

Tübinger Arbeitspapiere zur Internationalen Politik und Friedensforschung

Nr. 36

Andreas Hasenclever, Peter Mayer, Volker Rittberger

Is Distributive Justice a Necessary Condition for a High Level of Regime Robustness?

Center for International Relations/Peace and Conflict Studies,
Institute for Political Science,
University of Tübingen

Address: Melanchthonstr. 36, D-72074 Tübingen

Phone: ++49 (0)7071 29-78372

Fax: ++49 (0)7071 29-2417

WWW Homepage: <http://www.uni-tuebingen.de/ir/>

© Copyright: A. Hasenclever, P. Mayer, V. Rittberger
Tübingen 2000

ISBN 3-927604-33-X

Redaktion: Klaus Stodick, Volker Walz

WWW-Layout: Jürgen Plieninger

Editorischer Hinweis: Ab Nr. 23 sind die "Tübinger Arbeitspapiere" auch [auf dem Textserver der Universitätsbibliothek](#) als PDF-Dokument verfügbar. Wegen der Um- und Einarbeitungszeit ergibt sich leider ein Verzögerungseffekt: Dieser Text wird ab ca. Dezember 2000 auf dem Server zur Benutzung bereitstehen.

Table of Contents

[1. Introduction](#)

[2. Methodology: The Nature of Necessary Condition Hypotheses](#)

[2.1. The Meaning of Necessary Condition Hypotheses](#)

[2.2. The Practical Value of Necessary Condition Hypotheses](#)

[2.3. Testing Necessary Condition Hypotheses](#)

[3. The Justice Hypothesis: Problems of Conceptualization and Operationalization](#)

[3.1. Regime Robustness](#)

[3.2. Distributive Justice in International Regimes](#)

[4. The Nuclear Non-Proliferation Regime - a Counterexample to the Justice Hypothesis?](#)

[5. Conclusion](#)

[References](#)

[Footnotes](#)

1. Introduction [1]

The causal and normative relationship between order and justice is one of the most fundamental problems of political theory. According to Bull (1977: 90), three "ideal-typical doctrines" on this subject can be distinguished. Both conservatives and revolutionaries subscribe to the view that there is a trade-off between order and justice. They disagree, however, as to how this tension is properly resolved in political action. While conservatives believe in the normative priority of order over justice, revolutionaries regard themselves as justified in pursuing justice at the expense of order. By contrast, liberals deny the existence of the conflict which, in the opinion of conservatives and revolutionaries, forces us to take sides. For liberals, justice and order are mutually reinforcing attributes of a society, whether domestic or international, and ultimately order cannot be achieved without justice and *vice versa*.

The liberal doctrine is certainly attractive, as it teaches that we can have the best of both worlds. It may also be deemed naïve or ideological, however, and conservatives and revolutionaries have always been quick to attack liberalism on these grounds. The sub-field of political studies that has proven least susceptible to this optimistic view on the relationship between order and justice is international relations. Here, conservatism goes by the name of realism, which, for most of the time, has been the dominant school of thought informing the bulk of the theory-building activities and empirical research on the subject. Adducing the anarchical structure of the international system or flaws in human nature (or both), realists point out that order is a scarce good in international politics. Order is not incompatible with anarchy, but it is extremely difficult to achieve and to maintain under the conditions that obtain in the states system. Upholding or reconstructing orderly and peaceful relations among states, therefore, are tasks that absorb the energies of responsible leaders, leaving no room for the pursuit of less fundamental values such as justice. But not only is a sincere concern with international justice a luxury that even decisionmakers of powerful nations cannot afford, realists agree with revolutionaries that seriously pursuing a justice agenda in politics comes at a significant cost to oneself *and* others, adding that this cost is prohibitively high in the international arena. For the politics of international justice ultimately breed chaos and deadly conflict. Moreover, since justice presupposes a minimum amount of order, the revolutionary course in world politics is self-defeating: if the revolutionary's precepts are followed we are bound to end up with the worst of both worlds (Donnelly 1992, Smith 1986, Thompson 1992: 28, Welch 1993: 193f).

These objections do not directly challenge the liberal tenet that only a just order is a stable order or that justice is conducive to the stability of a given framework of order. Realists do not deny the trivial fact that a world order which every state or at least every great power perceives as just is *ceteris paribus* more secure and durable than one which leaves one or more key actors dissatisfied (Bull 1977: 91). Rather, they point to the unsurmountable obstacles to establishing such a

scheme (Rousseau 1964 [1782]). Liberals defending the "realism" of their outlook therefore argue that the self-appointed realists within international relations theory grossly exaggerate the transition problem. There is not "a jungle out there" dooming any attempt by states to reach consensus over principles of conduct that go beyond the most basic requirements of co-existence. Justice may be impossible in the state of nature, but world politics increasingly less resembles Hobbes's (1992 [1651]: ch. 13) gloomy depiction of a collective without common government. Far from conceding that they engage in some sort of utopian thinking liberals claim to have history on their side. For them, the cause of peace and justice is promoted by powerful social trends - democratization, the growth of economic interdependence, and international organization or governance - which have fundamentally changed the character of international relations since the seventeenth and eighteenth centuries, when realism may have provided a more adequate description of international politics (Doyle 1983, Kant 1970 [1795], Zacher 1992).

