations? Is any meeting between two members of a given pair of trade ministries an iteration in a game of trade policy? Or is a meeting in which a single issue is discussed an iteration in a game concerned with that particular issue of trade? The fact that international relations occur over time does imply that international relations is iterated. Without more definition, however, such a conclusion provides no clue as to what constitutes an iteration.

A somewhat more specific, if still implicit, justification for treating international relations as an iterated phenomenon is found in narratives that describe particular actions by one nation as giving rise to retaliatory actions by another.[91] These accounts clearly imply that the game of international relations consists of more than one iteration. A retaliation can presumably occur only in some iteration subsequent to the provoking iteration.[92] Thus, there must be at least two iterations in the underlying game.

The retaliation-based view of iteration also implies something about *how* to define an iteration. If policy action *B* is a retaliation for policy action *A*, then policy action *B* must occur in an iteration *after* the one in which policy action *A* occurred. The time at which *B* occurs thereby defines a time by which the iteration containing *A* must have *ended*. Suppose, for example, that the United States' and Soviet Union's nuclear forces, as of December 31, 1967, all used warheads with a single nuclear bomb in each warhead. On January 1, 1968, the United States begins to deploy warheads with multiple bombs in each warhead. On January 1, 1970, the Soviet Union retaliates by deploying warheads with multiple bombs in each warhead. Assuming the Soviet retaliation is a sequential rather than a simultaneous move,[93] we can thus reliably infer that the first iteration had ended by December 31, 1969.

The retaliation-oriented view has fundamental difficulties of its own, however, which stem chiefly from implicitly allowing a nation's retaliatory *action* to define what constitutes an *iteration*. The retaliatory view of iteration tells us only that the first iteration had ended at some point before the retaliation occurred; it does not tell us whether that first iteration ended moments before, or months before. In the example given above, the whole year of 1969 could permissibly be part of either the first or the second iteration. For example, if the first iteration is said to have ended on December 31, 1968, and the second iteration on December 31, 1970, then the U.S. action occurred in the first iteration and the Soviet action in the second. Alternatively, however, one could say that the first iteration did not end until December 31, 1969; the U.S. action would still occur in the first iteration and the Soviet action in the second. The retaliation-oriented view thus fails to provide an unambiguous definition of iteration.

The most serious limitation of the retaliation-oriented view, however, is that it restricts

---

91

92

93

the ability to define an iteration to those situations in which one nation has defected and the other has already retaliated. Suppose, for example, that, as before, the United States' and the Soviet Union's nuclear forces, as of December 31, 1967, all used warheads with a single nuclear bomb in each warhead. On January 1, 1968, as before, the United States began to deploy warheads with multiple bombs in each warhead. Now let us change our previously hypothesized example and assume that, as of January 1, 1970, the Soviet Union has *not* deployed any multiple-bomb warheads. Has the Soviet Union thereby cooperated for at least one iteration despite the U.S. defection? How can one define the iteration in the absence of a retaliation? One apparently needs to wait until some retaliatory action is taken before it can be said that one round has ended and a new round has begun.

To put the point a slightly different way, the length of an iteration in the retaliation-oriented view can expand or contract with the timing of the retaliation. Suppose that the Soviets do not deploy multiple-bomb warheads until January 1, 1988. At that point the retaliation-oriented definition can spring into action to determine that the prior iteration lasted roughly two decades. In the situation where retaliation occurred within two years, however, the iteration length was said to be two years. If the Soviets defect within x years, then we know that the prior iteration lasted no more than x years, but if the Soviets have not yet defected as of the date on which we examine the situation, then the retaliation-oriented view can provide us with no definition of an iteration. This seems an unsatisfactory, *post hoc* definition at best. Indeed, the whole notion of retaliation and iterations as defined thereby essentially assumes its conclusion. If the only way to define a line between two iterations is through a retaliation, then one will *always* "observe" a defection by the retaliating nation in the relevant iteration, with the iteration of two or twenty years as the situation changes.

Note also that the retaliation-oriented view ignores entirely the relationship between the parties in the period *preceding* the first defection, yet one might well be interested in whether that pre-defection period was a lengthy or brief one, whether it included unanswered defections, and so forth. Consider two different initial histories between two nations. In the first, the two nations have been interacting for 100 iterations; at iteration 25, one nation defected without a retaliatory defection by the other, and both have cooperated consistently since then. In the second history, two nations have interacted for two iterations, in both of which both nations cooperated. A new defection in each of these situations might lead to different results, or to differing degrees of confidence in predicting those results. The retaliation- oriented view is of no help in making such distinctions, however, because it completely ignores the question of how to determine how many iterations have preceded the defection in question.

D. *The Importance of a Rigorous Definition of Iteration*

A definition of iteration is a fundamental prerequisite to rigorous Institutionalist inquiry. Without such a definition, one cannot assign payoffs to real-world actions, and thus one can hardly proceed with modeling and true theory based upon game-theoretical concepts.   In addition, the evolution of cooperation, which plays a prominent role in Institutionalist metaphors, depends upon the interaction of various strategies and without some definition of iteration, one cannot determine which strategies a nation is employing. The Institutionalist focus on the IPD makes a definition of iteration especially important because different definitions of iteration can

change a situation from a PD into some other type of game. And within the set of relationships among payoffs that constitute a PD, changes in the definition of an iteration can vary the ratios of payoffs in ways that affect the likelihood and stability of cooperation. This Article now takes up in modest detail each of the following uses for a definition of iteration: defining payoffs and actions, determining the strategies employed by a nation, determining the type of game involved in a particular context, and determining the ratio of payoffs within a PD.

1. *Payoffs and Actions*

The payoffs in a particular game are crucial to an analysis of that game's outcome. Indeed, it is the rank-ordering of the payoffs in the PD that identifies the particular type of game. The different relationships among various payoffs are also the defining characteristic of such games as Chicken, Stag Hunt, and other games.[94] These game types are not simply arbitrary labels of convenience; rather, the names of various games reflect the differing dilemmas (or, sometimes, the lack thereof) that players face in a given situation.[95] Indeed, with *no* specification of payoffs, there is no cognizable game. With some particular specification of payoffs, in contrast, a generic game becomes a particular type of game, such as the PD.[96]

In the real world, time passes continuously. The IPD, in contrast, implies a world of discrete, periodic decision-making. An iteration occurs; players choose their next actions; then another iteration occurs, and so on. One must therefore convert continuous (real world) time into discrete time intervals by designating some duration for each iteration, especially if one wishes to calculate the benefits and costs from various combinations of outcomes. Suppose, for example, that a particular nation adheres to an agreement regulating pollution at its borders. Assume that benefits accrue whenever an industrial plant refrains from emitting a pollutant that it would have emitted absent the agreement. Such benefits accrue at various times--presumably more often on weekdays than on weekends (if the plant's output is greater during the workweek than on weekends), perhaps only rarely at night (if night shifts are rare), and so forth.

What are the payoffs in the related game matrix? One cannot answer this question, and thus one cannot incorporate the answer with other payoffs to construct a game matrix, without picking some period over which to accrue the benefits. Perhaps exactly $1 million in environmental damage to each nation is avoided each month. If each iteration lasts for one month, then $1 million worth of benefits should be included in the payoffs. If one chooses a year as the length of an iteration, then $12 million is each nation's payoff from joint cooperation. Regardless, one must choose *some* interval if one is to calculate payoffs.

Note that, so long as the players consider the absolute magnitude of payoffs in their decision making process, the definition of iteration may affect the choice of action. To take a non-game-theoretical example, a person might be willing to bet one dollar on the flip of a coin, but be unwilling to bet $1 million dollars on the coin toss, even where the ratios of the various

94

95

96

payoffs are the same in both cases, that is 1:1. The same result might obtain in the game-theoretical context, except that it is the choices made by a player and her opponent, not a coin toss, that determine the outcome. If the payoff matrix in Figure I above[97] represents dollars, a particular person might choose to cooperate as her first action in an iterated game. Perhaps, internally, she focuses upon the desire to set up a cooperative relationship so that, in the long run, she can accumulate a substantial fortune. If the units in the payoff matrix in Figure I represent millions of dollars, however, perhaps she would choose to defect as her first action. She might, for example, focus upon the possibility of earning $5 million in the first round (if only her opponent will choose cooperate just this once) and then assure herself a mere $1 million in each subsequent round by always defecting. A definition of iteration with a longer interval between iterations increases the absolute magnitude of the payoffs, and so, if a nation chooses its action based in part on the absolute magnitude of the payoffs involved, then the definition of iteration may affect that nation's choice of actions in its international relations.

2. *Inference of Strategies*

The choice of strategy--the rules, as in tit-for-tat, that specify what action a player will take in a particular set of circumstances in any given iteration--is obviously an important part of the analysis of an IPD. The analysis of strategies, for example, is clearly the focus of Axelrod's book.[98] More generally, analysts of game theory tend to set out many of the other characteristics of the game--payoff, actions, and so forth--as assumptions and then undertake an examination of the optimal strategy for a player to adopt given those assumptions.

Some definition of an iteration is necessary to infer the strategies pursued by nations. In the real world, national governments do not submit to political scientists the algorithms by which they define their strategies. Instead, the analyst attempting to test a game-theoretical hypothesis would presumably observe the *actions* of governments and then attempt to infer their strategies *ex post*. For example, if a nation consistently cooperated in the iteration following a cooperative action by its opponent, and consistently defected in the iteration after its opponent defected, then the analyst might infer that the nation had adopted a tit-for-tat strategy.

As with the calculation of payoffs, one cannot characterize a nation's actions (and thereby infer its strategy) in an IPD without some definition of iteration. Assume that we characterize a nation's action as whatever policy that nation adopts for the majority of the interval of time chosen for the iteration.[99] Suppose, for example, that the United States allowed Mexican goods to pass freely through U.S. points of entry for three weeks and a day, and then barred their entry for six days. If we define an iteration as two weeks in length, then the string of U.S. actions would be *C-C*.[100] During the entire two weeks of the first iteration, the U.S. chose the cooperative action of allowing the goods to pass, and in eight of the fourteen days of the second iteration, the U.S. again chose to cooperate. For a majority of each iteration, therefore,

---

97

98

99

100

the U.S. allowed the goods to pass freely.

If, instead, we define the iteration as one week in length, then the string of U.S. actions would be *C-C-C-D*. In the first three iterations the U.S. cooperated fully, while in the last iteration, the U.S. barred free passage 86 percent of the time. With the shorter iteration length, analysts have a different set of actions for which to account than they do in the case of a longer interval--most prominently, they must now explain a defection. The definition of iteration thus affects the analysis of the strategy chosen.

Difficulties arising from having an unclear iteration length could affect practitioners as well as analysts of international relations. Suppose that Israel believes that it has one month in which to respond officially to an official U.S. proposal regarding Middle East peace talks, while the United States believes that the relevant iteration length is two weeks. If Israel intends to respond in an unambiguously cooperative fashion after *three* weeks have passed, then the United States will see a failure to respond within its two-week deadline, and is likely to interpret that failure as a defection. Indeed, the United States, piqued, might soon send a signal of dissatisfaction that Israel interprets as evidence of an uncooperative posture on the part of the United States, with the possible result that Israel would reverse its response to the original proposal. As a result of a lack of clarity in the definition of an iteration, a situation in which both players intended to cooperate might become a situation in which both players ultimately defect.

3. *The Type of Game Involved*

The definition of an iteration may even affect the type of game that appears to be at issue. Although the PD is the game to which neo-liberal Institutionalists have paid the most attention,[101] the PD is far from being the only type of game.[102] The different incentive structures that characterize different types of games lead to different optimal actions in many cases.

Suppose, for example, that two nations must invest $2.5 million each in new technology, which takes two months to install, in order to prevent $500,000 in environmental damage to each nation in each month following installation. If the relevant iteration lasts only two months, the game presents very little in the way of a dilemma: should each nation spend $2.5 million to receive nothing? (Figure II below shows this situation.[103]) This is not the set of incentives captured in the PD.

FIGURE II: Limiting Environmental Damage:Two-Month Iteration Interval[104]

PLAYER B

101

102

103

104

		COOPERATE	DEFECT
		COOPERATE	DEFECT
PLAYER A	COOPERATE	(-2.5, -2.5)	(-2.5, 0)
	DEFECT	(0, -2.5)	(0, 0)

If, on the other hand, the relevant iteration lasts six months, then the situation does present a PD. (Figure III below shows this situation.[105]) If the type of game at issue matters--and this is the assumption at the root of the neo-liberal Institutionalists' interest in the PD as an operative metaphor--then the definition of iteration can affect the analysis in a crucial fashion.

FIGURE III: Limiting Environmental Damage: Six-Month Iteration Interval

PLAYER B

		COOPERATE	DEFECT
PLAYER A	COOPERATE	(1.5, 1.5)	(-0.5, 2)
	DEFECT	(2, -0.5)	(0, 0)

4. The Ratio of Payoffs

The definition of an iteration can also affect the ratios of payoffs within various PDs. (This phenomenon is essentially a lesser included example of changes in the type of game that may occur with changes in the definition of iteration.) Suppose, for example, that cooperation requires a nation to cease production of an ozone-depleting substance such as CFC sub11 , in favor of a costlier, but more "ozone-friendly," substitute.[106] A cooperating nation will incur some initial expenses in facilitating the use of the substitute (such as research and development expenses, domestic implementing legislation, industrial retooling, and so forth). Suppose that such expenses all occur within a single year; that such expenses produce no benefits in that first year; that these one-time expenses total $30 million; and that, forever after, the cost per nation of using the ozone friendly substitute is $55 million per year over the costs of using CFC sub11 . Suppose further that it is possible to calculate in monetary terms the annual benefits of this conversion (in terms of reduced skin cancer, damage to crops, and so forth) as $50 million per cooperating nation. If we ignore the time value of money and adopt an iteration length of four years, then the first iteration in this situation presents us with the same payoffs as in Figure IV

105

106

below, which is a PD.[107]

FIGURE IV: Ozone Reductions: Four-Year Iteration Interval

| | | PLAYER B | |
		COOPERATE	DEFECT
PLAYER A	COOPERATE	(105, 105)	(-45, 150)
	DEFECT	(150, -45)	(0, 0)

If we instead designate an iteration period of five years (still ignoring the time value of money), then the first iteration presents us with the situation shown in Figure V below.[108]

FIGURE V: Ozone Reductions: Five Year Iteration Interval

| | | PLAYER B | |
		COOPERATE	DEFECT
PLAYER A	COOPERATE	(150, 150)	(-50, 200)
	DEFECT	(200, -50)	(0, 0)

Both Figures IV and V portray PDs, but the ratios of payoffs vary. In Figure IV, for example, the ratio of the payoff to the first player in the *DC* outcome (when the first player defects while the second cooperates) to the payoff to that first player in the *CC* outcome (when both players cooperate) is 1.43 (150 divided by 105), while in Figure V the same calculation yields a ratio of 1.33 (200 divided by 150). The increase in the time interval covered by the iteration means that the "return" to defecting is higher in the four-year case than it is in the five-year case. If the players consider the ratios of these payoffs in their decision making processes, then the definition of an iteration will matter. If, for example, many players would cooperate when the *CC* outcome yields a payoff half as large as the *DC* outcome, but few players would cooperate when the *CC* outcome yields a payoff equal to just one-tenth of the *DC* outcome, then the ratio of these payoffs is relevant to the choice of action by the players. The definition of iteration can thereby affect the analysis.

107

108

III. Iteration and the Law of Treaties

The previous part examined the importance of defining iteration more rigorously if neo-liberal Institutionalism is to move from metaphor to a true theory. This part examines how the international law of treaties can provide one relatively clear and objective set of definitions of an iteration, and can provide two different methods to guide determinations of whether a nation's behavior at each of those iterations should be interpreted as a cooperate action or a defect action.[109]

Sub-part A briefly describes the phases of the treaty process as contemplated by the law of treaties. Sub-part B derives from these phases a view of the treaty process as consisting of three iterations (negotiation, signature, and entry into force), each presenting a set of payoffs in the form of a PD. Sub- part C sets forth and briefly compares two approaches to the problem of converting a nation's real-world behavior into a corresponding cooperate or defect action in each of the three iterations of the treaty process. The first approach focuses on the procedural mechanisms that indicate a nation's formal consent (or lack of consent) to various obligations at each of the three iterations in the treaty process, whereas the second approach focuses on the degree to which nations actually comply with their legal obligations in each iteration.

A. *The Law of Treaties*

A treaty is a textual specification of legal obligations at the international level. Just as a statute results from laws governing the legislative process and sets forth rules of law for citizens to follow, a treaty results from a general process governed by rules of law and sets forth particular rules for nations to follow. The body of legal rules generally governing the treaty process is known as the law of treaties.[110] The Vienna Convention on the Law of Treaties codifies long-standing customary law governing the validity and interpretation of treaties.[111] The Vienna Convention includes a cluster of rules addressing the pathway down which an idea for international cooperation must travel in order to become a full-blown treaty.[112] These rules specify that a nation first becomes partially bound by a treaty after signing the text at the conclusion of international negotiations, or by registering its accession to, that treaty.[113] A nation becomes fully bound by the text after that nation and a sufficient number of other nations have ratified, or acceded to, that text.[114]

Article 26 of the Vienna Convention sets forth in legal terms the center of gravity of the treaty process: "Every treaty in force is binding upon the parties to it and must be performed by

---

109

110

111

112

113

114

them in good faith."[115] The underlying idea, often described by the phrase *pacta sunt servanda*, is that nations indicating their full consent to a treaty are affirmatively bound to comply in good faith with each and every term in that treaty once that treaty has entered into force. A nation may indicate its full consent to be bound by a treaty by a variety of means, but ratification--a formal communication by the head of state of a signatory nation--is the most common means of indicating such consent.[116] Accession is a similar means of expressing full consent but involves a statement from a nation that was not a signatory to the treaty.[117] Entry into force is typically defined as occurring upon a sufficient number of ratifications or accessions.[118] An individual nation's ratification of, or accession to, a treaty is thus typically a conditional commitment by an individual nation to comply in full with all of the obligations set forth in a treaty. Once the aggregate ratifications of, or accessions to, a treaty by a sufficient number of nations occurs, then the condition for each nation's commitment is satisfied and a full-blown obligation commences upon the treaty's entry into force, requiring all ratifying or acceding nations to comply with each and every term of the treaty in good faith. Those nations that ratify or accede after the expression of full consent that triggers initial entry into force are typically fully bound--that is, the treaty enters into force with respect to them--within a fixed number of days from each nation's ratification or accession.[119]

The mere signature of a treaty text by a nation's authorized representative is typically not sufficient to indicate full consent.[120] Signature is nonetheless an act of some import in the treaty process. By the signatures of their representatives, nations indicate that a particular text is the authentic embodiment of the fruits of the negotiations.[121] A signatory is also obliged to comply with any provisions clearly intended to apply immediately, such as clauses expressly describing substantive obligations to be undertaken before the treaty enters into force, or setting forth the procedures that the parties are to use in binding themselves fully to all of the substantive obligations of the agreement.[122] More generally, a signatory must refrain from "acts which would defeat the object and purpose of [the] treaty."[123] Signature thus indicates partial consent, but by no means full consent, to abide by the terms of a treaty.

Even before a treaty is open for signature, nations must engage in negotiations to agree upon the relevant terms. The law of treaties implicitly prescribes certain rules of conduct for such negotiations by allowing a nation to invalidate its full consent to a treaty if that consent was induced by fraud or coercion or was procured by the corruption of its representative by another nation.[124] (Additionally, any nation may avoid its obligations under a treaty if the agreement

---

115

116

117

118

119

120

121

122

123

124

"conflicts with a peremptory norm of general international law."[125]) These implicit prescriptions dovetail with a nation's general obligation under international law to conduct its international legal affairs in good faith.[126] Although "good faith" may be an inherently vague concept, little argument is required to suggest that fraud, coercion, and the corruption of a nation's agents are not actions taken in good faith. One may therefore infer that nations have an obligation to avoid fraud, coercion, and corruption--and whatever else may constitute bad faith--in the conduct of their treaty negotiations.[127]

B. *The Treaty Process as an IPD: Iterations and the Prisoner's Dilemma*

The remainder of this part uses the international legal process described above to argue that the treaty process constitutes three iterations of an IPD and that one may derive from that process some concrete definitions of iterations and actions. The present sub-part argues that the treaty process consists of three well-demarcated iterations and that each iteration presents nations with a payoff structure that is a PD. The next sub-part focuses on how to generate a correspondence between real-world behavior connected with the treaty process and the dichotomous actions (cooperate or defect) of the IPD. By allowing the translation of both iteration and action from the real world to the game-theoretical world of the IPD, the treaty process can provide a starting place for an effort to move neo-liberal Institutionalism from metaphor to model.

The fundamental idea advanced in this sub-part is that the treaty process consists of a sequence of three iterations--negotiation, signature, and entry into force--and that each iteration presents nations with a PD involving the choice to cooperate with or defect from the underlying process of international cooperation. The starting point for the analysis is the acceptance of the argument, made by both neo-liberal Institutionalists and Realists, that international cooperation presents nations with a PD. I take as essentially axiomatic the idea that international treaties effectuate international cooperation, although there are arguments against such an axiom.[128] If treaties are a means to international cooperation, and if international cooperation presents nations with a PD, then it follows that the choice about whether to comply with a treaty's terms also presents nations with a PD.

This conclusion does not mean, however, that the treaty process necessarily presents an *iterated* PD. International relations is not inevitably a fractal structure: a process in international relations may *in the aggregate* present payoffs that constitute a PD while one *component* of the process presents a game of Chicken, another component a PD, and a third component some other type of game. To conclude that the treaty process constitutes an IPD, one must first propose a definition of the relevant iterations (which I do below in sub-part 1) and then determine whether the payoffs that a nation faces at *each* iteration constitute a PD (which I do in sub-part 2).

125

126

127

128

1. *What are the Iterations in the Treaty Process?*

As a result of its formal actions at three successive points in time, a nation participating in the unfolding treaty process takes on three successive obligations. By participating in negotiations, a nation becomes obliged to conduct those negotiations in good faith. By subsequently signing (or acceding to) a treaty, a nation becomes obliged to refrain from defeating the object and purpose of the treaty.[129] Upon the subsequent entry into force of a treaty, a nation that has ratified or acceded to that treaty becomes obliged to adhere in good faith to all of the treaty's terms until that treaty terminates. Except with respect to these three events, the international legal obligations of a given nation with respect to a particular treaty generally do not change at any point during the treaty process.[130] Conversely, a nation's failure to participate in treaty negotiations, to sign a treaty, or to ratify or accede to a treaty means that nation is not bound by the legal obligations imposed by the respective phases.

These successive stages of obligation suggest a three-iteration structure to the treaty process. The "negotiation" iteration begins with the commencement of negotiations on the relevant treaty and ends just before the treaty's signature. The "signature" iteration begins with the treaty's signature and ends just before the treaty's entry into force. The "entry-into-force" iteration begins with the treaty's entry into force and ends with the treaty's termination. Each phase requires nations to undertake a distinct formal activity: the sending of a national delegation to negotiations, the signature of the treaty, and the transmission of a ratification or accession. These activities are not only distinct but also independent; that is, participation in any one phase does not imply participation in any other phase. A nation can send a delegation but refrain from signing or from ratifying or acceding to the treaty; a nation may refrain from sending a delegation but later swoop silently into the signing ceremony and have its representative sign the draft; or a nation may refrain from signing an agreement but later, through accession, indicate its intention to be fully bound.

Each of the three iterations involves not only a distinct *formal* activity but also a distinct and independent *substantive* obligation: to negotiate in good faith, to refrain from defeating the object and purpose of the treaty, and to obey affirmatively all terms of the treaty in good faith. These substantive obligations therefore involve distinct behaviors to be undertaken in compliance with those obligations.

The obligations of the negotiation iteration are clearly independent of obligations tied in some way to the substance of the treaty. A nation could lie, threaten, and bribe in the negotiation phase, but then obey every clause in the resulting treaty to the letter, thereby fulfilling its obligations in the signature and entry-into-force iterations. Conversely, a nation could behave with perfect candor and propriety during negotiations but nonetheless either defeat the object and purpose of the treaty after signature or fail to comply with the treaty terms after entry into force.[131] The signature iteration and the entry-into-force iteration have obligations that are

---

129

130

131

independent not only from those of the negotiation iteration but also from one another. Both iterations involve obligations derived from the treaty itself, but compliance (or non-compliance) with one set of obligations does not imply compliance (or non-compliance) with the other set of obligations. It is not difficult to imagine a nation refraining from activities that defeat the object and purpose of a treaty but then failing to comply in good faith with all terms of the agreement upon entry into force. A treaty might, for example, have as its object and purpose the preservation of the ozone layer. Such a treaty might include among its mandates not only a complete cessation of the production of ozone-depleting chemicals but also contributions to an international fund for research into the effects of ozone depletion. Assume that the production ban is necessary to achieve the object and purpose of the treaty but that the contributions are not. A nation might refrain entirely from the production of ozone-depleting chemicals during the signature iteration (and even during the entry-into-force iteration) but fail to contribute to the international fund. Such a nation would thereby comply perfectly with its obligations in the signature iteration, but fail to comply with its obligations in the entry-into-force iteration.

Conversely, a nation might defeat the object and purpose of a treaty immediately after signature but comply with the treaty after its entry into force. Immediately after signing a treaty that bans testing of a particular weapon, for example, a nation might test such a weapon and thereby defeat the object and purpose of the treaty;[132] nonetheless, once the treaty entered into force, a nation might not only refrain entirely from such tests but also comply with every other term in the treaty.

A combination of international practice and international law makes these three phases--negotiation, signature, and entry into force--not only distinct and independent, as just discussed, but also strictly sequential. Negotiation precedes signature as a matter of definition. Nations must work out their differences in negotiations before they can indicate, through signature, their assent that the resulting text both indicates the fruit of their negotiating labors and sets forth an object and purpose that signatories must not defeat with their subsequent actions. Given that nations typically make no collective modifications to the signed text between signature and entry into force, the signature of a treaty typically signals the end of negotiations on that treaty's text.[133] Finally, signature precedes entry into force as a matter of international law. Nations first indicate their view that a text authentically expresses the fruits of negotiation through signature, and then later indicate their full consent to that text through ratification.[134] (A nation may also refrain from signing a treaty but indicate its full consent through accession.) Over time, enough ratifications or accessions accumulate for the treaty to enter into force--an event that occurs at a single, clearly defined moment. The entry-into-force iteration thus has a clear beginning, and the signature iteration has a clear conclusion.[135] Finally, entry into force precedes termination as a matter of definition: there can be no termination of a treaty unless there is a treaty in force to be terminated. The various iterations thus proceed in strict sequence.[136]

132

133

134

135

136

2. *Does a Nation Face a Prisoner's Dilemma at Each Iteration in the Treaty Process?*

As mentioned above, I agree with the neo-liberal Institutionalists' characterization of international cooperation as a PD with iterations. As long as a treaty is a means of effectuating international cooperation, a nation's decision to comply or not to comply with a treaty in force is a PD. When, however, one moves from the issue of how to characterize participation in a cooperative endeavor (that is, compliance with a treaty in force) to the issue of participation in a cooperative endeavor involving multiple iterations, then one must revisit the issue of exactly what incentives confront the participants at *each* iteration. One could, after all, define iterations in a process that reflects a PD as a whole without necessarily defining iterations that each reflect a PD. Imagine, for example, an overarching process that could be disaggregated into a two-iteration game. Assume that the payoffs in the first iteration are much smaller than the payoffs in the second and that the second iteration constitutes a PD. The first iteration could then take the form of a wide variety of games while, because of the small contribution to the overall payoffs made by that first iteration, the overarching process would remain a PD.

I note at the outset of this analysis that the definition of each iteration varies from iteration to iteration, as do the real-world behaviors available within each iteration. The legal implications and formal indicia of the negotiation iteration differ from those of the signature iteration; the obligations and formalities of signature differ from those involved in ratification and entry into force; and negotiation and entry into force also (unsurprisingly) differ from one another in form and substance. These differences necessitate defining each iteration separately, as I have done above.

These differences between a stylized IPD and an effort to translate a real-world phenomenon into an IPD are hardly unique to the treaty process, of course. One might, for example, consider a series of votes on an annual budget by a member of Congress, or the actions of adversaries possessing nuclear weapons, as potential candidates for representation by an IPD. The congressional votes are likely to be on budgets that are not identical to one another, just as the treaty process involves iterations that involve different formal actions or underlying legal obligations. The payoffs from those congressional votes are likely to vary from year to year with a variety of changes--in the particular bargains offered to the member in exchange for his or her vote, in the amount of attention that the electorate is paying to budgetary matters, and so forth--just as the payoffs to a nation from failing to participate in negotiations may differ from the payoffs to a nation from failing to comply with obligations to which it has indicated its full consent. A series of decisions between adversaries possessing nuclear weapons is even more likely to involve different actions over the years (a defect action in the early 1950s could not involve decisions about launching ICBMs, as they did not then exist) and different payoffs (as the technologies of the adversaries change and affect the deaths and damage that will result from decisions by one or more of the adversaries to use nuclear weapons).

One should also note as a preliminary matter that the divergence between the stylized IPD presented by game theoreticians and the treaty process does not make the treaty process simply an elaborate version of a single-move game. As discussed above, each iteration involves a distinct set of legal obligations, and the action taken in each iteration is independent from the

action taken in other iterations.

Having discussed these preliminary matters, I turn now to the determination of whether each of the three iterations in the treaty process presents nations with a PD, and thus whether one may fairly characterize the treaty process as an IPD.

Of the three relevant iterations, the signature iteration and the entry-into-force iteration are each almost certain to present nations with a PD as long as international cooperation itself is a PD in the specific case under consideration.[137]   The negotiation iteration is also likely to be a PD, but I base this conclusion upon general principles relating to the nature of negotiations rather than upon general principles assumed by IR theory. I now work backwards in the analysis from the entry-into-force iteration through the signature iteration to the negotiation iteration.

a. The Entry-Into-Force and Signature Iterations

Compliance in good faith with all of the terms of a treaty in force (that is, compliance during the entry-into-force iteration) is clearly a PD. If international cooperation is a PD, then one can hardly call the treaty process a means of effectuating international cooperation without the final legal result of the treaty process also being a PD.

Does a nation's decision about whether to refrain from defeating the object and purpose of the treaty during the signature iteration also qualify as a PD? If one views such an obligation as some kind of lesser version of the obligation that exists with respect to treaties in force, then it is natural to view the decision about whether to refrain from defeating the object and purpose of the treaty as a lesser version of the PD.

Consider as an example a treaty that prohibits commercial whaling in international waters with the purpose to prevent the extinction of various species of whales.[138]   As with other common-pool resource questions, compliance with such a treaty presents a PD. A nation that conducts whaling while all other nations do not benefits from that whaling unhindered by competitors; such a nation might prefer to preserve the hunted species of whale, but only if the cooperation of all other nations could be guaranteed. The obligation not to defeat the object and purpose of the treaty is essentially a less strictly binding version of the obligation to obey all the terms of the treaty. If defeat of the object and purpose of the treaty requires so much whaling as to extinguish one or more of the relevant species, then the relevant activity for defeating the object and purpose of the treaty would be intensive whaling--essentially the same activity that defeats compliance with all of the treaty's terms.[139]   When the treaty as a whole effects the solution to a PD, it is not difficult to imagine that refraining from the defect of its object and purpose--its essence--would also present a PD.

b. The Negotiation Iteration

137

138

139

With respect to negotiation, one might argue that less some direct connections between disobedience to the legal rules of the relevant iteration and the overall PD of treaty-effectuated cooperation make the negotiation iteration a PD. However, the good faith obligation of the negotiation iteration is less directly tied to the substance of the treaty that results than are the obligations of the signature and entry-into-force iterations. In the absence of a direct link between the negotiation iteration and the neo-liberal Institutionalists' assumption that international cooperation is an IPD, the conclusion that the negotiation iteration is a PD requires some additional assumptions. Immediately below, I discuss an example in which the decision about whether to comply with one's obligations in the negotiation iteration does present a PD. I also highlight the assumptions necessary for the negotiation iteration generally to be a PD and discuss briefly the plausibility of those assumptions.

Consider the following example. Two nations, *A* and *B*, are negotiating a bilateral environmental treaty aimed at reducing cross-border air pollution. The parties have already struck all value-maximizing bargains available to them in light of the transactions costs that they face. Nation *A* seeks an exemption for auto-manufacturing plants from provisions of the treaty currently on the table. Nation *B* similarly seeks an exemption for brick kilns. Nation *A* possesses a significant number of automobile plants near its border with nation *B* but few brick kilns, while nation *B* has many brick kilns near the *A-B* border but few auto-manufacturing plants.

Assume that all costs and benefits to either national government can be expressed in a single and inter-comparable unit, which we shall call the "chip" in deference to the frequency with which international negotiations are analogized to poker games. If nation A can obtain the auto-makers' exemption, nation A will gain 150,000 chips. (This increase might represent greater freedom for nation *A* to make concessions in these or other negotiations with nation *B*, a greater likelihood that the treaty will be ratified by domestic legislatures, greater campaign contributions to nation *A* from auto-makers, or any other of a variety of potential benefits.) For nation *B*, granting the exemption will lead to a cost of 120,000 chips. (This cost might represent not only the converses of the benefits listed above but also the health costs of greater trans-border pollution, the damage to buildings or agriculture from more acidic air, the loss of political face from conceding to the other side, and so forth.) With respect to the inclusion of the brick-kiln exemption in the treaty, nation B would reap benefits equivalent to 75,000 chips, while nation *A* would suffer costs of 60,000 chips.[140] Let us also assume that the transaction costs standing between the parties and a freely negotiated resolution of the exemptions can be measured at exactly 200,000 chips for each exemption.[141]

Against this backdrop, both nations are contemplating an act of bad faith. For nation *A*, that act of bad faith is a bribe equivalent to 140,000 chips (presumably converted into the local currency, or at least to Swiss francs) to an influential representative of nation *B*; the bribe will reliably lead to inclusion of the auto-manufacturing exemption in the treaty. For nation *B*, the

---

140

141

contemplated act of bad faith is the construction of a dazzling computer-graphics program to display data on the costs to brick-makers of complying with the existing treaty terms. Nation *B* knows that the data is a wild exaggeration of actual costs, and also knows that nation A will be so dazzled by the program and its data that nation *A* will agree to the exemption for brick-makers. Construction of the program will cost nation *B* the equivalent of 40,000 chips.

The situation described above presents nations *A* and *B* with a PD. As always, the payoffs are important only in a relative sense, so let us choose as a baseline the cooperate-cooperate outcome (that is, nation *A* does not bribe and nation *B* does not dazzle with false data) and assign these payoffs a value of 0 chips for nation *A* and 0 chips for nation *B*. If nation *A* negotiates in bad faith (that is, bribes the representative of nation *B*) while nation *B* negotiates in good faith, then nation *A* will net 10,000 chips in comparison to the baseline payoffs: the bribe costs nation *A* 140,000 chips but garners nation *A* gross benefits from the auto-maker's exemption of 150,000 chips, for a total net benefit to nation *A* of 10,000 chips. If nation *A* and nation *B* both negotiate in bad faith, then nation *A*'s net change in position is a loss of 50,000 chips: nation *A* will gain 150,000 chips from the resulting auto-makers' exemption, but it will need to spend 140,000 chips on the bribe and will additionally lose 60,000 chips from the brick-makers' exemption resulting from nation *B*'s computer-assisted dazzlings. If nation *A* negotiates in good faith, while nation *B* decides to go ahead with its scheme of dazzling graphics and fraudulent data, then nation *A* loses 60,000 chips from the resultant inclusion of the brick-makers' exemption in the treaty. The payoffs in chips for nation *A* are therefore ranked as follows, with nation *A*'s action listed first in each pair and the value of the payoff to nation *A* listed in parentheses: bribe/cooperate (10,000); cooperate/cooperate (0); bribe/dazzle (-50,000); cooperate/dazzle (-60,000). This hierarchy of payments, shown in Figure VI below, is a PD.