In the recent past, the study of the third pillar of the liberal peace - international governance - has been much furthered by focusing on international regimes (Kratochwil/Ruggie 1986). By creating convergent expectations for particular issue-areas of international relations (Krasner 1983: 2), international regimes offer the prospect of a piecemeal or patchwork solution to the problem of world order. Given that regime theory reflects ideas that are at the core of the liberal perspective on world politics (Keohane 1989b, 1990), [2] it is a remarkable fact that justice - both as an evaluative standard and as an analytical category - has never played a prominent role in the debate about international regimes (Baldwin 1993: 9, Hasenclever/Rittberger/Mayer 1997, Hurrell 1993: 65-68). This may in part be due to the tactical decision taken by leading neoliberal institutionalists such as Keohane (1984) to begin their theoretical enterprise with major concessions to realism. To strengthen their case these authors sought to demonstrate that the realist pessimism about international cooperation is not well-founded, *even if* one accepts for the sake of the argument central realist assumptions about world politics, including anarchy, the primacy of the state, and the irrelevance of moral motivations. [3]

In any case, discussions of the sources of regime robustness seldom explicitly or implicitly hark back to the earlier liberal claim that justice [4] fosters the stability of a political arrangement. There are some notable exceptions, though. For example, Puchala and Hopkins (1983: 66) contend that "[f]airer' regimes are likely to last longer, as are those that call for side payments to disadvantaged participants". They "expect that regimes founded on more egalitarian norms [...] would be more adhered to and less susceptible to revolutionary change". Similarly, O. Young (1994: 134) argues that "it is virtually impossible to achieve high levels of implementation and compliance over time through coercion" and that, therefore, "even great powers have a stake in the development of international institutions that meet reasonable standards of equity". Finally, in an attempt to account for the remarkable robustness many of the post-war regimes have exhibited in the face of the challenges posed by the collapse of the Soviet Union and the end of the cold war, Ruggie has pointed to the fact that these regimes are "multilateral" in character. Multilateralism, according to Ruggie (1992: 571), is an "institutional form which coordinates relations among three or more states on the basis of 'generalized' principles of conduct". He then proceeds to observe that "successful cases of multilateralism in practice appear to generate among their members [...] expectations of 'diffuse reciprocity'. That is to say, the arrangement is expected by its members to yield a rough *equivalence of benefits* in the aggregate and over time" (our emphasis) (see also Martin 1992: 784f). [5]

Not only are references to the role of justice as a promoter of regime stability scarce in the current literature, those who posit such a role usually do not take pains to spell out the rationale of this hypothesis. If they did, they would probably need to challenge the rationalistic assumptions that underpin the bulk of contemporary international relations theory (Hurrell 1993: 67). It is difficult (though perhaps not altogether impossible) to explain why rational (optimizing) egoists should exhibit a greater measure of loyalty to fair institutions than to biased ones, all other things equal. The hypothesis is more readily accounted for if it is assumed that individual and corporative actors have a *sense of justice* which in part motivates their social behavior. In a book-length study, Welch (1993) has shown how perceptions of injustice can help to explain the onset of major wars which are anomalies from a rationalistic perspective. In the context that he studies, the justice motive is a destructive force. But, as he points out, this effect is contingent on the institutional setting in which the sense of justice operates. A political order that is perceived to deny to some of the actors what they regard as their due tends to create dissatisfaction, resentment, and eventually violent conflict (or at least the withdrawal from cooperation). By the same token, an institutional arrangement that by and large meets these claims will *ceteris paribus* enjoy a greater measure of stability as it is not plagued by destructive impulses of this sort. Justice is conducive to regime robustness because an "actor operating on the basis of the justice motive will be *more* tolerant of other states' gains if they are perceived to be legitimate entitlements, and *less* tolerant of other states' gains if they are perceived to violate entitlements" (Welch 1993: 31, 191, see also Elster 1991: 120, Franck 1995: 8). [6]

One particular form that the conjecture that fairness promotes regime robustness assumes is the proposition that distributive justice is a necessary condition for a high level of regime robustness. This hypothesis is equivalent to the proposition that inequity in the allocation of benefits and burdens among regime members

guarantees ("is sufficient to") a low degree of institutional "staying power" in the face of exogenous shocks. It is this hypothesis which we shall focus on in this article. For convenience we will henceforth refer to it as the *justice hypothesis*. In the next section ([sec. 2](#)) we explore the meaning of hypotheses stating necessary conditions of some phenomenon of interest and contrast them with more familiar types of hypothesis. We also discuss ways of evaluating such hypotheses. [In the third section](#) we shift from matters of form to matters of substance, analyzing the two concepts that enter into our hypothesis: regime robustness and distributive justice. It will become obvious that the operationalization of these concepts is fraught with considerable difficulties. Although we are unable to present a rigorous test of the justice hypothesis in this article, [in the fourth section](#) we illustrate our methodological and conceptual argumentation by looking at an important security institution, the nuclear non-proliferation regime. We will suggest that this regime poses a serious challenge to the justice hypothesis, which, in principle, can be falsified by a single unambiguous counter-example.

[Back to the top](#)

2. Methodology: The Nature of Necessary Condition Hypotheses

The justice hypothesis is a proposition that depicts a certain state of affairs as necessary for some phenomenon to occur. Hypotheses of this kind may be called *necessary condition hypotheses*. Understanding and evaluating the justice hypothesis therefore requires clarity about the structure of necessary condition hypotheses. In this section we comment on three questions: 1) What is the meaning of necessary condition hypotheses and how do they differ from other types of hypothesis? 2) What is the potential use of a necessary condition hypothesis? 3) How can necessary condition hypotheses be tested?