FIGURE VI: Bilateral Environmental Treaty as a PD[142]

		PLAYER B	
		COOPERATE	DEFECT
PLAYER A	COOPERATE	(0, 0)	(-60, 35)
	DEFECT	(10, -120)	(-50, -95)

The analogous calculations for nation *B* lead to the analogous conclusion that the situation described above is consistent with nation *B*'s half of a PD. If nation *B* negotiates in bad faith (that is, dazzles the representative of nation *A* with the fraudulent computer program) while nation *A* negotiates in good faith, then nation *B* will gain 35,000 chips in comparison to the baseline payoffs: the computer program and associated false data cost nation *B* 40,000 chips but

garner nation *B* gross benefits from the brick-makers' exemption of 75,000 chips. If nation *B* instead negotiates in good faith, while nation *A* decides to go ahead with its bribery scheme, then nation *B* loses 120,000 chips as a result of the inclusion of the auto-makers' exemption in the treaty. If nation *A* and nation *B* both negotiate in bad faith, then nation *B* will suffer a net loss of 95,000 chips: it will spend 40,000 chips on computer-assisted dazzlings and lose 120,000 chips from the inclusion of the auto-makers' exemption, while gaining only 75,000 chips from the inclusion of the brick-makers' exemption. The payoffs in chips for nation *B* are therefore ranked as follows, with nation *B*'s action listed first in each pair and the value of the outcome to nation *B* listed in parentheses: dazzle/cooperate (35,000); cooperate/cooperate (0); dazzle/bribe (-95,000); cooperate/bribe (-120,000).[143] As with the calculations with respect to nation A, this hierarchy of payoffs for nation *B* is a PD and is shown in Figure VI.

The exact numbers of this extended example are arbitrary, but the general relationships among those numbers are not. For both nation *A* and nation *B*, the negotiation iteration presents payoffs in the rank-ordering of the PD. However, not just any set of numbers will do. If one varies some of the amounts by a sufficient amount, the resulting situation is not a PD. Note first that the numbers chosen above involve a situation in which an act of bad faith has a cost lower than the gains that it reaps for the party acting in bad faith. The bribe costs 140,000 chips; the result is a gain to the bribing nation of 150,000 chips. The computer graphics and associated data cost the equivalent of 40,000 chips; the result is a gain to the nation employing them fraudulently of 75,000 chips. Imagine instead that the nearly incorruptible representative of nation *B* required a bribe of 1,000,000 chips in order to support the auto-makers' exemption, or that the construction of the relevant computer program and associated data required an expenditure equivalent to 200,000 chips. In these cases, it is irrational for nation *A* or nation *B* to pursue their respective acts of bad faith: the gain from doing so (150,000 chips for nation *A* from the auto-makers' exemption and 75,000 chips for nation *B* from the brick-makers' exemption) are much less than the costs to do so (1,000,000 chips for nation *A* and 200,000 chips for nation *B*). No rational bargainer would spend more than it receives in return. If there were no other actions of bad faith available that cost less to employ than the gains that they reaped, then no rational nation would be tempted to employ bad faith. The defect-cooperate option for a nation (where its own action is listed first in the pairing) would then have a lower payoff than the cooperate-cooperate option, and so the negotiation iteration would not present a PD. (The hierarchy of payoffs necessary to constitute a PD requires the defect-cooperate payoff to be higher than the cooperate-cooperate payoff.) The treaty process as a whole would therefore not be an IPD of three iterations.

Second, note the amount of the transaction costs inhibiting free bargaining for each exemption: 200,000 chips in a situation where the total gross benefits to the party desiring a particular outcome are at most 150,000 chips. Suppose instead that the monetary equivalent in those transaction costs were only 1,000 chips. In that case, it might well be wisest for each nation simply to spend the 1,000 chips necessary to tear down the transaction-cost barriers and then negotiate a good-faith deal with the other nation, rather than to attempt to obtain the favored exemption through bad faith behavior. Suppose, for example, that nation *A* spent 1,000 chips to

143

dismantle the transaction costs relating to the auto-makers' exemption and then offered nation *B* 130,000 chips in concessions to accept that exemption. Nation *B* should accept, since 130,000 chips is an amount greater than the 120,000 chips in costs to nation *B* from accepting the exemption. Nation *A* would then have obtained the auto-makers' exemption by spending 1,000 chips in dismantling transaction-cost barriers and 130,000 chips in concessions to nation *B*. That total expenditure of 131,000 chips is less than the 140,000 chips necessary to obtain the auto-makers' exemption through bribery. If a good-faith means of obtaining the exemption is available for a lower cost than the bad-faith means, then the payoff from the cooperate-cooperate option will be greater than the payoff from the defect-cooperate option. In this example, the payoff in chips to nation *A* from the cooperate-cooperate option is 19,000 (150,000 minus 131,000) while the payoff to nation *A* from the defect-cooperate option is only 10,000 (150,000 minus the 140,000 chips necessary to effectuate the bribe). As discussed above, the cooperate-cooperate option must have a smaller payoff to a nation than does the defect-cooperate option if the negotiation iteration is to present a PD. Under this new set of assumptions, therefore, the treaty process would not present a three-iteration IPD.

Although the above paragraph simply lowers the transaction costs and demonstrates a case in which the result is a negotiation iteration that does not present a PD, one may be more precise about the effect of transaction costs on the payoff structure of the negotiation iteration. The net benefits of acting in bad faith provide one side of the comparison; the net benefits of paying the necessary transaction costs and then reaching a bargain in good faith provide the other side of the comparison. If the net benefits of acting in bad faith are greater than the size of the difference between the parties' valuations of the outcome that remains after the parties have overcome any transaction costs, then acting in bad faith is a more attractive option for one party than proceeding in good faith, and the negotiation iteration will be a PD. However, if the net benefits of acting in bad faith are less than the size of the post-transaction-cost difference, then good faith is cheaper than bad faith even for the party contemplating bad faith, and the negotiation iteration will not present a PD.

Thus far, this analysis has proceeded by example and by examining the necessary conditions for the negotiation iteration to present a PD. The final portion of the analysis of the likelihood that the negotiation iteration will present a PD involves the much more subjective project of deciding whether the necessary conditions discussed above will in fact be met. One may summarize the conditions discussed in the paragraphs above in a single, if somewhat complex, condition: for each party to a negotiation, some method of bad faith must provide the party contemplating bad faith with positive net benefits that exceed the difference between the two parties' evaluation of the resulting outcome, even after subtracting from this difference the transaction costs of a good-faith bargain reaching the same outcome. If the costs of effectuating bad faith are very low relative to the benefits that they can achieve, and the transaction costs of reaching a good-faith bargain are very high, and the difference between the two parties' valuation of the outcome is very small, then bad faith will present a desirable means to the party contemplating bribery or coercion or corruption or fraud. If the situation is the same with respect to the other party's contemplation of bad faith, then the negotiation iteration will present the parties with a PD. If instead the costs of effectuating bad faith are astronomical, and the transaction costs of reaching a good-faith bargain are infinitesimal, and the difference in the parties' valuation of the outcome is gigantic, then bad faith will not be a rational alternative to

good faith. The party contemplating bad faith will then not see such a defection as a rational alternative to good faith, and so the parties will not face a PD in the negotiation iteration.

It would of course be ludicrous to attempt a precise quantitative evaluation of every international treaty negotiation to determine whether the conditions for considering the negotiation iteration to be a PD are met. I hope, at least, that by identifying those conditions, others will be able to make more informed judgments for themselves. In addition, however, I would venture the personal opinion that the conditions necessary to make the negotiation iteration a PD are at least plausibly met by a variety of treaty negotiations. The presence of high transaction costs is perhaps the condition most likely to be met. International negotiations take place against a backdrop of significant linguistic and cultural differences in almost every case, and often one must include radically different political orientations in the mix of factors likely to make communications costly or ambiguous. On topics involving issues of economics or technology, some nations are clearly much more competent to undertake sophisticated analyses than other nations; the difference in capabilities may make for difficulties in communication--unless nations possessing fewer analytical resources simply take the word of other nations, a possibility that seems unlikely. The views of those not typically present at the table--legislatures, interest groups, the public generally--may be relevant but costly to determine as well.

To the degree that informational or cultural asymmetries result both in high transaction costs and in differential valuations of outcomes, there will be a tendency for situations with high transaction costs (which *favor* a correspondence between the negotiation iteration and the PD) to occur in conjunction with situations involving large differentials in valuation (which *disfavor* a correspondence between the negotiation iteration and the PD). An inherent counter-balancing may therefore occur. Differential valuations may also be more likely to occur in political negotiations, as opposed to negotiations involving market transactions, because political negotiations quite frequently involve the exchange of "goods" that have very few bidders or are inherently difficult to value. For example, a limited number of entities typically make plausible "bids" to enter a military alliance and the redistributional aims of governments, as with World Bank activities, may be very difficult to value monetarily. In such cases, no reference to market prices is possible to narrow differences between parties. To the degree that difficulties in monetizing exchanges make differential valuations between parties more likely, international political negotiations are more likely to involve differential valuations than exchanges between two parties engaged in a market-oriented transaction--even if the market-oriented parties have the same asymmetries of capability and culture that obtain in the analogous international political transaction.

Finally, there is the cost of bad-faith actions relative to the benefits that they yield. The greater the surplus of benefit over cost, the greater the likelihood that bad faith is rational, and thus the greater the likelihood that the parties face a PD in the negotiation iteration. To a degree, any broadening of the available means for pursuing a given end (that is, from good-faith actions alone to actions taken in both good faith and bad faith) will increase the possibility of making available some low-cost but high-yield activities. Some bad-faith activities, such as coercion or corruption, also seem likely to fall upon states or individuals vulnerable to certain opportunistic behavior--the nation with a military Achilles' heel, or the negotiator with a just-discovered secret

or a recent financial reverse--that may well yield results far beyond the effort necessary to effectuate them.

In my view, the assumption that the negotiation iteration is a PD is plausible. Such an assumption is hardly to be taken for granted, however, and the strength of the other assumptions in the chain of logic that binds the treaty process to characterization as an IPD certainly does make the nature of the negotiation iteration the most contestable link. In the end, however, as with all the factors involved in deciding whether the negotiation iteration presents a PD, one must simply judge for oneself.

3. *The Source of Iterations: Comparing the Treaty Process to the Enforcement Process*

The substantive terms of a treaty may serve as a short-hand method for defining international cooperation. One could start with the obligations set forth in a treaty, choose an iteration of some arbitrary length (say, a year), and examine whether or not nations comply with those obligations year after year.[144] The determination of just what the terms of the treaty mean--indeed, whether the text actually sets forth any obligations at all--would then presumably be a matter suitable for the attention of international lawyers.

It is worth noting, however, that the analysis above suggests a different definition. I argue that the various stages of the treaty process-- negotiation, signature, and entry into force--are the relevant iterations. The definition of those iterations flows from the treaty process itself, rather than from an arbitrary choice made by the would-be analyst. Judgments about which forms of national behavior correspond to cooperate actions thereby stem not only from an understanding of what constitutes compliance with the terms of a particular treaty but also from an understanding of the procedural aspects of the treaty process. The relevant behavior of nations includes not only their behavior after a treaty enters into force but also their behavior during pre-signature negotiations and during the post-signature (but pre-entry into force) period.

C. *The Treaty Process as an IPD: Actions*

This Article has argued that the beginning and end of each of the three iterations in the treaty process are relatively easy to determine, and that each iteration is not only procedurally but substantively distinct from the others. The Article has also argued that each iteration presents a nation with a PD. The treaty process, then, may be viewed as three sequential iterations in an IPD.

Moving neo-liberal Institutionalism from metaphor to true theory requires a clear definition of "iteration," but such a definition does not by itself complete the transition. At the very least, one must also construct a method for translating international behavior into game-theoretical actions. The treaty process can be of assistance in this endeavor as well.

At each of the three iterations of the treaty process, there are two ways to advance an

---

144

essentially dichotomous characterization of international behavior that corresponds to the two actions available in the typical presentation of the PD. One method for determining how international behavior translates into game-theoretical actions focuses on formal international legal behavior; the other focuses on compliance with substantive standards of international law. From the formally-oriented perspective, cooperate actions would consist of participation in negotiations (in the negotiation phase), the signing of the treaty (in the signature phase), and the ratification of the treaty (in the entry-into-force phase). From the compliance-oriented perspective, a cooperate action consists of conducting negotiations in good faith (in the negotiation phase), refraining from activities that defeat the object and purpose of the treaty (in the signature phase), and affirmatively complying in good faith with all provisions of the treaty (in the entry-into-force phase).[145] Under either method, the treaty process provides definitions of iteration and actions that offer some hope of matching the metaphorical abstractions of the PD to real-world situations with sufficient fidelity to serve as the foundation for a true theory.

1. *A Formally-Oriented Perspective on the Correspondence Between Real-World Behavior and Game-Theoretical Actions*

One perspective on the correspondence between international behavior and game-theoretical actions emphasizes the formalistic behavior of parties. At each iteration, a party either has or has not taken certain formal, readily observable steps: sending a representative to international negotiations, for example, or signing, ratifying, or acceding to a treaty. A correspondence between these formal real-life actions and the two game-theoretical actions of the PD is not difficult to construct. Sending a representative, signing a treaty, and ratifying or acceding to a treaty would, in this view, correspond to a cooperate action for the negotiation, signature, and entry-into-force iterations, respectively. Accordingly, the failure to send a representative, to sign a treaty, or to ratify or accede to a treaty would correspond to a defect action in the relevant iteration.

2. *A Compliance-Oriented Perspective on the Correspondence Between Real-World Behavior and Game-Theoretical Actions*

Another perspective on the correspondence between actual international behavior and game-theoretical actions emphasizes the behavior of parties with respect to the obligations incurred at each iteration rather than with respect to the formal indications of participation or consent. At each iteration, a party assumes a particular set of obligations with which they may or may not comply. A party that sends a representative to negotiations must negotiate in good faith; a party that signs a treaty must refrain from taking actions that defeat the object and purpose of the treaty; and a party that ratifies or accedes to a treaty must affirmatively comply in good faith with all the treaty's terms after entry into force. In a compliance-oriented perspective on the treaty process, the fulfillment of these obligations corresponds to a cooperate action for the negotiation, signature, and entry-into-force iterations, respectively. Failing to comply with the obligations involved in a particular iteration corresponds to a defect action in the relevant iteration.

145

3. *Comparing the Formally-Oriented and Compliance-Oriented Perspectives on the Correspondence Between Real-World Behavior and Actions in the IPD*

This Article has pointed out some natural correspondences between the treaty process, on the one hand, and the stylized IPD that lies at the heart of the Institutionalist version of IR theory, on the other. Below I explore in more detail the difficulties of operationalizing these correspondences.[146] One cannot fairly judge the utility of the treaty-oriented effort to operationalize the IPD until one has wrestled with these complexities. Nonetheless, some comparison of the formally-oriented perspective with the compliance-oriented perspective is useful even at this preliminary stage.

The natural reaction of the IR theorist to the formally-oriented perspective is that "merely formal" adherence to the particulars of an international legal process does not reveal much about actual behavior. The portion of neo-liberal Institutionalism that draws upon Realism takes a skeptical attitude towards international legal obligations. To a neo-liberal Institutionalist, the mere statement of an intention to negotiate, or of a commitment to comply with a treaty, means very little. However, if there were truly no cost attached to the formally-oriented actions of adherence to the cooperative goals of the treaty process, then nations would always undertake such actions. Why, after all, should a nation not at least appear cooperative by sending delegates to negotiations, or by signing, ratifying, or acceding to a treaty, regardless of its true intentions?

The fact of the matter is that many nations refrain from taking one or more of the formally-oriented actions corresponding to cooperative behavior. Not every nation sends delegates to every potentially relevant international negotiation aimed at producing a treaty text; not every nation signs the resulting treaty; and not every nation ratifies or accedes to the treaty. One must therefore infer that some cost attaches to these formally-oriented actions, whether in monetary terms (the cost of sending delegates to international negotiations, for example), in domestic political terms (as it was, for example, in decisions by the United States not to ratify the Versailles Treaty or SALT II), or in international political terms (an unfavorable reaction from other nations upon later failing to adhere with the relevant international legal obligations). In fact, there are even occasions when some nations wish very much to become formally bound by a treaty but have their wishes actively resisted by other nations, already formally bound, whose consent is necessary for the would-be member to join. China wishes to become formally bound by the treaties constituting the World Trade Organization, for example, and Romania wishes to join NATO. In each of these cases, however, present members of the WTO and NATO wish to prevent, or at least delay, the increases in membership. If nations will wage costly political battles over formal membership in such treaty-constituted organizations, then formal participation in those organizations must mean something.

Regardless of the precise implications of formal participation considered in isolation, the formally-oriented view possesses the important advantage of clarity. Whether a nation has sent delegates to a negotiating session, or has signed, ratified, or acceded to a treaty is relatively easy

---

146

to determine, both from the perspective of other nations and from that of the would-be empiricist. The compliance-oriented perspective, on the other hand, is more difficult to apply. Whether a nation has complied in good faith with all terms of a treaty, for example, is likely to involve a much more complex inquiry than the issue of whether the ratified treaty has entered into force. The determination of entry into force typically involves glancing at a treaty provision concerning the number of nations that must ratify or accede before entry into force, and then counting down a list of ratifying and acceding nations to see if that number has been reached. The determination of compliance, in contrast, requires comparing the actual behavior of each fully bound nation with the substantive rules of a particular treaty, which rules often involve complex activities such as the flow of international trade or the deployment of military weapons.