2.1. The Meaning of Necessary Condition Hypotheses

A familiar type of general hypothesis in the social sciences states a relationship of covariance between two variables, which may or may not be interpreted as a causal relationship. For example, the level of income of a person may be hypothesized to increase with the number of years he or she has spent at school or other institutions of formal education. This hypothesis is a causal one if it is assumed that there is a causal mechanism or pathway that links the level of education with the level of income later in life (Little 1991: 25). If the hypothesis is interpreted causally, the variable specifying the cause is called the "independent variable", and the variable specifying the effect the "dependent variable". In view of their characteristic form, such hypotheses - whether or not they are given a causal interpretation - may be called *correlational hypotheses*. Correlational hypotheses state (or imply) that the greater the value of some variable A, the greater (or alternatively: the smaller) [[7](#)] the value of some other variable B. [[8](#)] Hence, the variables that enter into a correlational hypothesis must be measured at the ordinal-scale level or higher. Correlational hypotheses may be deterministic or probabilistic. If the hypothesis is deterministic it is assumed that, for any units x and y within the range of application of the hypothesis, the truth of the proposition that x has a greater value on A than y guarantees the truth of the proposition that x has a higher value of B than y. By contrast, a probabilistic correlational hypothesis is consistent with a (usually unspecified) number of deviations from this pattern.

The justice hypothesis as presented [in the first section](#) of this article is not a correlational hypothesis. It does not relate varying degrees of justice that might be realized in a regime to varying degrees of robustness certain or likely to be displayed by this regime. As a necessary condition hypothesis it states that some phenomenon (here: a high level of robustness) will not occur or be in existence *unless* some other phenomenon (here: a just distribution of gains and burdens in the regime) is present. Necessary condition hypotheses take the form: B only if A. Every necessary condition hypothesis is logically equivalent to a corresponding hypothesis that states a sufficient condition: "B only if A" is true if and only if "if non-A, then non-B" is true. Thus, the proposition that distributive justice is necessary for a high degree of regime robustness implies (and is implied by) the proposition that if a regime is unjust then it is moderately robust at best, or more simply: that unjust regimes fall short of a high level of robustness. Since logicians refer to if-then statements as "conditionals" (e.g. Quine 1982: 21), we may call hypotheses of this kind *conditional hypotheses*. [[9](#)] Conditional hypotheses may be rarer than correlational ones in contemporary social sciences (Dessler 1992: 3, n. 4), but this does not mean that they are absent or generally uninteresting. A case in point is the much debated democratic-peace hypothesis, whose proponents

hold that shared democratic institutions are a sufficient condition for a dyad of states to refrain from using military force against one another (or, to put it differently, that for any two states to fight each other in war it is necessary that at least one of them is not a democracy) (Brown/Lynn-Jones/Miller 1996).

In principle, conditional hypotheses lend themselves to a probabilistic interpretation just as well as correlational ones (Little 1991: 27). Thus, one might render the intended meaning of the justice hypothesis as follows: Unjust regimes are extremely likely to display a low-to-medium level of robustness. However, understanding necessary condition hypotheses in this way seems to miss the pragmatic point of such formulations. Scholars identifying some phenomenon as a necessary condition for some other phenomenon are deliberately making a bold claim, which, in our view, is unduly watered down if interpreted in probabilistic terms right from the start. *Consequently, we will understand and examine the justice hypothesis as a deterministic proposition, according to which injustice in a regime strictly precludes a high level of robustness.* [10]

If conditional hypotheses (in the standard interpretation) are strong in the sense that they are deterministic, this does not mean that they are strong in every respect. In particular, a conditional hypothesis does not represent a form of superparsimonious single-factor explanatory model of the phenomenon in question (even when it specifies only one sufficient or necessary condition). By accepting a hypothesis that states a *sufficient* condition for some phenomenon (if A then B), we do not commit ourselves to the view that A is the one cause of B and that no other variable is causally relevant to the occurrence of B. [11] Nothing whatsoever is implied by such a hypothesis regarding the possibility that there are further processes or events which are sufficient for B or that there are variables which influence the occurrence of B in the absence of A - variables which could be of the highest practical importance if A happens to be a rare event. Thus, the democratic-peace hypothesis is consistent with the possibility that, e.g., states with a nuclear second-strike capability do not fight each other or that states which share membership in a military alliance tend to settle their conflicts peacefully. Similarly, someone who advances and defends a hypothesis that specifies a condition A as *necessary* for some phenomenon B to occur (B only if A) does not *ipso facto* advocate a monocausal explanation of B. This would only be the case if he or she made the additional claim that A is a sufficient condition for B as well (i.e. if he or she asserted the *biconditional* A if and only if B). Contending that distributive justice is a necessary condition for the occurrence of a high level of regime robustness is not to rule out that justice is only part of a larger set of necessary conditions which, in conjunction but not singly, produce in international regimes a pronounced resilience to exogenous shocks.

2.2. The Practical Value of Necessary Condition Hypotheses

General causal hypotheses, if corroborated, may serve at least three practical purposes. They can be used to *explain* events or regularities, to *predict* the outcomes of processes, and to *alter* aspects of reality. These practical goals are to a certain extent interdependent: Often we are able to explain a given event most persuasively if we could have predicted this event had we been informed about the causally relevant factual attributes of the situation. Similarly, the ability to deliberately change some aspect of the world requires the ability to predict the outcomes of at least one of our behavioral options. Because of this partial interdependence, hypotheses that are useful for one purpose tend to be useful for the other two purposes as well and *vice versa*. [12] At first sight, it might seem that necessary condition hypotheses do not score well on this count because they do not satisfy the requirements of a standard model of scientific explanation, which - owing to these interdependencies - is relevant to the purposes of predicting and manipulating outcomes as well. In the following we try to show that this criticism is at best only partially justified and that necessary condition hypotheses can be valuable for explanatory, predictive, and technical purposes.

The model we are referring to is Hempel's (1965) famous "deductive-nomological model of scientific explanation". [13] According to Hempel's analysis, explanation is a deductive argument, where a statement describing the event to be explained (the *explanandum*) follows logically from two sets of premises (the *explanans*): one composed of statements describing the causally relevant aspects of the situation (initial conditions) and one composed of one or more (empirically corroborated) universal hypotheses.

THE DEDUCTIVE-NOMOLOGICAL MODEL OF EXPLANATION

Fa	initial condition (a is an F)
(x) (Fx → Gx)	universal hypothesis (every F is a G)
<hr/>	
Ga	Explanandum (a is a G)

Although necessary condition hypotheses of the form: B only if A are universal hypotheses, they do not seem to fit this model. The reason is that the complex statement [(B only if A) and A] does *not* entail B. Suppose we know, or accept the hypothesis, that only just regimes are robust and we also have reliable evidence of the fact that the food aid regime is a just regime - then this does not suffice to warrant the conclusion that the food aid regime is robust. Both premises might be true and yet the conclusion false.

Nevertheless, hypotheses stating a necessary condition for some phenomenon of interest *can* be used for explanatory purposes if explaining is construed along the lines of the deductive-nomological model. This is because, as we have seen earlier, every necessary condition statement B only if A is equivalent to (or synonymous with) a statement expressing a sufficient condition: viz. if non-A then non-B. As a result, although the law-like statement that B only if A does not help to explain the occurrence of B, it may well help to explain why, in a given situation, B *failed* to occur. If, for example, we wish to account for the low robustness of some commodity cartel, we could - if we have reasons to believe that the justice hypothesis is true - rightly point to the skewed distribution of benefits from cooperation that this regime has institutionalized. [14] Obviously, we could also use the justice hypothesis to *predict* that a given unjust regime will not prove robust when exposed to exogenous shocks. The predictive power of the hypothesis is of particular *practical importance*, because it has implications for the appropriate design of international institutions. Not only do states have intrinsic (moral) reasons for basing their cooperative arrangements on just principles; they also have compelling instrumental reasons for doing so, provided they are interested in a stable international order underpinned by, and composed of, stable international regimes.

Up to this point, we have assumed that the deductive-nomological model is an exhaustive account of explanation. If we relax this - hardly tenable - assumption and take into account more recent developments in the philosophy of science, we can strengthen the case for the explanatory power of necessary condition hypotheses by showing that hypotheses of the form B only if A may not only be useful in accounting for non-B, *but for B as well*. Hempel (1965: ch. 12) was well aware that all genuine explanation in science cannot be analyzed in terms of the deductive-nomological model. He therefore proposed a second model - the "inductive-statistical model" - which is appropriate when events are explained on the basis of probabilistic generalizations rather than strictly universal (deterministic) laws. The basic structure of inductive-statistical explanations is similar to that of deductive-nomological ones except that the relation between the *explanans* and the *explanandum* is no longer one of logical entailment. Rather, if the explanation is supposed to be persuasive, the probabilistic generalizations together with the initial conditions make it *very likely* for the *explanandum* to occur.

THE INDUCTIVE-STATISTICAL MODEL OF EXPLANATION

Fa	initial condition (a is an F)
p (G F) ¹⁵ near 1	probabilistic hypothesis (nearly all F are G)
=====	<i>makes very likely</i>
Ga	Explanandum (a is a G)

[15]

Hempel's postulate that statistical explanations are only valid when the premises make the occurrence of the *explanandum* very likely (i.e. when the conditional probability of the *explanandum* given the *explanans* is close to 1) has provoked much controversy. This so-called "high-probability criterion" was introduced to rule out that both the occurrence and the non-occurrence of the *explanandum* might be explained with reference to the same hypothesis. For example, the statistical hypothesis that 60% of all heavy smokers contract lung cancer might be used to explain both that Jones, a chain-smoker, became ill and that Smith, who smokes just as heavily as Jones, did not. After all, it was not unlikely that Smith would be spared by the disease. The high-probability criterion has been criticized by philosophers such as Suppes (1970) and Salmon (1984) on the grounds that it is at odds with the use of the term "causal explanation" in ordinary language. For we often accept explanations which rely on a statistical hypothesis that attributes a much lower probability to the event we wish to explain. Thus, even if only 6 out of 10 heavy smokers eventually contract lung cancer, the fact that Jones has long been a heavy smoker provides us with a perfectly adequate explanation of his disease (Hausmann 1991: 36f). What is critical is not the (conditional) probability of the *explanandum* as such but the *increase* in this probability that comes with the presence of the supposed causal agent. Under certain conditions, it is enough for heavy smoking to *make it more likely* that a lung cancer develops in order for this fact to support a statistical explanation of a person's illness. In philosophical terminology: it may be enough that heavy smoking have "positive statistical relevance" for the occurrence of cancer. [16]

This theory has important consequences for our discussion of the explanatory value of necessary condition hypotheses. This is because necessary condition hypotheses, although they are most naturally interpreted as specifying a deterministic relationship, have *probabilistic implications*. Certainly, the statement "B only if A" does *not* imply that $p(B \& \text{w}n j; A)$ is high (let alone close to 1) and therefore cannot support an inductive-statistical explanation of the occurrence of B as modelled by Hempel. It does imply, however, that $p(B \& \text{w}n j; A)$ is higher than $p(B)$ (assuming that A is a "true" condition, i.e. one which is not universally met in the population). [17] Consequently, since they entail that *A has a positive statistical relevance for B*, hypotheses of this type can be used in explanations not only of non-B, but of B as well. Salmon's (1984: 31f) own example refers to a nervous disease called paresis. It is held that this disease, the most famous victim of which was Nietzsche, affects only persons who suffer from an advanced form of untreated syphilis - which is, of course, a necessary condition hypothesis -, although only a small fraction of these persons actually contract paresis. Under these circumstances, it would be perfectly appropriate to explain the fact that Nietzsche came to suffer from paresis with reference to his advanced untreated syphilis, even though at no time it was "very likely" that he would develop this disease (Hausmann 1991: 43f). [18]