In the formally-oriented perspective, the relevant national behaviors are quite similar with respect to a given phase, regardless of the treaty involved. The compliance-oriented perspective, in contrast, focuses on national obligations, a category that comprises a wide range of behaviors. The relevant behaviors in the entry in force iteration are often specific to the rules of a particular treaty--one treaty may involve the production of ozone-depleting substances,[147] another the destruction of chemical weapons,[148] a third the extradition of a foreign power's citizens[149]--and thus may require an understanding of the technical, legal, and political context underlying each treaty evaluated. "Object and purpose" likewise varies from treaty to treaty, and determining whether behavior violates this standard may be quite difficult to verify. This determination is further complicated when a nation's conduct is not readily observable: activities may be widely dispersed, such as international whaling,[150] or subject to concealment, such as with the production of weapons of mass destruction.[151]

Nonetheless, many analysts of international relations are likely to believe that actual behavior relating to compliance *vel non* with various rules is a more meaningful index than one focused on particular formal actions.[152] Certainly, actual adherence to various international legal obligations could prove a much more costly set of activities, and thereby in some sense a more reliable indicator of cooperativeness, than simply indicating an obligation to adhere. If one accepts the additional proposition that the treaty process has a goal or purpose, and that compliance with treaties aids in the realization of that purpose, then purely formal indicia of participation in the treaty process bear less significance to the inquiry than a more subjective analysis of a nation's actual actions with respect to its obligations.

The amount of actual compliance with treaties that have entered into force has, of course, been a perennial source of debate among scholars of international relations and international law, and has yielded no indication of any consensus.[153] One should not, however, allow the difficulty

---

147

148

149

150

151

152

153

of determining treaty compliance to drive all hope from the scene--or at least not to serve as an argument of any special force as applied against the use of *treaty* compliance as a gauge of international cooperation. Neo-liberal Institutionalists are interested in international cooperation. *Any* study of cooperation must involve a determination as to whether nations are actually cooperating, and *any* study of important cooperative efforts must examine nations with large incentives to defect and to hide their defections. Unless one wishes to abandon or trivialize neo-liberal Institutionalism--or, for that matter, the study of international cooperation from any perspective--one must confront the difficulty of characterizing national actions as cooperative or uncooperative, which in turn requires one to study nations' compliance with *some* rule or norm of international behavior. An examination of treaty compliance, as opposed to compliance in other milieus, in fact, presents relatively *few* difficulties in specifying what constitutes cooperation. Treaties set forth express rules, often in great detail, and the formalities of the treaty process generally provide the analyst with an easy way to determine, at least to a first approximation, just who is supposed to be participating in the cooperative effort.

As an illustration, consider the question of whether the United States and the Soviet Union cooperated during the Cold War to reduce the risks of nuclear war. During the 1970's and afterwards, several relevant arms-control treaties entered into force, leaving one to inquire whether the two nations complied with these treaties. Many of the relevant rules are quite specific. Consider the following provisions from the ABM treaty:

> [E]ach Party undertakes . . . not to deploy in the future radars for early warning of strategic ballistic missile attack except at locations along the periphery of its national territory and oriented outward.[154]

> The Parties agree not to deploy phased-array radars having a potential (the product of mean emitted power in watts and antenna area in square meters) exceeding three million, except as provided for in Articles III, IV and VI of the Treaty, or except for the purposes of tracking objects in outer space or for use as national technical means of verification.[155]

The issue of the signatories' compliance with these provisions highlights the utility of treaties, as opposed to other potential sources, for defining the contours of international cooperation. After the ABM Treaty had entered into force, the Soviet Union constructed a radar near the town of Krasnoyarsk at a site known as Abalakovo.[156] Note how concisely, and with what relatively objective support, one may argue that this radar is preemptively barred:

> [T]he Abalakovo radar did not exist at the time the ABM treaty was signed and is of a type generally used for early warning.

> . . .

---

[154]

[155]

[156]

The radar is located approximately 400 miles from the nearest Soviet border--the border with Mongolia. More important, rather than looking outward toward the Mongolian boarder, Abalakovo is oriented to the northeast across some 3,000 miles of Siberia.[157]

All available evidence contradicts the Soviet assertion that Abalakovo is designed primarily for space tracking or verification, although it might be able to track some future Soviet space-shuttle operations. Abalakovo does not appear to be principally designed for space tracking, nor does it add a significant new capability to existing Soviet space-tracking facilities. The Abalakovo radar faces northeast, a direction that would not allow it to be used for tracking satellite launches from Tyuratam or Plesetsk. (Soviet satellites orbit primarily at inclinations ranging from 47 to 83 degrees . . . .) Furthermore, if the Soviets had designed the radar for space tracking, they would have made the angle of the face more obtuse and the radar faces smaller.[158]

The same author also makes a judgment about the cooperativeness *vel non* of the Abalakovo radar in the absence of any governing treaty (an issue one might think of as the "strategic significance" of the Soviet action, or as the degree of cooperation or defection that the action might have indicated in the absence of a treaty).[159] This analysis occupies three full pages of text, not the half-dozen or so sentences comprising the analysis of compliance with the treaty.[160]

As long as compliance is a sign of cooperativeness, then, treaties have an advantage in allowing the precise specification of standards and thereby easing the task of judging whether actual behavior constitutes cooperation or defection for the purposes of the PD. This point is implied more broadly by a somewhat different question. Did the United States and the Soviet Union cooperate during the Cold War to reduce the risks of conventional war between their own forces? I am inclined to say that they cooperated less in this area than they did with respect to strategic nuclear arms, simply because the two nations signed no treaties regulating conventional weapons until much later in the Cold War than they did with respect to nuclear arms control.[161] This inclination reveals the utility of the treaty process in evaluating international cooperation. But were I to ask myself about each nation's behavior with respect to each area of armaments, I would face another problem that demonstrates the utility of treaties: absent a treaty on the subject, I have no clear way of determining how to judge whether the United States and the Soviet Union cooperated with respect to conventional arms control.[162] I would be forced to invent some fairly general "rules" based upon the general notion that arms races are uncooperative endeavors--a "rule," for example, that a large increase in the numbers of certain types of weapons, or intensive technological efforts respecting a class of weapons, would be uncooperative behavior. Nonetheless, I would be unable to advance any particular, quantitative

---

157

158

159

160

161

162

judgments as to what constituted (or would have constituted) cooperation with respect to conventional armaments.

IV. Complexities Relating to the Treaty Process and the Iterated Prisoner's Dilemma

Part III of this Article advanced the argument that the treaty process provides a useful and relatively objective method for operationalizing the concept of iteration in the context of international relations. It argued that the treaty process is divisible into three distinct phases readily viewed as three iterations and that the payoffs at each iteration constitute a PD (and thus that the treaty process presents an IPD), and further, that the actions available at each iteration fit naturally into the cooperate-defect dichotomy presented by the typical PD. Part III also argued that one might view the correspondence between game-theoretical actions and real-world behavior from either of two different (but not necessarily mutually exclusive) viewpoints, with one viewpoint emphasizing formal indicia relatively invariant from treaty to treaty and the other emphasizing compliance with the particular international legal obligations arising from the relevant treaty.

The present part of the Article focuses on some of the particularities and difficulties of generating a correspondence between the real world of the treaty process and the game-theoretical world of the IPD. Sub-part A discusses some complexities raised by the attempt to fit particular phases of the treaty process into a single iteration despite the fact that those phases unfold over time. Sub-part B addresses the additional complication that nations may selectively take on or comply with the obligations flowing from participation in the treaty process.

These complexities can arise with respect to any of the iterations in the treaty process, blurring the crisp abstraction of the treaty-process-as-IPD described in Part III. Though I do not argue the point further, I believe it likely that similar complexities would arise in any effort to operationalize the iteration, payoffs, and actions of the IPD in the real world of international relations. One may therefore view this part, like the previous part, as an attempt to apply the theories of neo-liberal Institutionalism to the real world. However, while Part III promoted international law as a basis for operationalizing the abstractions of IR theory, this part emphasizes the difficulties this approach entails.[163] Because the relevant background is international law rather than some other area of international politics, one may also view this part as a further effort to explore the nexus between international law and theories of international relations.

A. *Difficulties in Compressing Time Into a Single Iteration*

Activities in the real world unfold over time, but the choice of an action in the world of game theory occurs at a single point in time (or perhaps in some mathematical world so abstract as to be entirely timeless). Regardless of whether one adopts the formally-oriented or the compliance-oriented perspective, the difference between actual and theoretical worlds creates

---

163

certain difficulties, or at least certain choices, in operationalizing any definition of game-theoretical iterations and actions.

The formally-oriented perspective assumes the existence of a particular time by which the relevant formal activity must have occurred in order for a nation to have chosen a cooperate action in the iteration under consideration. In the compliance-oriented perspective, the difficulty stems from the fact that compliance with any particular obligation must be judged bi-modally to have occurred (or not occurred) for the iteration even though the actual unfolding of behavior over time may be more subtly graded. In employing either perspective, one may base a bi-modal judgment on formal activity or substantive compliance at a single point in time during the relevant iteration or continuously over some other period. Both the formally-oriented and compliance-oriented perspectives offer a number of plausible alternatives for the single point, such as the beginning or the middle or the end of the iteration. When actions are relevant not as discrete occurrences but as part of a cumulative course of behavior, a cooperate action may require that the condition to be met occurs continuously during the relevant iteration or simply that it be met more often than not. Choosing among these measures may not be easy, and choosing any of them may involve some distortions.

Consider as an example the negotiation iteration. The formally-oriented perspective on the correspondence between real-world and game-theoretical actions focuses upon whether a nation has sent a delegate to international negotiations aimed at agreeing upon a treaty draft. Does a cooperate action consist of sending a delegate to the first meeting? The last meeting? Any meeting? All meetings? Most of the meetings? The first three questions implicate a point-in-time standard; the last two, a cumulative standard. Point-in-time assessments require less effort to determine, but cumulative assessments may more accurately reflect the true extent of cooperation. A nation that sent a delegation to only the first meeting would not have chosen a cooperate action under the cumulative standard suggested above, but if one examined only the first meeting, one might reach the opposite conclusion.

The compliance-oriented perspective on the negotiation iteration focuses upon whether a nation has conducted its negotiations in good faith. Leaving aside the question of exactly what behavior constitutes good faith, there remains the question of just how much good faith is enough to constitute a cooperate action in a compliance-oriented view of the negotiation iteration. Good faith throughout the first meeting? During most of the first meeting? Throughout the last meeting? Throughout every meeting? Throughout most of the meetings? As with the formally-oriented perspective, a nation's actions may be more easily characterized using a point-in-time metric, but the result may be less accurate than one obtained by a cumulative measure.

Similar questions of definition present themselves with respect to the other iterations, for they too unfold over some continuous period of time. One nation may sign a treaty months before another does. Are both nations cooperating, or is only the first? A treaty may take years to enter into force and may then remain in force for years. Is a nation taking a cooperate or a defect action in the signature iteration if that nation ratifies a treaty five years after signature? Is a nation that ratifies a few months after entry into force cooperative or uncooperative? Similar difficulties may occur after entry into force. If a nation plainly violates a treaty for a year but

then complies perfectly for the next nine years, is that nation cooperative or uncooperative? Choosing between a point-in-time metric and a cumulative metric and choosing which particular measurement to use within each type of metric may present a difficult task in any iteration.

The differences between the Charter of the United Nations[164] and the Vienna Convention with respect to the duration of the signature phase can serve as an example of the potential variation among treaties with respect to different iterations along formally-oriented lines. Fifty of fifty-one original members of the United Nations signed the U.N. Charter on June 26, 1945; the fifty-first (Poland) had the U.N. Charter signed on its behalf on October 15, 1945.[165] By its terms, the U.N. Charter was to enter into force after ratification by the permanent members of the Security Council and by a majority of the other signatories.[166] This entry into force occurred on October 24, 1945, less than four months after the first signature.[167] Every signatory that had not ratified by that date had ratified the U.N. Charter by the end of 1945.[168] The treaty process with respect to the Vienna Convention, in contrast, unfolded in a more leisurely fashion. Thirty-two of the participants in the United Nations Conference on the Law of Treaties signed the Convention on May 23, 1969, the day the treaty was opened for signature.[169] Of the remaining conference participants, twenty-five signed the Convention at a later date, with the last signature occurring on April 12, 1976.[170] The Vienna Convention states by its terms that entry into force shall occur after thirty-five nations have ratified or acceded to it.[171] That entry into force proved to be January 27, 1980, more than ten years after the first signatures.[172]

The treaty process surrounding the U.N. Charter unfolded more rapidly than that of the Vienna Convention. One may therefore expect that the difficulties in compressing time into a single iteration would be greater with respect to the Vienna Convention than with respect to the U.N. Charter. More than 95% of the signatures on the U.N. Charter occurred on the first day that the treaty was open for signature, and all signatures occurred within four months.[173] In contrast, less than two-thirds of the signatories signed on the first day that the Vienna Convention was open for signature, and the final signature occurred almost seven years later.[174] The U.N. Charter entered into force within four months of the first signature.[175] The Vienna Convention entered into force more than ten years after the first signature, and quite a few of the Vienna Convention's signatories had not ratified that instrument even twenty-five years after signing it

164

165

166

167

168

169

170

171

172

173

174

175

and fifteen years after its entry into force.[176]

Both the U.N. Charter and the Vienna Convention are agreements of indefinite duration. The default rule, implicit in the Vienna Convention itself, is that treaties are of indefinite duration--that is, a treaty that has entered into force will remain in force indefinitely[177]--and neither of these treaties specifies a termination date. The difficulties of compressing a continuous span of time into a single iteration is especially clear with respect to an entry-into-force iteration of indefinite duration. The Vienna Convention sets forth a default method by which the parties may terminate a treaty, but that method involves the unanimous consent of the parties and is thus unlikely to be predictably employed.[178]

As the length in real time of an iteration increases, the differences among various standards of measurement (point-in-time versus cumulative, for example) increase accordingly. For cumulative metrics such as "compliance throughout the period" or "compliance throughout most of the period," one needs to know the duration of the period before making the relevant judgment, so these metrics are useless when the treaty is still in force and has no specified termination date.

If the analyst were willing to consider only those treaties that have been terminated, the *ex ante* indefiniteness of treaties would not present a problem *ex post*, as the treaties under scrutiny all have known termination dates. The drafters of some treaties have elected, in accordance with the option provided (but not taken advantage of) by the Vienna Convention, to specify a definite termination date.[179] The length of the entry-into-force iteration of these agreements is therefore known, though one could not, of course, characterize the compliance-oriented behavior of the parties as a cooperate or defect option until the end of the relevant period.

Fortunately, a number of treaties of indefinite duration suggest a plausible division of the entry-into-force phase into multiple iterations. The International Convention for the Regulation of Whaling, for example, sets forth its crucial substantive provisions in a schedule that the International Whaling Commission (ICW) may amend "from time to time."[180] Since 1949, the ICW has met once a year (except for 1977 and 1978 when it met twice) and has made at least some changes to the schedule in nearly every year it has met.[181] Thus, a year might be the appropriate length for each of a series of entry- into-force iterations for the ICRW. The Hague Convention on the Taking of Evidence states that it shall remain in force for five years but then shall "be renewed every five years" thereafter if no party has denounced the treaty; a series of entry-into-force iterations of five years' duration thus seems appropriate in this case.[182] Many

---

176

177

178

179

180

181

182

other examples of iteration lengths suggested by the text or practice associated with a treaty may be derived from a variety of treaties.[183]

B. *Difficulties in Characterizing Activities as Strictly Dichotomous*

The discussion in Part III assumed that the acts of nations are purely dichotomous--every act is either a cooperate action or a defect action. With respect to the formally-oriented perspective on the correspondence between real-world and game-theoretical actions, the discussion proceeded as if nations either send a representative to international negotiations or do not, as if nations either sign a treaty or do not, and as if nations either ratify a treaty or do not. With respect to the compliance-oriented perspective on the correspondence between real-world and game-theoretical actions, the discussion in Part III proceeded as if nations either negotiate in good faith or do not, as if nations either refrain from defeating the object and purpose of a treaty or do not, and as if nations either comply with all the obligations of a treaty in good faith or do not.

Precisely dichotomous characterizations of behavior during the treaty process may not be so easily discernible. Any attempt to call a given action a cooperate or defect action may face two conceptual difficulties: the difficulty of deciding what is the treaty text, and the breadth of compliance at issue in an compliance-oriented view of the negotiation and entry-into-force iterations.

1. *What is the Treaty Text? Reservations, Amendments, and Textual Ambiguities in the Signature and Entry-into-Force Iterations*

A variety of international legal phenomena render the simple cooperate or defect dichotomy potentially inaccurate. Sometimes a nation enters a reservation to a treaty, after which that nation is not bound by the portion of the text to which it objects, and other nations are not bound to observe the reserved obligations in their international relations with the reserving nation. Another party-nation may then possess fewer obligations towards the reserving nation than it possesses towards other signatories. A policy of the non-reserving nation could thereby violate its obligations with respect to other non-reserving nations but not violate (lesser) obligations with respect to the reserving nation. Similarly, a majority of nations may propose amendments to a treaty to which a minority of nations validly objects; nations that consent to such an amendment and nations that do not end up with different obligations under the treaty text. Finally, sometimes the text of a treaty is ambiguous and nations dispute the meaning of the relevant text; a nation that adopts one interpretation of an ambiguous provision arguably has a different set of obligations than a nation adopting a conflicting interpretation of that provision. I address each of these possibilities in turn.

a. Reservations

The Vienna Convention broadly authorizes a nation to enter a reservation as a formal

---

[183]

indication that its consent in the signature or entry-into-force iterations extends to a limited portion of the obligations contained within a treaty.[184] A reservation may not be entered with respect to a given obligation, however, if the treaty explicitly or implicitly prohibits such a reservation, or if "the reservation is incompatible with the object and purpose of the treaty."[185] A permissible reservation concerning a particular treaty obligation exempts the reserving state from any international legal duty to comply with that obligation.[186] Nations not making such a reservation remain bound by the portion of the text at issue.

In a situation in which some but not all nations enter a reservation, the simple dichotomy described in Part III cannot easily survive the additional complications. From the formally-oriented perspective, nations signing, ratifying, or acceding to the treaty have consented to different sets of obligations--the notion that every nation is identically bound vis-a-vis every other no longer applies. In effect, no single treaty exists. From the compliance-oriented perspective, the effect of reservations is that nations acquire different sets of obligations against which judgments of compliance or non-compliance are made.