2.3. Testing Necessary Condition Hypotheses

Falsificationists such as Popper (1968) argue that scientific progress is a process of trial and error in which our body of knowledge is gradually cleared from false beliefs. Science involves formulating empirically falsifiable theories and subjecting them to hard tests. These tests do not aim at verifying or validating these theories but at refuting them. Theories that have proven false are eliminated, and new conjectures are put to the test. This meta-scientific perspective builds on the insight that there is a fundamental asymmetry between the verification and the falsification of universal generalizations including scientific theories. This asymmetry consists in the fact that falsification of a universal hypothesis in the light of empirical evidence is, in principle, possible, whereas verification is not. A single black swan can show that the hypothesis that all swans are white is false, whereas any number of white swans cannot show that the hypothesis is true, if only for the fact that we have no way of knowing what future swans will look like. Popper proceeds to argue that positive, theory-confirming evidence (the observation of a white swan) is less valuable than negative, theory-infirming evidence (the observation of a swan that is not white), for only the latter contributes to scientific progress in the sense just described.

This latter view has recently been criticised as inadequate for the social sciences (King/Keohane/Verba 1994: 100-105, Van Evera 1997: 43f). These critics point out that both positive and negative evidence is informative and alters our state of knowledge. Whatever the merits of this criticism with regard to the evaluation of social scientific hypotheses in general, however, *in the case of (deterministically interpreted) necessary condition hypotheses searching for counter-examples rather than trying to amass supportive evidence seems to be the strategy of choice*. As it will become obvious, this strategy is less destructive and less insensitive to confirming evidence as it may seem at first sight and is indeed capable of providing a hypothesis with some measure of positive support. The falsificationist strategy involves two steps. First, potential falsifiers are derived from the hypothesis. A potential falsifier is a singular (non-general) and verifiable statement which is inconsistent with the hypothesis under scrutiny (e.g. "There is, at some specific time and place, a non-white swan."). Hence, if a potential falsifier is

verified, the hypothesis from which it has been derived is falsified. Second, the truth or falsity of the potential falsifiers is established by empirical observation. If one of the potential falsifiers turns out to be true, the hypothesis (at least in its present form) is refuted; if none of the potential falsifiers derived proves to be an actual one, one starts all over again, i.e. new potential falsifiers are derived and confronted with the facts. The process only stops, when the attempt to falsify the hypothesis has met with success. As long as it continues, though, positive support for the hypothesis accumulates if only as a side-effect of an activity that is directed at disconfirming it. Thus, the corroboration of the hypothesis *is* a possible outcome of the falsificationist strategy (Opp 1995: 190-192). [[19](#)]

A necessary condition hypothesis B only if A states only one thing: there are no Bs which are also non-As. To say that distributive justice is necessary for a high level of regime robustness is tantamount to saying that there are no highly robust unjust regimes. Thus, every statement of the form "X is a highly robust unjust regime" is a potential falsifier of the justice hypothesis. Testing this hypothesis involves completing this sentence schema with names ("the international trade regime", "the ozone regime", etc.) and confronting the resulting statements ("The international trade regime is a highly robust unjust regime." etc.) with the empirical data. As soon as we come across a true statement of this form, we may reject the hypothesis, provided we can assert the truth of the falsifying statement with a sufficient degree of certainty.

But how do we select the names to enter into the schema, i.e. *which cases do we look at?* With regard to any given general hypothesis we can distinguish three kinds of singular (particular) statements: statements that (if true) falsify the hypothesis, statements that (if true) instantiate the hypothesis (without verifying it), and statements that (if true) are irrelevant to the truth or falsity of the hypothesis. An instantiating statement is one that is not just consistent with the hypothesis in the sense that it does not contradict the hypothesis. Rather, it provides an "example" for the hypothesis or another general statement that is implied by the hypothesis. Only (true) instantiating statements provide support for the hypothesis, whereas other statements consistent with the hypothesis are simply irrelevant. Since the justice hypothesis implies (indeed, is equivalent to) the universal generalizations "Unjust regimes are at best moderately robust" and "Highly robust regimes are just", both regimes that are unjust and brittle and regimes that are just and very robust instantiate or confirm the hypothesis. By contrast, observations of just regimes that are not robust neither confirm nor disconfirm the justice hypothesis.

Table 1: *The justice hypothesis: falsifying, instantiating and irrelevant cases*

regime cases	not very robust	highly robust
unjust	instantiating	falsifying
just	irrelevant	instantiating

Both distributive justice and robustness are complex concepts which require careful analysis and operationalization, before any meaningful test of the justice hypothesis can come off ([see sec. 3 below](#)). Moreover, even after a valid operationalization has been established, measurement of the two variables is likely to be difficult, costly, and uncertain. This fact makes it highly probable that few (if any) regimes are readily allotted to one of the four cells of the above matrix. At the same time, the very difficulty and costliness of the measurement exercise requires that we employ our resources efficiently and follow a research strategy that makes use of our prior, unsystematic knowledge of the justice and robustness of individual regimes. [[20](#)]