One approach to the problems introduced by reservations would be to retain a rigid binary characterization of the choice open to nations, whereby a nation entering any reservations whatsoever would be judged to have chosen a defect action. This approach has the advantage of clarity but presumably sacrifices accuracy, at least in the case of whatever might be called a "minor" reservation. Alternatively, one might make a judgment as to the magnitude of each reservation and characterize only those nations making significant reservations as having chosen a defect action in the relevant iteration, albeit at the expense of adding the further complication of assigning degrees of significance. A nation that enters a truly minor reservation, or a reservation more in the nature of a clarification, would under this approach still be considered to have chosen a cooperate action.

Some commentators examining the utility of the PD in characterizing international relations have criticized what they see as the limits of a strictly dichotomous choice of actions.[187] At least two more complex alternatives are possible. One might abandon the stark binary nature of a cooperate-defect characterization and allow a wider range of actions, such as "ratify," "ratify with reservations," and "do not ratify."[188] Unfortunately, however, little game-theoretical work exists with respect to games involving even the slightly broader range of choice allowed by three, as opposed to two, actions.

One might apply a more detailed analysis that characterizes a treaty as involving several "sub-games," each of which corresponds to an important substantive treaty provision. A treaty aimed at protecting the ozone layer, for example, might include a series of provisions regulating particular classes of ozone-depleting chemicals, such as cholorofluorocarbons (CFC's) and

---

184

185

187

188

halons, as well as a provision related to an international fund for scientific research into ozone layer depletion. One might characterize such a treaty as involving three sub-games--one related to CFC's, one to halons, and one to the fund. In each sub-game, any reservation with respect to the relevant provision would constitute a defection, but would not affect the characterization of actions in other sub-games. For example, a reservation with respect only to the provisions concerning CFC's would be a defect action with respect to the CFC-related sub-game, but would not interfere with cooperate actions in the sub-games involving halons and the research fund.[189] The analysis of such linked sub-games, however, involves considerably more complexity than the analysis of a single game.[190]

I should also note that my analysis of reservations has so far assumed that non-reserving nations simply take note of any reserving nation's objections. Occasionally, however, a significant reservation will cause other nations to re-negotiate the treaty.[191] Such a situation carves another facet into the problem of characterization. The reserving nation has defected in the ratification iteration, yet that "defection" defines the cooperative, consensus position in the next iteration. One could simply accept that this switch is the price the analyst pays for relying on consensus as the yardstick of cooperation. The alternative approach necessitates the use of hindsight: an examination of the process as a whole can reveal which reservations are ultimately cooperative rather than uncooperative.

b. Amendments

Like reservations, amendments raise the question of what constitutes *the* treaty text.[192] A nation may ratify a treaty without reservations, but then refuse to consent to an amendment proposed by another party after entry into force. A nation refusing to consent to an amendment is typically not bound by that amendment, but obligations previously binding such a nation remain in place. The obligations with respect to which consent has formally been expressed (and thus against which compliance is to be judged in the compliance-oriented model) thus differ from nation to nation. The approaches applicable to resolving the complications of reservations--characterizing refusal to consent to an amendment as a defection, constructing a range of action that includes more than two choices, or differentiating the various provisions of a treaty into various sub-games--suggest themselves here as well.

In the case of amendments, however, one may face additional difficulties in the form of potential overlap between iterations. Because the amendment occurs after signature of a complete treaty text, an amendment marks the commencement of a sort of mini-treaty process--negotiation of the treaty amendment, signature of the negotiated text of that amendment, and ratification of the amendment--that raises issues of both formal adherence and actual compliance at the various stages of the amendment process. This mini-treaty typically unfolds during (and thus overlaps) the entry-into-force iteration of the treaty that is the subject of

---

189

190

191

the amendments. Such simultaneity or overlapping of iterations is in one sense an opportunity for analysts of international cooperation--it provides more iterations to observe--but it can also make for a messy analysis. Should one include the various iterations involving the amendment within the entry-into-force iteration (assuming that the amendment follows entry into force, which would be the typical course of events), or should one instead consider the amendment process as its own set of iterations occurring parallel to other evaluations relating to the underlying treaty?

In a sequential view, one would consider the entry into force of an amendment to signal the end of the entry-into-force iteration of a previous period, and then simply judge compliance in the next period according to the rules set forth by that amendment. If the iterations in the amendment process occurred rapidly enough, and if one believed that the amendment process dominated the participants' perceptions of whether others were cooperating generally during the relevant period, then one could insert a set of formal iterations related to the amendment between the entry-into-force iterations, and simply ignore any potential problems of simultaneity. The sequence of iterations would then run: treaty negotiation, treaty signature, treaty entry into force, amendment negotiation, amendment signature, and amendment entry into force. The advantage of thereby obtaining a longer sequence of iterations is that various behaviors too difficult to detect in shorter strings might make themselves apparent.

According to a parallel view of amendments, in contrast, one would simply look at another three-iteration process (relating to the amendment) generated by the same process (the relevant treaty) that also gave rise to an analytically distinct, if in part temporally overlapping, sequence of three iterations. One string of iterations would consist of the sequence: treaty negotiation, treaty signature, and treaty entry into force. Another string of iterations would consist of the sequence: amendment negotiation, amendment signature, and amendment entry into force. This approach sacrifices the opportunity to include a few potential iterations directly within the stream of other iterations (for example, including the negotiations relating specifically to the amendment in the entry-into-force iteration of the pre-amended treaty). This parallel approach has the advantage, however, of avoiding possible analytical difficulties from the simultaneous evaluation of two iterations (that is, a set of pre-entry-into-force iterations relating to the amendment and an entry-into-force iteration relating to the original treaty).

c. Ambiguous Provisions

The third difficulty that might arise in attempting to categorize real-world behavior dichotomously, at least with respect to the compliance-oriented view, involves the interpretation of ambiguous provisions. If a provision is ambiguous, then, by definition, at least two behaviors are potentially consistent with compliance. Because the international legal system frequently fails to provide definitive resolutions even when confronted with specific disputes, there may be instances in which a nation undertakes behavior that appears arguably, but not definitively, non-compliant. The proper categorization of such behavior in the compliance-oriented view is unclear.[193]

---

[193]

As in the discussion of reservations and amendments above, a variety of approaches is possible. One might characterize behavior raising any shadow of non-compliance as a defect action; one might construct an additional category of action ("arguable non-compliance," for example) beyond the simple comply/do-not-comply dichotomy; and one might divide the provisions of a treaty into sub-games.

There are at least two complexities inherent to an approach strictly parallel to that discussed with respect to reservations and amendments. First, a judgment about the degree to which behavior is compliant remains necessary, and such a judgment may well involve a need to resolve the very difficulty at issue: how is one to know whether behavior is "arguably non-compliant" without knowing anything about whether such behavior is compliant at all? Second, one might argue that a nation undertaking non-compliant behavior is in some circumstances initially acting in a cooperative manner even if that nation's behavior is eventually agreed to be non-compliant. A nation that in good faith took a particular position on an ambiguous provision and then participated fully in the relevant dispute-resolution procedures would be cooperating in the normal sense of the word, despite the fact that the eventual resolution of the dispute involves an interpretation of the relevant provision that makes that nation's behavior legally impermissible. To characterize such a nation as having chosen a defect action might thus seem unduly harsh, but to characterize that nation's behavior as a cooperate action, despite the fact that the treaty may be interpreted to proscribe the relevant behavior, seems unsatisfactory as well.

2. *What is the Breadth of Compliance in the Negotiation and Entry-into-force Iterations?*

Reservations, amendments, and ambiguities present potential difficulties because the text of the treaty that gives rise to the obligations of nations that have signed, ratified, or acceded to that instrument can effectively differ from nation to nation, depending on whether and how that nation has entered reservations, consented to amendments, or resolved textual ambiguities. A similar problem can arise with respect to a nation's compliance with its obligations even in the absence of reservations, amendments, or ambiguities: a nation may be legally obliged to comply with exactly one set of obligations but may actually comply with only some of those obligations. The characterization of that partial compliance then involves a potentially difficult judgment about whether the nation has chosen a cooperate or a defect action for the relevant iteration. For reasons that I explain below, the signature iteration does not present this particular problem, but the negotiation and entry-into-force iterations may.

Assume, for example, that a treaty obliges a nation to cease its production of CFC's by the year 2000 and to contribute $1,000,000 to an international research fund. The nation signs the treaty in 1996, ratifies the treaty in 1997, and sees the treaty enter into force in 1998. Assume further that this nation does cease production of CFC's by the year 2000 but contributes no money to the international fund. Is the nation's behavior a cooperate action or a defect action? In contrast to the examples discussed above, it is not the legal obligations of this nation, but the degree of compliance with those obligations, that is at issue. The nation is formally bound to the single version of the treaty that has entered into force, but its degree of compliance with that text is unclear.

One might construct a similar example with respect to good faith in conducting treaty negotiations. Suppose that one nation lies to all of the other nations concerning its ability to reduce its production of CFC's but truthfully describes its legitimate inability to contribute to the international fund. Once again, it is unclear whether such behavior represents a cooperate action or a defect action.

As with reservations and amendments, one might take an all-or-nothing approach and characterize the slightest failure to comply with all of a nation's obligations as a defect action. With respect to compliance after entry into force, however, a slightly more generous standard is suggested by the law of treaties itself: certain breaches of a treaty are material, and release other nations from their obligations with respect to the breaching party, while other breaches are immaterial and trigger no such release.[194] Guided by the particulars of the standard for determining when a breach is material or immaterial, one might consider only material breaches as defect actions, while allowing immaterial breaches to be characterized as cooperate actions equivalent to adherence.[195]

As with reservations and amendments, one could identify three separate characterizations of national behavior (that is, "comply," "breach immaterially," and "breach materially") corresponding to three separate game-theoretical actions (that is, "cooperate," "cooperate imperfectly," and defect). A more elaborate set of actions, however, comes at a price. As noted above, relatively few game-theoretical studies examine games with more than two actions available to each player.[196] A tripartite characterization of compliance actions would therefore substantially narrow the range of applicable theoretical models. One would have a richer theory, but fewer existing models to draw upon in elaborating that theory.

Partial adherence to a given obligation generates this difficulty when the behavior is either partial good faith in the negotiation iteration or partial compliance in the entry-into-force iteration. One should note that the signature iteration, in contrast, is unlikely to present problems of this sort. The relevant obligation in the signature iteration is to refrain from activities that defeat the object and purpose of the treaty, implying a single object and purpose. A nation thus may either act so as to defeat the single object and purpose of the treaty, or act so as not to defeat that object and purpose. Questions arising from partial adherence in the signature iteration will therefore not arise.

V. Conclusion

Part I of this Article set forth the Article's two basic building blocks: the IR "theory" of neo-liberal Institutionalism and the multi-part schema that Duncan Snidal has advanced for judging the intellectual sophistication with which IR theorists employ game theory. Part II examined one building block in light of the other by emphasizing the importance of iteration to

---

194

195

196

the evolution of cooperation generally and to the description of international relations by neo-liberal Institutionalists in particular. Concluding that the definition of iteration was critical to, but essentially ignored by, Institutionalist theory, the Article explored in Part III the possibility of using the treaty process to generate various definitions of iteration that suggested outcomes resembling the dichotomy of the two-action PD. Finally, Part IV examined and elaborated upon various phenomena that complicate the effort to apply the game-theoretical concepts of iteration and action to the treaty process.

Taking the achievement of a true theory of neo-liberal Institutionalism to be a desirable objective of a self-consciously rationalistic branch of political *science*, Part III noted a number of characteristics of the treaty process that bode well for its utility in operationalizing the bedrock of Institutionalist thought. Part IV, in turn, explored some significant obstacles that any effort actually to conduct such an operationalization would face, but many--perhaps even all--of those obstacles would be present in roughly similar form in *any* effort to judge international cooperation in terms of compliance with norms or rules or procedures in a multilateral setting. On the whole, then, I believe that the analysis of this Article strongly supports the proposition that the treaty process should be treated as the most fruitful method for pursuing the operationalization of the rationalistic core of neo-liberal Institutionalism.

Certainly the treaty process is a more-than-viable candidate for interdisciplinary efforts of international legal scholars and theorists of international relations. If the treaty process can help neo-liberal Institutionalists advance their use of game theory--and thus, in many ways, advance their entire perspective--from mere metaphor to something approaching a true theory, then international law will have done IR theory a useful service. And if international legal scholars assist political scientists in making specific examinations of the treaty process informed by the IR theorists' concern with systematic aspects of the substance and process of international cooperation, then an interdisciplinary perspective will also have done international legal scholarship a useful service.

Of course, whether international legal scholars and IR theorists will, separately or together, undertake much in the way of a theoretically informed examination of the treaty process remains an open question. Perhaps the advocates of neo-liberal Institutionalism lack even a pretension of developing a true theory, and the invariably challenging efforts to develop a more rigorous definition of iteration will founder for lack of interest. Certainly, in recent years, it has been the Neorealists, rather than the neo-liberal Institutionalists, who have deployed the advances in game theory made by mathematicians and economists over the last decade.[197] Even the Neorealists, however, have thus far focused much more upon converting those advances into the vocabulary of international relations than they have focused upon actually operationalizing the relevant theories into sustained empirical investigations. Institutionalist scholars employing tools that, like the IPD, were familiar to economists in the mid-1980's have tended to focus on microeconomic theories not directly related to the IPD or even to game theory generally.[198] This

---

- Professor of Law, UCLA School of Law; Visiting Professor of

neglect of the IPD and its operationalization by IR scholars, and especially by neo-liberal Institutionalists, leaves neo-liberal Institutionalism in a methodologically precarious position, at least so long as neo-liberal Institutionalism rests upon a characterization of international cooperation as an IPD.

Thus far, neo-liberal Institutionalism has neither elaborated upon the rationalist grounding of the theory as initially advanced in the IPD, nor advanced an alternative rationalist grounding for that theory, nor argued that the theory needs no rationalist grounding at all. Perhaps neo-liberal Institutionalism is more than a single move in a larger game of IR theory that, without progressing, becomes ever more elaborate; perhaps IR theory itself is simply part of a larger game of social science theorizing and grant-chasing. Nonetheless, the neo-liberal Institutionalists remain the only group of IR theorists that allows a role for international law in its theory and also professes an interest in the methodologies of rational choice. The use of international law in the pursuit of a more rigorous rational-choice methodology therefore seems a game worth the candle.

Footnotes

• The seminal Institutionalist work is Robert Keohane, After Hegemony: Cooperation and Discord In The World Political Economy (1984). For some other overviews of neo-liberal Institutionalism by adherents, see Robert Keohane & Lisa Martin, The Promise of Institutionalist Theory, 20 Int'l Sec. 39 (1995); Charles Lipson, International Cooperation in Economic and Security Affairs, 37 World Pol. 1 (1984). For some criticisms of neo-liberal Institutionalism, see John

---

Mearsheimer, The False Promise of International Institutions, 19 Int'l Sec. 5 (1994); Joseph Grieco, Anarchy and the Limits of Cooperation: A Realist Critique of the Newest Liberal Institutionalism, 42 Int'l Org. 485 (1988).

• For a summary of the various schools of international relations, and of their evolution and inter-relationships, see John K. Setear, Responses to Breach of a Treaty and Rationalist International Relations Theory: The Rules of Release and Remediation in the Law of Treaties and the Law of State Responsibility, 83 Va. L. Rev. 1, 2-3 nn.2-5, 6-9 nn.8-9 (1997) [hereinafter Setear, Breach of Treaty].

• Several scholars of international law have examined the interplay between neo-liberal Institutionalism and international law. See Kenneth Abbott, Modern International Relations Theory: A Prospectus for International Lawyers, 14 Yale J. Int'l L. 337 (1989) (surveying neo-liberal Institutionalism and examining its potential applicability to various problems in international law); Anne-Marie Slaughter Burley, International Law and International Relations Theory: A Dual Agenda, 87 Am. J. Int'l L. 205 (1993) (surveying the history of ideas in IR theory and international law since World War II and proposing a "dual agenda" involving the use of both IR Liberalism and neo-liberal Institutionalism in examining international law); John K. Setear, An Iterative Perspective on Treaties: A Synthesis of International Relations Theory and International Law, 37 Harv. Int'l L.J. 139 (1996) (arguing that the law of treaties governing the validity of, and degree of obligations in, treaties reflects extensive concern with the prominent rationalist-IR concept of iteration); Setear, Breach of Treaty, supra note 2 (examining the international legal rules governing responses to breach of an international treaty from various rationalist theories of international relations including neo-liberal Institutionalism); Edwin Smith, Understanding Dynamic Obligations: Arms Control Agreements, 64 S. Cal. L. Rev. 1549 (1991) (using, *inter alia*, an earlier version of neo-liberal Institutionalism, known as "regime theory," to explain the lack of legalistic formalism in the U.S.-Soviet relationship involving arms-control treaties). It should be noted that the author formerly known as Anne-Marie Slaughter Burley is now known as Anne-Marie Slaughter.

A few political scientists, for their part, have noted the potential relevance of international law to their own work. See Robert Keohane, Compliance with International Commitments: Politics Within a Framework of Law, 86 Am. Soc'y Int'l L. Proc. Ann. Meeting 176, 176 (1992) [hereinafter Keohane, Compliance with International Commitments] (noting critical importance of international legal rules in determining whether nations obey principles and rules of international regimes); Charles Lipson, Why Are Some International Agreements Informal?, 45 Int'l Org. 495 (1991) (noting that various constraints imposed by the formality of the treaty process may make informal agreements useful in situations requiring flexibility or minimal publicity).

4. See infra Part I.B-C.

5. See Burley, supra note 3, at 207 (stating that "[t]he discipline of international relations was born after World War I" and "weaned on Political Realism"); see also Steven Forde, Varieties of Realism: Thucydides and Machiavelli, 54 J. Pol. 372, 372 (1992) (stating that Realism has "a heritage which reaches back to ancient Greece" but agreeing that "Realists have dominated the postwar study of international politics"); Stanley Hoffmann, The Political Ethics of International Relations 6 (1988) (describing Realism as "probably the most distinguished school of thought in the history of international relations").

6. For influential postwar formulations of Realism, see Edward H. Carr, The Twenty Years' Crisis 1919-1939: An Introduction to the Study of International Relations, at vii, 63-88 (2nd ed. 1946) (stating his goal as "counteracting the glaring and dangerous defect of nearly all thinking, both academic and popular, about international politics in English-speaking countries from 1919 to 1939--the almost total neglect of the factor of power"); Hans J. Morgenthau, Politics Among Nations 4-15, 293-327 (6th ed. 1985).

7. See Burley, supra note 3, at 209-39 (surveying the growth of opposition to the Realist school of IR).

8. For an excellent survey of the intellectual tides of post-war IR theory (and also of international law), see

generally id.

9. Problems of collective action are intimately associated with the production of public goods, although such problems may also arise in connection with the production of related but distinct "club goods" or goods whose production leads to a "tragedy of the commons." See Setear, An Iterative Perspective on Treaties, supra note 3, at 174-76, 178-79 nn.160-61.

10. For a description of the relationship between the public-goods problem and the PD, see id., supra note 3, at 176-80. A leading neo-liberal Institutionalist, Robert Keohane, makes use of the PD as a metaphor for international cooperation. See Keohane, After Hegemony, supra note 1, at 74-78. Keohane expressly compares game-theoretical and market failure models of international interactions, announcing that the two theories reach the same conclusions as to the utility of international institutions. Id. at 97. One should note that Keohane draws upon the works of those analyzing public goods but believes that international cooperation, though characterized by a PD, rarely involves a true public good. Id. at 76-77.

11. Setear, An Iterative Perspective on Treaties, supra note
3, at 185-89.

12. Id. at 186-87.

13. See Robert Axelrod, The Evolution of Cooperation (1984).

14. Id. at 42, 52-53.

15. See Setear, An Iterative Perspective on Treaties, supra note 3, at 182 n.173 (discussing enthusiasts of the tit-for-tat strategy and arguing against treating that strategy as a panacea); see also Axelrod, supra note 13, at 39 (discussing strategies that would have been superior to tit-for-tat strategy in various tournaments).

16. Robert Axelrod & Robert O. Keohane, Achieving Cooperation

Under Anarchy: Strategies and Institutions, 38 World Pol. 226 (1986).

    17. Setear, An Iterative Perspective on Treaties, supra note
3, at 184.

    18. Stephen D. Krasner, Structural Causes and Regime Consequences: Regimes as Intervening Variables, 36 Int'l Org. 185, 185 (1982).

    19. Robert Keohane, International Institutions: Two Approaches, 32 Int'l Stud. Q. 379, 383 (1988).

    20. See John Gerard Ruggie, International Responses to Technology: Concepts and Trends, 29 Int'l Org. 557, 570 (1975) (defining regimes as "a set of mutual expectations, rules and regulations, plans, organizational energies and financial commitments, which have been accepted by a group of states"); Oran R. Young, International Regimes: Problems of Concept Formation, 32 World Pol. 331, 332-33 (1980) (describing regimes as social institutions, which may or may not be associated with "explicit organizational arrangements"). A definition in the transitional period between a focus on regimes and a focus on institutions defined the former as "[f]ormal international organizations and codified rules and norms" and let the word "institutions" include not only regimes but also less formal behaviors that are nonetheless "recognized by participants as reflecting established rules, norms, and conventions." Robert O. Keohane, International Institutions and State Power, at vii (1989).

    21. For some works surveying the actual functioning of institutions from an Institutionalist perspective, see Miles Kahler, International Institutions and the Political Economy of Integration (1995); Institutions for the Earth (Peter M. Haas et al. eds., 1993).

    22. Setear, An Iterative Perspective on Treaties, supra note
3, at 184.

23. For the seminal IR-theory discussion of all of these ideas within the neo-liberal Institutionalist framework, see Keohane, After Hegemony, supra note 1, at 89-109.

24. Burley, supra note 3, at 209-210.

25. Duncan Snidal, The Game Theory of International Politics, in Cooperation Under Anarchy, 25, 29-36 (Kenneth Oye ed., 1986) [hereinafter Snidal, Game Theory]. Snidal has frequently employed game theory in his own work, in a fashion that probably falls somewhere between "analogy" and "model" in his own terminology. See, e.g., Duncan Snidal, Relative Gains and the Pattern of International Cooperation, 85 Am. Pol. Sci. Rev. 701, 704-10 (1991); Duncan Snidal, Coordination Versus Prisoner's Dilemma: Implications for International Cooperation and Regimes, 79 Am. Pol. Sci. Rev. 923, 924 (1985).

26. Snidal, Game Theory, supra note 25, at 29-36.

27. Id. at 30.

28. Id. at 31.

29. Id.

30. Id. at 31-32.

31. Id.

32. Id. at 32.

33. Id.

34. Id. at 33.

35. Id.

36. Id. at 34-35.

37. Id. at 36 (emphasis added).

38. Id. at 26-27.

39. See, e.g., Peter C. Ordeshook, Game Theory and
    Political
Theory, at ix, passim (1986) ("[T]his research seeks to satisfy
a rigid definition of "theory," and not some ambiguous criteria
of good journalism and insightful comment."); Michael Taylor,
The Possibility of Cooperation, passim (1987) (using
mathematical game theory models to attack the liberal theory of
the state). For a description of more recent efforts to employ
sophisticated game theory (though not necessarily of efforts to
test the resulting models and true theories), and a discussion
of the confinement of these efforts largely to Realists
discussing security issues, see Setear, Breach of Treaty, supra
note 2, at 123-25.

40. See Robert Jervis, Realism, Game Theory and
    Cooperation,
40 World Pol. 317 (1988). Some state the problem in terms of
related metaphors in public goods or collective action. Elinor
Ostrom, Governing the Commons: The Evolution of Institutions for
Collective Action 2-7 (1990).

41. Cf. Snidal, Game Theory, supra note 25, at 30.

42. Keohane, After Hegemony, supra note 1, at 67-69, 73-77.

43. Id. at 82.

44. Id. at 83. Keohane also includes "market failure" as a
relevant literature in suggesting a functional explanation for
institutions. As an example, he cites the information
difficulties creating a (flawed) market for chronically
malfunctioning automobiles. Id.; see also George A. Akerlof,
The Market for "Lemons": Quality Uncertainty and the Market
Mechanism, 84 Q.J. Econ. 488 (1970).

45. Keohane, After Hegemony, supra note 1, at 104; see also
infra note 47 (discussing Keohane's use of evidence of
retaliations in his arguments for the relevance of the
tit-for-tat strategy). In this portion of his discussion,
Keohane also examines two other incentives for compliance:

precedent and reputation. He calls precedent a "weak reed to lean on." Keohane, After Hegemony, supra note 1, at 105. He is more sanguine about reputation, but presumably only where there is the prospect for retaliation. A reputation for cheating is a disadvantage only if nations retaliate against those with a poor reputation.

46. Keohane, After Hegemony, supra note 1, at 77 (comparing the small number of actors in the international system with the number of actors in problems of collective action). Snidal, though not focused exclusively on neo-liberal Institutionalism, also notices a correspondence between the nation-state in IR and the player in game theory: "[B]y assuming that power-maximizing states are the principal actors, game theory subsumes the Realist position .... [G]ame theory is equally consistent with a modified structural approach." Snidal, Game Theory, supra note 25, at 35. Much of this "modified structural approach," however, has strong similarities, in terms of both its proponents and its ideas, to neo-liberal Institutionalism. See Robert O. Keohane, Theory of World Politics: Structural Realism and Beyond, in Political Science: The State of the Discipline 503, 530-31 (Ada Finifter ed., 1983) (setting forth the tenets of modified structural Realism and using Axelrod to argue for the importance of "non-power factors").

47. Although the quoted passage above discussing "retaliatory linkage" does mention the tit-for-tat strategy, see supra text accompanying note 45, this passage is only a description of the possibility for retaliation. It is not an assertion that there is an analogy because, as predicted by the tit-for-tat strategy, nations in fact retaliate. Indeed, Keohane explicitly states that "the GATT provisions for retaliation have been invoked only once, and then ineffectively." Keohane, After Hegemony, supra note 1, at 104-05. Certainly one does not typically defend a model by showing that the activity it predicts is extremely rare. On the metaphorical or analogical level, however, the rarity of a "predicted" phenomenon could simply be an observation interesting to the writer or a part of the "imperfection" of the comparison.

48. Snidal, Game Theory, supra note 25, at 26.

49. Id. at 40.

50. Ken Binmore, Fun and Games: A Text on Game Theory 354-55
(1992) (giving mathematical proof that a finitely repeated PD
has a unique subgame perfect equilibrium in which both players
choose to be uncooperative).

51. See Axelrod, supra note 13.

52. For a treatment of this issue that goes into
substantially more depth as to the relevant game theory, see
Setear, An Iterative Perspective on Treaties, supra note 3, at
174-89 (discussing neo-liberal Institutionalism and iteration).

53. See Keohane, After Hegemony, supra note 1, at 68-75
(discussing how the use of rational choice theory combined with
a belief in egoist players leads some scholars to predict that,
based on the uniterated PD, cooperation will not normally arise
in the international arena).

54. But see infra note 85.

55. See Robert S. Pindyck & Daniel L. Rubinfeld,
Microeconomics 446 (2d ed. 1992).

56. The absolute values in Figure I are arbitrary.
However,
the *relative* values define the rank-ordering and make this
payoff matrix a PD.

57. For a review of examples of international cooperation
set
against the backdrop of the PD, see Setear, An Iterative
Perspective on Treaties, supra   note 3, at 180 n.163.

58. Binmore, supra note 50, at 311-12.

59. For a somewhat more elaborate treatment of this
rationale, see Setear, An Iterative Perspective on Treaties,
supra note 3, at 178 n.158. See also Eric Rasmusen, Games and

Information: An Introduction to Game Theory 27-29 (reprinted with corrections 1991) (discussing the dominant strategy in a one-shot PD).

60. See Rasmusen, supra note 59, at 28 (discussing the dominant strategy equilibrium).

61. Id. at 32-35 (discussing the Nash equilibrium).

62. See Anatol Rapoport et al., The 2 x 2 Game 104 (1972).

63. See Binmore, supra note 50, at 357-58 (giving algebraic proof that strategy CC is a Nash equilibrium in a two player infinite horizon PD).

64. See Setear, An Iterative Perspective on Treaties, supra note 3, at 189 n.208 (discussing "shadow of the future" as one of five methods offered by Axelrod to improve chances of cooperation).

65. Oxford English Dictionary of Quotations 296 (3d ed. 1979).

66. Axelrod, supra note 13.

67. See id. at 30 (round-robin tournament), 52 (ecological tournament).

68. See id. at 33 (describing successful strategies as those that waited to defect until opponent had defected in some previous iteration).

69. See Setear, An Iterative Perspective on Treaties, supra note 3, at 187, 189.

70. See supra notes 58-62 and accompanying text.

71. Axelrod, supra note 13, at 12-16.

72. Setear, An Iterative Perspective on Treaties, supra note

3, at 186.

73. See Axelrod, supra note 13, at 30 (round-robin tournament), 52 (ecological tournament). Axelrod's round-robin tournaments consisted of 200 iterations between each player, id. at 30, while his ecological tournament consisted of 1,000 "generations" of such 200-iteration rounds, id. at 52.

74. For a general discussion of future discounting, see Edith
Stokey & Richard Zeckhauser, A Primer for Policy Analysis 160-65 (1978).

75. See Setear, An Iterative Perspective on Treaties, supra note 3, at 188 (discussing the implicit assumption of no future discounting used in Axelrod's work tournaments).

76. Keohane, After Hegemony, supra note 1, at 76 (citation omitted).

77. Id.

78. Id.

79. Id.

80. Id.

81. See Snidal, Game Theory, supra note 25, at 51. Indeed, Snidal suggests that one should reduce the simultaneity of bargaining in order to increase the prospects for cooperation--as one would expect given the positive relationship between iteration and cooperation. Id. at 51 (arguing that the analysis of iteration "suggests that states may promote cooperation (in Prisoner's Dilemma) by...dividing negotiations into larger numbers of smaller sequential steps").

82. Michael McGinnis has argued that cooperation in such games, if they are all PDs, may evolve "through cooperation on each game alone, through comprehensive linkage, and through a quid-pro-quo arrangement" in which each player agrees to cooperate on some issues while the other player defects in

exchange for like treatment on other issues. Michael D. McGinnis, Issue Linkage and the Evolution of International Cooperation, 30 J. Conflict Res. 141, 151 (1986). McGinnis seems to be engaged in what those using Snidal's framework might call "modeling without theory." He uses quite rigorous techniques in a mathematical sense but makes no effort, beyond mentioning the existence of various perceptions of "issue linkage," to develop any correspondence between international behavior and his model.

83. A somewhat similar issue involves the proper number of actors to include as necessary to determine the payoffs to participants. Axelrod's work, and most other work in the tournament genre, involves more than two players in each tournament, but determines the payoffs by aggregating payoffs from pairwise interactions. As Cornes and Sandler note:

> Axelrod (1984) has investigated the evolution of cooperation in repeated games in a study that draws on the results of a computer tourney, in which he invited contestants to submit computer programs to play a repeated Prisoner's Dilemma. In general, it appeared that a simple tit-for-tat strategy performed extremely well, by virtue of its ability to secure cooperation from the other player. Such a result is certainly interesting, although its relevance to public goods problems is debatable. For one thing, we have pointed out that, even within the binary choice framework, the public goods problem may not conform to the Prisoner's Dilemma. Second, n-player games are more complicated than two-player games. Axelrod's tournament consisted of rounds of two-player contests, whereas our ultimate interest is in public goods models with many players. It is not clear that it makes much sense for an individual to punish or reward 99 others to secure their cooperation in later plays of the game.

Richard Cornes & Todd Sandler, The Theory of Externalities, Public Goods, and Club Goods 141-42 (1986) (citations omitted).

One can also explore a PD involving payoffs dependent upon the choices of more than two actors. See, e.g., Taylor, supra note 39, at 82-108; Ordeshook, supra note 39, at 213-14. This

analysis is rather more complex.

84. Keohane, After Hegemony, supra note 1, at 76.

85. Axelrod, supra note 13, at 12 (noting that the ability
   to
communicate in a PD is irrelevant). The prisoners in the PD are
typically said to be held incommunicado in separate cells. So
long as the prisoners have no credible sanctions available with
which to threaten each other, however, they could actually be
placed in the same cell. If they do have credible sanctions
available, and if negotiations would allow the conveyance of
those threats, then the entire analysis must shift into an
analysis either of "cooperative" games or of some
non-cooperative game that is not a PD. See id.

86. Keohane, After Hegemony, supra note 1, at 76 (emphasis
added).

87. Kenneth Abbott has briefly discussed the differences
among the standard PD, second-chance PD, and iterated games.
Kenneth Abbott, "Trust But Verify": The Production of
Information in Arms Control Treaties and Other International
Agreements, 26 Cornell Int'l L.J. 1, 5-12 (1993). According to
Abbott, "[i] n the standard PD game, neither player can change
its move after learning what   the   other   has   done....   If
players have the ability to revise their moves in response to
the   actions   of   others,   however,   the   situation   is   much
different." Id. at 9.

88. See supra notes 76-87 and accompanying text. The only
other reference to iteration in the context of international
relations (as opposed to its role in the purely game-theoretical
context) that I have found in the same work also treats the
definition of iteration quite casually:

> By elevating injunctions to the level of principles
> and rules, furthermore, regimes construct linkages
> between issues. No longer does a specific
> discriminatory agreement constitute merely a
> particular act without general significance; on the
> contrary, it becomes a "violation of GATT" with

serious implications for a large number of other
issues. In the terms of Prisoner's Dilemma, the
situation has been transformed from a single-play to
an iterated game.

Keohane, After Hegemony, supra note 1, at 89-90. This passage
confuses issue linkage with iteration. The act of reaching a
specific discriminatory agreement is (or is not) part of an
iterated game regardless of whether GATT exists. It is the
significance of the act, and the contours of the likely
response--not the presence (or absence) of iteration--that
depend upon the elevation of an injunction to a rule.

89. See, e.g., Keohane, After Hegemony, supra note 1, at 76
("Such bargaining typically occurs not only in one bargaining
episode but in several, over a period of time. Negotiations on
international monetary arrangements, trade, and energy take
place continuously and are expected to continue indefinitely
into the future.").

90. Cf. R. Harrison Wagner, The Theory of Games and the
Problem of International Cooperation, 77 Am. Pol. Sci. Rev. 330,
333-34 (1983) (criticizing the use of a one-shot PD as a
metaphor for international relations in light of the fact that
nations may have prior understandings about how the game is to
be played).

91. See, e.g., John Conybeare, Trade Wars: A Comparative
Study of Anglo-Hanse, Franco-Italian, and Hawley-Smoot
Conflicts, 38 World Pol. 147, 150 (1985) ("Trade games are
invariably iterated; a state can always respond to the other
state's previous move (for example, by imposing a retaliatory
tariff), thereby prolonging a trade conflict for as long as it
sees advantages in doing so."); George Downs et al., Arms Races
and Cooperation, 38 World Pol. 118, 129 (1985) ("The prospect of
iteration plays a central role in arms decisions because there
is always the possibility that any rate of arms increase or
technological innovation will inspire a new race or intensify
the existing one.").