Since we attempt to refute the hypothesis, we do not select regimes which we believe to be just. For just regimes are either irrelevant to, or instances of, our hypothesis. Neither do we choose regimes which, given our present state of knowledge, are likely to prove both unjust and brittle. What we aim to do is to identify *plausible* potential falsifiers (Opp 1995: 194f). [[21](#)] *The cases of prime interest are apparently unjust regimes that seem to be nonetheless highly robust.* If no such cases are available, we turn our attention to seemingly unjust regimes that are not manifestly brittle and to apparently robust regimes which show no clear signs of satisfying criteria of distributive justice. [[22](#)] If none of these potential falsifiers turns out to be true, we move on to other regimes. Now regimes will be given priority that, even though less likely to be both unjust and highly robust, still hold (in the light of our present knowledge) a reasonable chance of displaying these attributes, etc. The process does not and cannot produce a definite confirmation of the hypothesis. *With every potential falsifier that is refuted, however, the degree of corroboration increases and we are justified in placing more confidence in the hypothesis.* Note also that, since, ideally, we proceed from more to less plausible potential falsifiers, the testing procedure approaches its saturation point at maximum speed as it were (although it is impossible to provide hard criteria for deciding whether or not this point is reached in a concrete situation). [[23](#)]

Of course, at any time a potential falsifier may be confirmed by the case study. Even then we need not immediately drop the hypothesis, though. This is for two reasons: *1)* We cannot rule out that the failure of the hypothesis in this case is due to measurement error. Consequently, we lay this case aside as an "anomaly" with a view of redoing the case study, once improved measures are available. This decision is the more legitimate the later the anomaly occurs in the testing sequence. *2)* While the first option leaves the hypothesis intact for the time being, the second involves its modification. Here we regard the failure of the hypothesis as an indication that we may have overstated its range of applicability (cf. King/Keohane/Verba 1994: 101, 103f). For example, if the justice hypothesis is falsified by a security regime, we may hypothesize that the relationship it specifies holds for low- but not for high-politics. Subsequently, we shift our attention to economic and environmental regimes, seeking as before to identify institutions which are both unjust and highly robust. [[24](#)]

[Back to the top](#)

3. The Justice Hypothesis: Problems of Conceptualization and Operationalization

From the logical point of view, testing a (deterministic) necessary condition hypothesis is a fairly straightforward business. Such a hypothesis specifies a logically possible (internally consistent) state of affairs - B and non-A - as empirically impossible, and the task of testing amounts to trying to show that, contrary to the hypothesis, this state of affairs does exist in empirical reality, that there *are* non-As which are Bs. Problems of representativeness and selection bias do not arise. Insufficient degrees of freedom, multicollinearity and the omission of variables are not an issue. In case of the justice hypothesis, however, these methodological advantages are coupled with quite substantial problems of measurement, which affect both variables - robustness and distributive justice - and are apt to quickly dissolve the initial impression that the hypothesis is easy to evaluate empirically.

3.1. Regime Robustness

Robustness (in our usage) is a property of international regimes that is similar to but more specific than regime stability. While stability may be defined rather broadly as the tendency of a regime to endure (whatever the causes of this tendency), robustness (or resilience) [[25](#)] is defined more narrowly as the capacity of a regime to absorb exogenous shocks or to withstand exogenous challenges (Hasenclever/Mayer/Rittberger 1996: 4-11). Robustness is internally related to the notion of an event that threatens the existence or the effectiveness of the regime from outside. *A regime that is resilient does not lose the loyalty or the support of its members when, due to some process or event in the physical or social environment of the regime, the prospects for cooperation deteriorate and incentives for deviant (non-cooperative) behavior increase.* This notion can be further developed in the light of the most common definition of international regimes, which depicts regimes as institutions composed of four types of components:

"Regimes can be defined as sets of implicit or explicit principles, norms, rules, and decision-making procedures around which actors' expectations converge in a given area of international relations. Principles are beliefs of fact, causation, and rectitude. Norms are standards of behavior defined in terms of rights and obligations. Rules are specific prescriptions or proscriptions for action. Decision-making procedures are prevailing practices for making and implementing collective choice." (Krasner 1983: 2)

With this definition in mind we may explicate the concept of regime robustness as follows: Given an exogenous challenge, the member states demonstrate their continual allegiance to the regime concerned by maintaining its principles and norms and continuing to comply with its provisions (some of which may have been adapted to the new situation). Changes *within* the regime (i.e. changes only in rules and procedures) in response to an exogenous shock do not indicate cracks in a regime's robustness provided that the altered injunctions are complied with as before; changes *of* the regime (i.e. changes in principles and norms), however, normally do, especially when these changes cannot be accounted for in terms of the explicit purpose of the regime, but merely reflect changes in the underlying power structure (cf. Krasner 1983: 3f). Still, a regime that is continued and complied with even after drastic revisions in its normative substance have been implemented has proven more robust than one that (in response to a similar shock) is dismantled altogether.

Several *methodologically relevant consequences* emerge from this definition of the term "regime robustness". To begin with, robustness is a dispositional concept describing a property that shows (or fails to show) *only when* certain conditions are met (i.e. only when an exogenous shock has occurred). As a result, it is impossible to rank all regimes by their degree of robustness owing to a lack of commensurability. Regime P which has survived a serious exogenous shock cannot be said to be more (or less or equally) robust than regime Q which, so far, has sailed in fine weather. Regime X which has faced and mastered two serious exogenous challenges has not proven more (or less or equally) robust than regime Y which has confronted - and survived - only one such challenge etc. Subsets of regimes are comparable, though, and partial orderings can be constructed. There are two issues here which require closer examination: *1)* By what criteria do we decide which (if any) of two commensurable regimes is more robust? *2)* When are two regimes commensurable?

As to the first issue, we have already outlined the answer. *Differences in the robustness of commensurable regimes* emerge from their memberships' responses to the exogenous challenges. These responses have two dimensions: *i)* continuity/change in the level of compliance with the regime; *ii)* possible normative regime change. A regime is most robust (relative to the strength of the shock it faced) if the level of compliance does not go down after the shock and no changes are made to the normative core of the regime. A regime is least robust if it breaks down in response to the shock. All regimes in a given class of commensurables fall somewhere on the continuum that is marked by these two poles. The table represents an ordinal scale that emerges naturally from these considerations.

Table 2: Scale of the variable "regime robustness"

Membership's Response to the Shock	Value of Variable 'Robustness'
no change of the regime and level of compliance not reduced	high degree of robustness
either change of the regime or lower level of compliance (but not both)	medium degree of robustness
change of the regime and lower level of compliance	low degree of robustness
breakdown of regime-based cooperation	not robust (brittle) [26]

As to the issue of *commensurability*, two regimes are comparable with respect to robustness if they encountered similar or equivalent (i.e. similarly strong) exogenous shocks. Moreover, for two regimes to fall into the same class of commensurables, the number of similar shocks they have confronted should not vary too much. For it would not be plausible to attribute the same level of robustness to two regimes, when one of them has survived, say, four shocks (of a given strength) before it collapsed in the wake of the fifth, whereas the other did not even survive the first one. Having undergone a similar number of similar shocks is sufficient for two regimes to be comparable with regard to robustness. It is not necessary, though. Obviously, a regime which survives a severe exogenous shock has proven more robust than a regime which collapses in the wake of a less serious challenge. In addition, to avoid perverse attributions we stipulate that regimes which are ineffective in the sense that their injunctions are generally ignored ("dead-letter regimes") are brittle by definition. Similarly, regimes which break down in the absence of any external stimulus are defined as being not robust. [[27](#)]

Another methodologically relevant consequence flowing from our definition of the term "robustness" is that we have to be clear what we mean by an *exogenous shock* and how we measure the strength of an exogenous challenge. We have already noted that an exogenous shock is an event which occurs in the environment of a regime and tends to reduce the willingness or ability of the regime members to cooperate in the issue-area governed by the regime. A robust regime prevents this tendency from prevailing. Since shocks of identical strength may be absorbed by one regime and cause the breakdown of another (reflecting the fact that some regimes are more robust than others), shocks must be defined independently of the outcome of the situation which they are predicated of. To do so, two (not mutually exclusive) approaches suggest themselves: an inductive one and a deductive one. The *inductive approach* searches for events that seem to have put historical regimes under considerable strain - events that appear to have weakened given regimes if not prompting their collapse - and then generalizes on these observations. The result is a (possibly hierarchically ordered) list including such events as a sudden deterioration of the overall political relationship of central regime members, a sharp recession, a dramatic change in the identity of the actors in the issue-area etc.

By contrast, the *deductive approach* draws on theories of international cooperation and regimes, specifying - at an abstract level - the kinds of events that, according to these models, tend to destabilize an ongoing cooperation among states. Subsequently, these abstract event-types are interpreted in terms of more concrete real-world phenomena. Perhaps most useful in this regard is a substream of the regimes literature which has been labeled the "situation-structural

approach" (Zürn 1992). *Situation-structuralists* use simple models from the theory of strategic choice - such as the Prisoner's Dilemma or the Battle of the Sexes - in order to depict and analyze the constellations of interests (payoff structures) that characterize given issue-areas in international relations (Hasenclever/Mayer/Rittberger 1997: 44-59). Drawing on results from game-theoretical research they use these models (together with other situational variables) to study such problems as the determinants of regime formation or the variation in regime form. Most important from our point of view are the theorems that pertain to the stability of cooperation and the variables that affect it. Thus, situation-structuralists make a fundamental distinction between two types of regimes: coordination regimes and collaboration regimes (Stein 1983).

Coordination regimes arise in situations such as the Battle of the Sexes; collaboration regimes are responses to situations resembling the Prisoner's Dilemma. This basic variable (regime type) affects the modal stability of the regime: Institutionalizing solutions to coordination problems, *coordination regimes* direct their members to adopt behavioral patterns which represent Nash equilibria of the game. As a consequence, no actor has an incentive to withdraw from the convention, once it is established, so long as no-one else defects. This property should make coordination regimes inherently stable. By contrast, cooperation in a Prisoner's Dilemma is not a Nash equilibrium and members have an incentive to cheat on their partners. This fact makes much greater demands on an institution designed to secure cooperation among sovereign states and is likely to result in *collaboration regimes* exhibiting a lower average degree of stability than coordination regimes.

A COORDINATION GAME: BATTLE OF THE SEXES

		Actor B	
		C	D
Actor A	C	1/1	3/4 _{P, N}
	D	4/3 _{P, N}	2/2 _N

A COLLABORATION GAME: PRISONER'S DILEMMA

		Actor B	
		C	D
Actor A	C	3/3 _{P+}	1/4 _P
	D	4/1 _P	2/2 _{N,M}

Legend: *C* = cooperative option; *D* = non-cooperative option (defect); *M* = maximin solution; *N* = Nash equilibrium; *P* = Pareto-optimum; *P+* (hatched cells) = qualified Pareto-optimum/cooperative outcome/outcome possibly institutionalized by regime (i.e. solution that is both Pareto-efficient and superior to the "natural outcome" resulting from independent decision-making); 4 = individually most desired outcome; 1 = individually least desired outcome [28]

Situation-structural analyses also provide clues as to what kind of events (changes in the structure of the game) are exogenous shocks to a regime. For each type of situation (Battle, PD, etc.) there are several factors which are hypothesized to play an important role in facilitating or inhibiting processes of regime formation. We can therefore define a *shock* as any event external to the regime which causes or involves the dissolution of a facilitating factor or the emergence of an inhibiting one. [29] The factors that facilitate and inhibit regime formation in coordination situations and in situations resembling the Prisoner's Dilemma overlap to a large extent, although they are not identical (Zürn 1992: 219). The following table derives from some of these factors a (most likely incomplete) typology of (abstract) shocks to a regime and provides examples of their (more concrete) realizations in world politics.