92. See also supra notes 79, 87 and accompanying text
(describing reversal of government policy). There is actually

an alternative view. In the PD as traditionally conceived, the players choose their actions simultaneously. In other versions of the game, however, one player chooses and reveals her action before the other player does so. In such a game, with the same payoffs as the PD and with just one iteration, the result should always be mutual defection. See supra notes 58-62 and accompanying text. One might call the move of the player choosing second a "retaliation" since it follows the defection of the first-moving player. Note, however, that in this alternative, one-iteration view, there is no period of initial cooperation. This alternative view is therefore inconsistent with accounts like Conybeare's in which a period of initial cooperation is followed by one nation's defection, which in turn gives rise to the other nation's retaliatory defection. See Conybeare, supra note 91, at 150. For Conybeare, a trade "conflict" arises out of initial cooperation, followed by one nation's raising its tariffs, followed by a retaliatory tariff erected by the other nation. Id. See also Binmore, supra note 50, at 134-35, who argues that sequential and simultaneous move games often must be analyzed differently.

93. See supra note 92 for the distinction between sequential
and simultaneous move games.

94. For a taxonomy of two-player games, see Rapoport et al.,
supra note 62, at 15-35.

95. Id.

96. Id.

97. See supra p. 655.

98. Axelrod, supra note 13, passim.

99. Alternative decision rules could also be adopted, however. One such alternative would be to characterize each nation's action as the action which was in effect on the last day of the relevant period, regardless of when that action was undertaken.

100. Note that, under the alternative decision rule, described supra at note 99, the sequence of U.S. actions in this example would be *C-D* (rather than *C-C*) even though the duration of the iteration is still two weeks.

101. See, e.g., Keohane, After Hegemony, supra note 1, passim.

102. In addition to the PD, other games include Chicken, Harmony, Deadlock, Stag Hunt, and Coordination. Rapoport et al., supra note 62, at 15-35.

103. A contribution costs $2.5 million and yields nothing, and each player's payoff is independent of the other's action, so Figure II is easy to derive.

104. The numbers in Figures II-V represent payoffs in millions of dollars.

105. When both nations contribute, they each pay out $2.5 million in the first two months (with no benefits) and then each generates $500,000 in benefits for the next four months (with no costs). Because the good at issue--the prevention of environmental damage--is a public good, both nations garner the benefits of each nation's generated benefits. See supra notes 9-10 and accompanying text (discussing public goods).

If both contribute, then each pays $2.5 million and garners $1 million monthly for four months, for a total net benefit to each nation of $1.5 million ($4 million for the four months of benefits, minus $2.5 million).

If only one contributes, then the contributing nation loses $500,000--a $2.5 million outlay offset by just $2 million in benefits (four months of $500,000 monthly benefits). The non-contributing nation makes no outlays but still receives $2 million in benefits.

If neither nation contributes, then the status quo persists.

106. For some of the actual treaties regulating the production and consumption of ozone-depleting substances, see Vienna Convention for the Protection of the Ozone Layer, Mar. 22, 1985, T.I.A.S. No. 11,097, 26 I.L.M. 1529 (entered into force Sept. 1, 1988); Montreal Protocol on Substances That Deplete the Ozone Layer, Sept. 16, 1987, 26 I.L.M. 1550 (entered into force Jan. 1, 1989).

107. Costs for each nation are its $30 million in start-up expenses during the initial year, plus three years of annual expenditures at $55 million per year, for a total of $195 million in costs over the four years. Benefits accrue for three years at $100 million per year--the combination of the $50 million in annual benefits generated by a nation's own actions and the $50 million generated annually by the other nation--for a total of $300 million per nation in benefits. The net benefit for each nation from mutual cooperation is therefore $105 million ($300 million in benefits minus $195 million in costs).

If only one nation contributes, then the contributing nation expends $195 million in costs for a benefit of $150 million ($50 million annually for three years). The contributor thus loses $45 million. The non-contributing nation obtains $150 million in benefits for no expenditures.

108. Costs for each nation are its $30 million in start-up expenses during the initial year, plus four years of annual expenditures at $55 million per year, for a total of $250 million in costs over the five years. Benefits accrue for four years at $100 million annually--the combination of the $50 million in annual benefits generated by a nation's own actions and the $50 million generated annually by the other nation--for a total of $400 million per nation in benefits. The net benefit for each nation from mutual cooperation is therefore $150 million ($400 million in benefits minus $250 million in costs).

If only one nation contributes, then the contributing nation expends $250 million in costs for a benefit of $200 million ($50 million annually for four years). The contributor thus loses $50 million. The non-contributing nation obtains $200 million in benefits for no expenditures.

The calculation of each nation's benefits as the sum of the benefits generated by them both is driven by the characterization of the benefits as being public goods. See supra notes 9-10 and accompanying text (discussing public goods).

109. The discussion below proceeds as if a single treaty were the focus of the analysis. International relations, however, provide several examples of treaties addressing the same issue expressly or implicitly linked together, including agreements governing protection of the ozone layer, whaling, and the possession of strategic nuclear weapons. See Setear, An Iterative Perspective on Treaties, supra note 3, at 217-27 (discussing these three sets of treaties from an iterative perspective). One might well be able to string together the iterations involved in each of the linked treaties into a larger series of iterations representing the regulation by treaty of the particular subject matter.

110. See, e.g., Mark Janis, An Introduction to International Law 14-15 (2d ed. 1993) (the law of treaties sets forth "the accepted rules respecting the making, the effect, and the amendment, invalidity, and termination of agreements among states"); Malcolm N. Shaw, International Law 560 (1991) (discussing various definitions of the law of treaties).

111. Vienna Convention on the Law of Treaties, May 23, 1969, 1155 U.N.T.S. 331 [hereinafter Vienna Convention]. Although the United States is not a party to the Vienna Convention, the U.S. Department of State has recognized the Vienna Convention as the "authoritative guide to current treaty law and practice." S. Exec. Doc. No. 92-L, at 1 (1971). *The Restatement of the Foreign Relations Law of the United States* "accepts the Vienna Convention as, in general, constituting a codification of the customary international law governing international agreements." Restatement (Third) of the Foreign Relations Law of the United States 145 (1987).

112. For a lengthier and more complete description of the

treaty process, see Setear, An Iterative Perspective on Treaties, supra note 3, at 147-55 (discussing three crucial events that together specify four iterations); id. at 191 (discussing the four-phase process); id. at 193-95 (discussing increasing obligations as phases advance).

113. Vienna Convention, supra note 111, art. 18.

114. Id., art. 26.

115. Id.

116. Janis, supra note 110, at 18-22; see also Setear, An Iterative Perspective on Treaties, supra note 3, at 149 nn.41-42 (stating that ratification is the most common indication of full consent, but that additional indicators are also recognized).

117. Janis, supra note 110, at 19; see also Vienna Convention, supra note 111, arts. 14-15 (describing accession in language nearly identical to that used to describe ratification).

118. Where ratification is the means of expressing full consent to be bound, the default rule is that ratification by *all signatories* is necessary for entry into force. Vienna Convention, supra note 111, art. 24(2) ("Failing any [provision in the treaty to the contrary], a treaty enters into force as soon as consent to be bound by the treaty has been established for all the negotiating States."). The parties may nonetheless change this rule. See id., art. 24(1) ("A treaty enters into force in such manner and upon such date as it may provide or as the negotiating States may agree."). For examples of treaties that modify the default rule, see Setear, An Iterative Perspective on Treaties, supra note 3, at 150 n.45. For examples of treaties that count both ratifications and accessions against the total necessary for entry into force, see Vienna Convention, supra note 111, art. 84(1); Convention on International Trade in Endangered Species of Wild Fauna and Flora, Mar. 3, 1973, art. XXII(1), 27 U.S.T. 1087, 993 U.N.T.S. 243 [hereinafter CITES]; Convention on Biological Diversity, June 5, 1992, art. 36(1), 31 I.L.M. 822. For an

example of a treaty that allows both ratification and accession but does not treat them equally for the purpose of determining entry into force, see International Convention for the Regulation of Whaling, Dec. 2, 1946, art. X(4), 62 Stat. 1716, 161 U.N.T.S. 72, as amended, Nov. 19, 1956, 10 U.S.T. 952 [hereinafter ICRW] (treaty enters into force upon ratification of six signatories, including five specifically named). The ICRW refers to accession as "adherence." Id., art. X(2).

119. See, e.g., Vienna Convention, supra note 111, art. 84(2); CITES, supra note 118, art. XXII(2); Convention on Biological Diversity, supra note 118, art. X(3); see also ICRW, supra note 118, art. X(4) (nations ratifying or adhering after treaty enters into force are bound as of date of ratification or notice of adherence).

120. See Janis, supra note 110, at 17-18; Setear, An Iterative Perspective on Treaties, supra note 3, at 149. One might contrast this situation with the typical state of affairs with respect to a private contract in a common-law country, where signature and a full expression of consent to be bound are typically the same event.

121. Vienna Convention, supra note 111, art. 10(b).

122. Id., art. 24(4). According to Reuter, the legal basis for compliance with such provisions "lies in the consent implicit in the very adoption of the text of the treaty." Paul Reuter, Introduction to the Law of Treaties 53 (Jose Mico & Peter Haggenmacker trans., 1989). This type of provision is frequently found in treaties establishing international organizations; such treaties typically provide for the immediate creation of a commission to prepare the constitution and describe the operation of the organization. See id.

123. Vienna Convention, supra note 111, art. 18. A nation that accedes to a treaty assumes the same obligation. Id. Accession therefore places the acceding nation in the same position as signatories in terms of generating an obligation not to defeat the object and purpose of the treaty. If this were not so, then an acceding nation would have expressed its intention to be fully bound without taking on any obligation to

be partly bound, and acceding nations would then have a different set of pre-entry-into-force obligations from ratifying nations (who are by definition signatories, and thus already partially bound).

124. Id., art. 49 ("If a State has been induced to conclude a treaty by the fraudulent conduct of another negotiating state, the State may invoke the fraud as invalidating its consent to be bound by the treaty."). Article 50 of the Vienna Convention treats corruption of a state's representatives at the bargaining table. See id., art. 50 ("If the expression of a State's consent to be bound by a treaty has been procured through the corruption of its representative directly or indirectly by another negotiating State, the State may invoke such corruption as invalidating its consent to be bound by the treaty."). Article 51 declares that coercion of a state's representative invalidates the expression of the state's consent, whereas article 52 declares that coercion of the state itself invalidates the treaty as a whole. See id., art. 51 ("The expression of a State's consent to be bound by a treaty which has been procured by the coercion of its representative through acts or threats directed against him shall be without any legal effect."); id., art. 52 ("A treaty is void if its conclusion has been procured by the threat or use of force in violation of the principles of international law embodied in the Charter of the United Nations.").

125. Id., art. 53.

126. See J.F. O'Connor, Good Faith in International Law (1991); Oscar Schachter, International Law in Theory and Practice 129 (1991) (stating that nations assuming international obligations generally intend to carry out such obligations in good faith). Once a treaty enters into force, those who have expressed full consent are, absent any invalidating factors, clearly bound to comply with that treaty in good faith. Vienna Convention, supra note 111, art. 26 ("Every treaty in force is binding upon the parties to it and must be performed by them in good faith."); see also id., art. 31(1) ("A treaty shall be interpreted in good faith."). I am discussing here the obligations of the parties during negotiations and thus the

obligations that exist before the relevant treaty has entered
into force. One must therefore look beyond articles 26 and 31
of the Vienna Convention for a good-faith obligation applicable
to pre-signature negotiations, although one may of course argue
that the good-faith obligations to comply with and to interpret
treaties in force is a manifestation of a general obligation to
conduct one's international legal affairs in good faith. See
O'Connor, supra, at 124; David Koplow, Parsing Good Faith: Has
the United States Violated Article VI of the Nuclear
Non-Proliferation Treaty?, 1993 Wis. L. Rev. 301, 368-71
(discussing good faith in the international legal context).

The Vienna Convention mentions good faith in connection not
only with compliance obligations but also in connection with the
assertion that an obligation to comply may not exist when the
expression of a nation's consent to be bound violates "its
internal law regarding competence to conclude treaties." Vienna
Convention, supra note 111, art. 46(1). Such an assertion is
well-founded only when the violation "was manifest and concerned
a rule of its internal law of fundamental importance." Id. "A
violation is manifest if it would be objectively evident to any
State conducting itself in the matter in accordance with normal
practice and in good faith." Id., art. 46(2).

127. For a discussion of what constitutes a good-faith
     effort
to pursue treaty negotiations aimed at nuclear disarmament,
rather than what constitutes good faith in the actual conduct of
negotiations, see Koplow, supra note 126, at 367-79.

128. I have made one such argument myself, though somewhat
elliptically. See Setear, An Iterative Perspective on Treaties,
supra note 3, at 211 (arguing that the rule of jus cogens is
sensible because it invalidates treaties that represent a
defection from the international system). One might also
imagine treaties of alliance among aggressive nations as
something other than cooperation from the point of view of the
international system. See, e.g., Treaty of Nonaggression
Between Germany and the Union of Soviet Socialist Republics,
Aug. 23, 1939, Germany-U.S.S.R., reprinted in Nazi-Soviet
Relations 1939-1941, at 76-77 (Raymond J. Sonntag & James S.
Beddie eds., 1948).

129. In a report to the General Assembly of the United Nations, the International Law Commission offered an alternative proposal that would have required states to refrain from defeating the object and purpose of a treaty upon their agreement to enter negotiations for the construction of that treaty. Restatement (Third) of the Foreign Relations Law of the United States § 312, reporter's note 6 (1987). This proposal, however, was rejected by the delegates to the United Nations Conference on the Law of Treaties (the conference that produced the Vienna Convention). Id. Had that proposal been accepted, the three-iteration view of the treaty process proposed immediately below--negotiation, signature, entry into force--would be more difficult to maintain because a nation's obligations in the negotiation and signature phases would overlap substantially.

130. The Vienna Convention does, however, set forth a possible exception concerning the primary obligation at issue during the signature iteration. A nation that has signed a treaty is obliged to refrain from defeating the object and purpose of the treaty "until it shall have made its intention clear not to become a party to the treaty"; moreover, a nation that has ratified or acceded to a treaty is bound "pending the entry into force of the treaty and provided that such entry into force is not unduly delayed." Vienna Convention, supra note 111, art. 18.

Undue delay of entry into force would certainly require some reformulation of the three-iteration framework presented here. All ratifying or acceding nations would by undue delay be relieved of their obligation not to defeat the object and purpose of the treaty. The three-iteration framework of the text, which assumes that a signing or acceding nation is obliged not to defeat the object and purpose of the treaty throughout the signature iteration, would therefore need some reworking.

A signatory that makes clear its intention not to become a party to the treaty complicates the analysis in a different fashion. In that case, the action of an individual nation relieves *that nation* from its obligation not to defeat the object and purpose of the treaty but does not affect the

obligations of other nations. Contrast this situation with undue delay of entry into force, in which the inaction of a group of nations (in failing as a group to register enough ratifications or accessions for entry into force) simultaneously relieves *all those nations* of their obligations not to defeat the object and purpose of the treaty. Below, I address a structurally similar problem in which the relevant nations retain their obligations throughout the iteration but simply fail to adhere to them. See infra note 131 and accompanying text.

131. One might argue that it would constitute bad faith for a
nation to participate at all in negotiations if that nation knew in advance that, after signing any resulting treaty, it would take steps to defeat the object and purpose of that treaty. Such behavior suggests fraud. A nation could, however, approach negotiations with reasonable skepticism about the likelihood that bargaining will produce an acceptable solution. During negotiations, that nation could honestly express its views and seek a compromise, but later find its skepticism confirmed when the text proves not to represent a solution to which that nation wishes to consent. This situation would seem to me to constitute good faith, not bad faith. More generally, if good faith in negotiations *by itself* implied a duty to sign the resulting treaty, then actual signature would be superfluous from a legal standpoint. That is clearly not the case, however.

132. This is the example given in the Restatement (Third) of
the Foreign Relations Law of the United States § 312, cmt. i (1987). The Restatement remarks that "[i]t is often unclear what actions would have [the] effect [of defeating the object and purpose of the agreement]." Id.; see also Certain German Interests in Polish Upper Silesia (Ger. v. Pol.), 1926 P.C.I.J. (ser. A), No. 7, at 30 (May 25) (Germany cannot transfer state property to private persons in territory that a signed but unratified treaty was to transfer to Poland).

133. I note here three complexities, resulting from atypical
situations, that I treat in more detail below. First, a

particular nation may qualify its acceptance of the rules set forth in a particular text by entering reservations at the time that nation signs or ratifies or accedes to the treaty. Therefore, even nations at the same stage of the treaty process do not invariably share identical obligations. See infra Part IV.B.1.a. Second, nations sometimes collectively amend the text of treaties after that treaty's entry into force, and such amendments do not always fully bind all those originally fully bound by the treaty. See infra Part IV.B.1.b. Third, nations do not always sign an agreement simultaneously, nor do they always ratify or accede to an agreement simultaneously. This complication blurs the otherwise sharp boundaries between the various iterations. See infra Part IV.A.

134. On rare occasions, however, these two iterations merge.
See Setear, An Iterative Perspective on Treaties, supra note 3, at 149 n.41 (describing the possibility that signature constitutes full consent).

135. Not all ratifying or acceding nations necessarily ratify
or accede before the entry into force of the treaty. When some signatures occur after other signatures, the possibility that indications of full consent will occur both before and after entry into force means that there is at least the potential for an overlap in time between the various iterations. This overlap renders the task of defining iterations more difficult. See infra Part IV.A.

136. In an earlier work, I described the treaty process as consisting of four iterations--negotiation, signature, ratification, and compliance--and stated that "[a] decision to participate in negotiations, to sign, to ratify, or to comply after entry into force is cooperative." Setear, An Iterative Perspective on Treaties, supra note 3, at 191. The view set forth here, in contrast, involves only three iterations--negotiation, signature, and entry into force. Both the three-iteration and the four-iteration views of the treaty process involve a negotiation and signature iteration. The three- iteration view also includes an "entry-into-force" iteration, while the four- iteration view includes (in addition

to negotiation and signature) a "ratification" iteration and a "compliance" iteration.

The differences between the three-iteration view and the four-iteration view stem from my further thoughts (prompted by comments from those who read earlier versions of the current piece) on the difference between a formally oriented view of the treaty process and an obligation-oriented view of that process. The "entry-into-force" iteration is intended as equivalent to the "compliance" iteration in my older view, but I use "entry into force" rather than "compliance" because, as I now realize, *all* of the iterations should (and, in the three-iteration view, clearly do) generate compliance obligations.

The reason for dropping the "ratification" iteration is similar. Ratification is a distinct formal step that an individual nation takes, but that step does not *by itself* result in any new international legal obligations. Ratification (or accession) is a *necessary* step on the road to full-blown treaty obligations, but a nation's ratification is not a *sufficient* step to result in such obligations: one nation's ratification (or accession) leads to entry into force only if enough *other* nations have also ratified (or acceded to) the treaty. Until that entry into force, a ratification does not involve any international legal obligations not already incumbent upon a signatory. Because ratification (or accession) does not by itself generate new international legal obligations, the use of ratification as a measure of the boundaries of an iteration is inappropriate.

A part of the earlier work that focused on obligations is consistent with my current analysis. See id. at 193-201 (describing the increase in obligations at "each of the four iterations" in the treaty process but in fact discussing only three changes in obligations).

137. As previously mentioned, I treat as axiomatic the assumption that international cooperation is a PD. See supra text accompanying note 128.

138. See ICRW, supra note 118.

139. Some activities--for example, neglecting reporting requirements--may violate the treaty's terms but not defeat its object and purpose. If no such requirements existed, there would be no difference between the obligation to refrain from defeating the object and purpose of the treaty and the obligation to comply in full with all the terms of a treaty after it enters into force.

140. All of these values are assumed to take into account any discounting related to probabilistic phenomena, such as the likelihood that nations will cheat on treaty provisions, or on intertemporal phenomena, such as the fact that the benefits of the agreement are likely to accumulate over time.

141. I discuss below the effects of modifying this (and other) assumptions about the chip values.

Note that, in a world without transaction costs, both exemptions would presumably be granted. Nation A could, for example, offer 130,000 chips to nation B to grant the auto-manufacturing exemption; that amount is more than the value to nation B of not allowing the exemption (120,000 chips) but less than the value to nation A of obtaining the exemption (150,000 chips), so the transfer of 130,000 chips from nation A to nation B in exchange for the exemption would increase the net wealth of both parties. Nation A should therefore make such an offer, and nation B should accept the offer. A similar argument would apply to the brick-making exemption desired by nation B, though the range of wealth-increasing transfers in that case would be between 60,001 chips and 74,999 chips.

142. The numbers in Figure VI represent the payoffs in thousands of chips. Cf. supra Figures II-V.

143. Note that I assume throughout this analysis that nation B garners no benefits from the bribe given to its representative, nor from the expenditures necessary to construct the computer program and create the data therefor. One might argue that nation B will benefit in some way when the bribed representative spends the money, at least if that representative

does so in nation B, or that the expenditures relating to the computer program will bring some benefit to the macroeconomy of nation B. Nonetheless, I simply assume that these outlays exclusively represent costs to nation B.

144. See generally Setear, An Iterative Perspective on Treaties, supra note 3, at 153 n.63 (discussing treaties specifying particular times for reevaluation of success or operation of provisions); id. at 224-27 (discussing nuclear arms control agreements encouraging future iterations).

145. The analogous treatment of ratification in the formally-oriented perspective and compliance in good faith with all treaty obligations in the compliance-oriented perspective assumes that the treaty in question actually enters into force. Ratification does not by itself oblige a nation to comply in good faith with all provisions of a treaty. See supra notes 118-119 and accompanying text.

146. See infra Part IV.

147. For representative examples of such treaties, see supra note 106.

148. See, e.g., Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare, June 17, 1925, 26 U.S.T. 571, 94 L.N.T.S. 65; Convention on the Prohibition, Production, Stockpiling, and Use of Chemical Weapons, opened for signing Jan. 13, 1993, U.N. GAOR, 47th Sess., Supp. No. 27, app. i, U.N. Doc. A/4/27 (1992).

149. See, e.g., Extradition Treaty, May 4, 1978, U.S.-Mex., 31 U.S.T. 5059.

150. See, e.g., ICRW, supra note 118.

151. See, e.g., Treaty on the Limitation of Anti-Ballistic Missile Systems, May 26, 1972, U.S.-U.S.S.R., 23 U.S.T. 3435 [hereinafter ABM Treaty].

152. Realists, who comprise the dominant school in the post-World War II study of international relations, do not believe that international legal rules independently influence national behavior at all. See Burley, supra note 3, at 207-08. Realists are generally skeptical of the assertion that nations comply with international law, and any compliance with international legal rules actually observed is seen by Realists as a coincidence stemming from extra-legal considerations of national interest. The relevance of formal statements of an intent to adhere to international legal rules, without any consideration of whether nations actually comply with those rules, is therefore likely to be seen by Realists as an empty endeavor.

153. See, e.g., Abram Chayes & Antonia Chayes, The New Sovereignty: Compliance with International Regulatory Agreements (1995); Keohane, Compliance with International Commitments, supra note 3, passim.

154. ABM Treaty, supra note 151, art. VI(b).

155. Agreed Statements Regarding the Treaty on the Limitation of Anti-Ballistic Missile Systems, May 26, 1972, U.S.-U.S.S.R., para. F, 23 U.S.T. 3456. Article III allows two clusters of ABM systems within certain geographical limits; article IV permits the development and testing of ABM systems and their components; article V restricts pre-existing and future weapons systems to minimize their potential use in ABM systems. ABM Treaty, supra note 151, arts. III-V.

156. See generally Ralph Earle II & John Rhinelander, The Krasnoyarsk Radar--A "Material Breach" of the ABM Treaty?, Arms Control Today, Sept. 1988, at 9; Catherine Girrier & Allan Krass, Disproportionate Response: American Policy and Alleged Soviet Treaty Violations (1987).

157. Greg Dalton, Violation: The Soviet Abalakovo Radar, in Compliance and the Future of Arms Control 105, 107 (Gloria Duffy et al. eds., 1988)

158. Id. at 108 (footnote omitted; parentheses in

original).

159. Id. at 109-12.

160. Id.

161. See Treaty on Conventional Armed Forces in Europe, Nov.
19, 1990, S. Treaty Doc. No. 102-8 (1991), reprinted in 30
I.L.M. 6 (1991). This treaty, commonly referred to as the CFE
Treaty, is a multilateral treaty among all the states that were
members of the NATO Alliance and the Warsaw Pact at the time of
its negotiation. David Sloss, It's Not Broken, So Don't Fix It:
The International Atomic Energy Safeguards System and the
Nuclear Nonproliferation Treaty, 35 Va. J. Int'l L. 841, 877
(1995).   The treaty drastically reduces the number of
conventional forces stationed in Europe. Id.

162. Countries may, of course, engage in tacit bargaining,
seeking to influence other states through behavior rather than
through formal agreements or diplomatic exchanges. See George
W. Downs & David M. Rocke, Tacit Bargaining, Arms Races, and
Arms Control (1990). This sort of behavior, which   occurs
without any statement as to its meaning, produces difficulties
in judging to what extent the behavior constitutes cooperation.

163. I realize the presentation of both sides of an issue
within a single article, especially in the absence of a
definitive resolution of this tension, breaks in some ways with
contemporary academic practice. I hope the reader will agree
that the gains afforded by this joint treatment exceed any
losses.

164. Charter of the United Nations, June 26, 1945, 59 Stat.
1031, 3 Bevans 1153, reprinted as amended in International Law:
Selected Documents 1 (Barry E. Carter & Phillip R. Trimble eds.,
1991) [hereinafter U.N. Charter].

165. See Charter of the United Nations, at n.2, in United
Nations, Multilateral Treaties Deposited with the
Secretary-General (ST/LEG/SER.E) (last modified Oct. 28, 1997)
<http://

www.un.org/Depts/Treaty/final/ts2/newfiles/part_boo/i_boo/i_1.ht
ml> [hereinafter Status of U.N. Charter].

166. U.N. Charter, supra note 164, art. 110(3) ("The
     present
Charter shall come into force upon the deposit of ratification
by the Republic of China, France, the Union of Soviet Socialist
Republics, the United Kingdom of Great   Britain and Northern
Ireland, and the United States of America, and by a majority of
the other signatory states.").

167. Status of U.N. Charter, supra note 165.

168. See id.

169. See Vienna Convention on the Law of Treaties, in
     United
Nations,   Multilateral   Treaties   Deposited   with   the
Secretary-General (ST/LEG/SER.E) (last modified Oct. 28, 1997)
<http://
www.un.org/Depts/Treaty/final/ts2/newfiles/part_boo/i_boo/i_1.ht
ml> [hereinafter Status of Vienna Convention].

170. See id.

171. Vienna Convention, supra note 111, art. 84 ("The
     present
Convention shall enter into force on the thirtieth day following
the date of deposit of the thirty-fifth instrument of
ratification or accession.").

172. See Status of Vienna Convention, supra note 169.

173. See Status of U.N. Charter, supra note 165.

174. See Status of Vienna Convention, supra note 169.

175. See Status of U.N. Charter, supra note 165.

176. See Status of Vienna Convention, supra note 169. The
United States, though not a signatory to the Vienna Convention,
has recognized the Convention as authoritative. See supra note

111.

177. See Vienna Convention, supra note 111, arts. 42(2), 54,
56(1); see also Setear, An Iterative Perspective on Treaties,
supra note 3, at 153 n.62 (interpreting article 54 of the Vienna
Convention); id. at 201-03 (discussing rules governing
termination from the iterative perspective).

178. See Vienna Convention, supra note 111, art. 54 ("The
termination of a treaty ... may take place ... at any time by
consent of all the parties after consultation with the other
contracting States.").

179. See id. ("The termination of a treaty ... may take
place
... in conformity with the provisions of the treaty."). For an
example of such a treaty, see Interim Agreement on Certain
Measures With Respect to the Limitation of Strategic Offensive
Arms, May 26, 1972, U.S.-U.S.S.R., art. VIII(2), 23 U.S.T. 3462
(stating that the treaty shall remain in force for five years
unless replaced sooner by more complete measures limiting
strategic offensive arms).

180. ICRW, supra note 118, art. V(1).

181. See Setear, An Iterative Perspective on Treaties,
supra
note 3, at 222.

182. Hague Convention on the Taking of Evidence, Mar. 18.,
1970, art. 41, 23 U.S.T. 2555, 847 U.N.T.S. 231; see also
Message from the President Regarding the Treaty on the
Limitation of Underground Nuclear Weapon Tests, S. Exec. Doc.
No. 94-N, at viii (1976) (treaty between U.S. and U.S.S.R. to be
automatically extended every five years unless one party
notifies the other of termination); European Convention for the
Protection of Human Rights and Fundamental Freedoms, Nov. 4,
1950, art. 65, 213 U.N.T.S. 221 (a party may terminate the
treaty only after remaining a party to the treaty for five
years).

183. See, e.g., U.S.-Egypt Treaty Concerning the Reciprocal Encouragement and   Protection of Investments, Sept. 29, 1982, art. XIII, 21 I.L.M. 927 (1982) (treaty to remain in force for ten years, after which one year's written notice will be sufficient for termination); Treaty on the Non-Proliferation of Nuclear Weapons, July 1, 1968, arts. VIII(3) and X(2), 21 U.S.T. 483, 729 U.N.T.S. 161 (operation of the treaty is to be reviewed every five years; the treaty is to remain in force for twenty-five years, at which time it will be reevaluated); Final Document on the Extension of the Treaty on the Non-Proliferation of Nuclear Weapons, 34 I.L.M. 959 (1995) (Treaty on Non-Proliferation of Nuclear Weapons was extended indefinitely at the end of the twenty-five year period).

184. Vienna Convention, supra note 111, art. 19 ("A state may
when signing, ratifying, ... or acceding to a treaty, formulate a reservation.").

185. Id. For a brief discussion of some treaties with specific terms prohibiting some or all reservations, see Setear, An Iterative Perspective on Treaties, supra note 3, at 151 n.50.

186. The Vienna convention defines a reservation as a nation's "unilateral statement ... purport[ing] to exclude or to modify the legal effect of certain provisions of the treaty in their application to that State." Vienna   Convention, supra note 111, art. 2(1)(d); see also Janis, supra note 110, at 20 (acknowledging that a reservation releases a party from the obligation against which the reservation is made).

187. See Stephen Weber, Cooperation and Discord in U.S.-Soviet Arms Control 5 (1991) (arguing that limitations on strategies available to players in a PD make that game an incomplete metaphor for U.S.-Soviet relations involving nuclear weapons); Jervis, supra note 40, at 329-332 (criticizing the use of only two choices and suggesting that a continuum of actions is available to nations); Wagner, supra note 90, at 331 (arguing that games must have more than two actions per player to model problems in international cooperation).

188. One may construct a game involving more than two

choices
per player that can fairly be called a PD. See Brian Betz,
Response   to   Strategy   and   Communication   in   an   Arms
Race-Disarmament Dilemma, 35 J. Conflict Resol. 678, 679-80
(1991) (discussing a PD with six choices per player); Norman
Frohlich & Joe Oppenheimer, When Is Universal Contribution Best
for the Group? Characterizing Optimality in the Prisoners'
Dilemma, 40 J. Conflict Resol. 502, 505-08 (1996) (describing a
PD with continuous rather than dichotomous choices). One may
also examine a PD involving more than two players at a time.
See Frohlich & Oppenheimer, supra, at 508-12 (developing a
general model   applicable to any number of players); Russell
Hardin, Collective Action as an Agreeable n-Prisoners' Dilemma,
16 Behav. Sci. 472 (1971) (discussing a ten-person PD); Daniel
G.   Arce   M.,   Stability   Criteria   for   Social   Norms   with
Applications to the Prisoner's Dilemma, 38 J. Conflict Resol.
749 (1994) (discussing a three-person PD). For a comparison of
bilateral and multilateral bargains within the framework of a PD
using examples from international relations, see Robert Pahre,
Multilateral Cooperation in an Iterated Prisoner's Dilemma, 38
J. Conflict Resol. 326 (1994).

      189. I assume for the sake of argument that these
reservations would be allowed under the terms of the treaty
itself and the Vienna Convention.   Cf. Setear, An Iterative
Perspective on Treaties, supra note 3, at 151 n.50 (describing
limitations on reservations).

      190. James Morrow, Game Theory for Political Scientists 128
(1994).

      191. See Wayne Sandholz, Choosing Union: Monetary Politics
and Maastricht, 47 Int'l Org. 1 (1993) (examining movement for
monetary   union   in   Europe   that   culminated   with   Maastricht
Treaty);   Time   Magazine,   May   31,   1993,   at   17   (describing
Denmark's   vote   in   favor   of   the   Maastricht   Treaty   after
concessions by other nations excluded Denmark from common
monetary and defense policy).

      192. The Vienna Convention provides that "[a] treaty may be
amended by agreement between the parties."   Vienna Convention,
supra note 111, art. 39.

193. The process of resolving treaty disputes about ambiguous
provisions may itself provide an opportunity for examining a set
of iterations occurring within the entry-into-force iteration
for a particular treaty. The default procedure for resolving
such disputes, as set forth in the Vienna Convention, is clearly
iterative in nature. See Vienna Convention, supra note 111,
art. 65; see also Setear, An Iterative Perspective on Treaties,
supra note 3, at 191-93 (describing dispute resolution under the
Vienna Convention). Many treaties modify this default procedure
in a similarly iterative manner. See Setear, An Iterative
Perspective on Treaties, supra note 3, at 215-16 (describing
dispute-resolution procedures set forth in various treaties as
supplements to, or substitutions for, procedures in the Vienna
Convention).

194. Vienna Convention, supra note 111, art. 60.

195. For an extensive discussion of this standard in light of
Institutionalist and Realist theories of international
relations, see Setear, Breach of Treaty, supra note 2, at 15-68.

196. See supra notes 187-188 and accompanying text.

197. See Setear, An Iterative Perspective on Treaties, supra
note 3, at 123-25 nn.229-30.

198. By the early 1980's, Oliver Williamson and others,
building upon the work of Ronald Coase, had developed
"transaction cost economics" and various notions of "relational"
contracting. See Oliver Williamson, Markets and Hierarchies
(1975); Charles J. Goetz & Robert E. Scott, Principles of
Relational Contracts, 67 Va. L. Rev. 1089 (1981); Oliver
Williamson, Transaction-Cost Economics: The Governance of
Contractual Relations, 22 J.L. & Econ. 233 (1979); see also
Ronald H. Coase, The Problem of Social Cost, 3 J.L. & Econ. 1
(1960) (examining the impact of various allocations of legal
rights in worlds of zero and non-zero transaction costs); Ronald
H. Coase, The Nature of the Firm, 4 Economica 386 (1937)

(postulating that firms exist to minimize the costs of contracting). A number of IR scholars have recently begun to adapt this work to the study of international political issues. See, e.g., Jeffry A. Frieden, International Investment and Colonial Control: A New Interpretation, 48 Int'l Org. 559, 564-67 (1994) (describing a "relational contracting" approach to various theories of investments by "home   countries"   in "host countries" and to conflict or cooperation between home and host countries); David A. Lake, Anarchy, Hierarchy, and the Variety of International Relations, 50 Int'l Org. 1, 10-18 (1996) (applying the theory of relational contracting to issues of "governance" such as alliances and empire).

Another recent effort applies theories of monopoly, refined in important ways between the mid-1950's and the mid-1970's, to the currently lively topic of democratic states and belligerent international behavior. David A. Lake, Powerful Pacifists: Democratic States and War, 86 Amer. Pol. Sci. Rev. 24 (1992); see also A.C. Harberger, Monopoly and Resource Allocation, 44 Amer. Econ. Ass'n Papers & Proc. 77 (1954) (recapitulating and refining methods of estimating deadweight loss from monopolies); Ann O. Kreuger, The Political Economy of the Rent-Seeking Society, 64 Amer. Econ. Rev. 291 (1974) (examining the effects of international trade restrictions on competitive rent-seeking and the resulting increase in deadweight losses); Gordon Tullock, The Welfare Costs of Tariffs, Monopolies, and Theft, 5 Western Econ. J. 224 (1967) (expanding deadweight losses to include the concept now called "rent-seeking").