Table 3: *Types of exogenous challenges to international regimes*

Type of Exogenous Shock	Paradigmatic Realization/Example
Shrinking shadow of the future: a. rapidly increasing discount rate b. growing doubts as to the continuation of the game	a. sharp recession in member states (for environmental regimes) b. rapidly declining fishery
Change in the distribution of power (e.g. hegemonic decline)	U.S. in the 1960s and 1970s
Breakdown of mutual trust/increased sensitivity for relative gains	events that led to the Second Cold War
Emergence of new relevant actors unfamiliar with or unwilling to play by the rules of the game	disintegration of the Soviet Union
Substantial changes in preference orderings or utilities in major actors	end of the Cold War (for North-South regimes or for NATO)

It is best to regard the two approaches to defining exogenous shocks to international regimes not as mutually exclusive but as complementary. The deductive approach has the virtue of being more systematic than the inductive one. It offers the prospect of a theory of exogenous shocks which defines the most important categories of shocks and possibly also includes propositions about the relative weights of the various types of shock. At present, this theory is still in its early

stages, however. Moreover, even if all the types and weights were specified, there would still be the problem of relating these abstract concepts to empirical phenomena, and in this translation process unsystematic, intuitive and inductively acquired knowledge about exogenous challenges is likely to play an indispensable role. Conversely, a purely inductive approach is plagued by problems of causal attribution and therefore is likely to remain unsatisfactory as long as the supposed challenges are not understood theoretically and analyzed with the aid of categories derived from a general model of international cooperation. Unfortunately, neither of the two approaches - whether used singly or in combination - is likely soon to come up with a compelling solution to the problem of how to assess the relative strength of given exogenous shocks. In the meantime, it seems reasonable to treat the five types of shocks derived from the situation-structural theory of regimes as equivalent. Since these types are not mutually exclusive, the challenge posed by a given event (say, the end of the Cold War) is the more serious the more types of shock it exemplifies simultaneously.

Fortunately, not all of the above mentioned problems need to be solved in order to devise a meaningful test of the justice hypothesis. In particular, it is not necessary to order all (or a large sample of) regimes by their level of robustness. [30] A reasonable *procedure* is to begin by looking for cases that are highly robust by the measures specified in this section (i.e. regimes which have undergone several shocks or confronted a challenge simultaneously exemplifying several types of shock *and* have done so without either significant normative change or a reduction in the compliance following suit). Subsequently the regimes thus identified are examined with respect to their degree of distributive justice as defined in the next section. Any of these highly robust regimes that clearly fails to meet the criteria of distributive justice constitutes a counter-example to the justice hypothesis which must be regarded as a serious anomaly if not an actual falsifier of this hypothesis. In our empirical illustration [in the fourth section](#) of this article we shall proceed this way. [31]

3.2. Distributive Justice in International Regimes

Most students of international regimes agree that states create regimes in order to reap gains from cooperation. Depending on the substance of the principles, norms, rules and procedures that constitute the regime and on the character of the issue-area, the nature, size and allocation of these gains vary considerably across the range of actual and hypothetical institutions. The members of any given regime may benefit more or less equally, or some may do better than others, possibly even to the extent that some participants would have been better off without the regime. *Distributive justice* is a standard of evaluation which applies to the way institutions - whether domestic or international - allot benefits and burdens to the actors who participate in them. According to the justice hypothesis, only regimes which conform to this standard are highly robust. To test the hypothesis it is therefore necessary to clarify what is meant by "distributively just" in this context. [32]

In order to specify the defining features of a just regime we have the choice between two approaches which might be labeled "externalist" and "internalist", respectively. *Externalists* define "distributive justice" by an allocation principle (or a set of such principles) which is not necessarily identical with what the actors themselves regard as just. The term "distributive justice" is either nominalistically defined ("I henceforth use the word 'distributive justice' to mean such and such") or interpreted in terms of some (supposedly) objective conception of justice. It is usually regarded as the task of moral philosophy to provide us with a conclusive answer to the question of what is objectively just (whether or not this answer is congruent with the actors' moral beliefs). [33] Externalists therefore often rely on philosophical accounts of justice such as Rawls's (1971) theory of justice as fairness with its famous two principles when conceptualizing the variable distributive justice for regimes (e.g. Zürn 1987). The external approach may be perfectly and perhaps singly adequate for purposes of normative evaluation, i.e. when we wish to know whether or not a given regime is morally desirable and therefore deserves our support. However, it is *not* the best choice for the purpose of testing the justice hypothesis, which is not a normative statement but an *essentially analytical proposition about a causal relationship in the social world*. Although using some external standard as an operational definition of distributive justice and testing the hypothesis on that basis might produce an interesting empirical result in itself, the *validity* of the operationalization remains in doubt as long as it cannot be ruled out that this externally defined variable called "distributive justice" is uncorrelated with the normative beliefs and perceptions of justice entertained by the actors themselves. To put it differently, distributive justice (in this context) should be defined with a view to the actors' subjective conception of justice, because *otherwise we have no reason to expect the causal mechanism underpinning the justice hypothesis - viz. the justice motive - to operate in the first place*. Obviously, people's sense of injustice is aroused only by what *they* regard as unjust. Any definition of the term "distributive justice" as employed in the justice hypothesis, therefore, should be based on an internal standard of evaluation, i.e. on a distributive principle or on a combination of such principles which is considered valid and pertinent by the actors themselves. *Internalists* need not assume that

conceptions of justice vary from person to person (each having his own ideas about what is just) or that these understandings themselves are perfectly variable (any conceivable distributional pattern qualifying as a potentially just distribution). On the contrary, although there is often disagreement on what is just in a given situation, important aspects of the idea of distributive justice are not controversial. Consequently, there is no need to begin with a *tabula rasa* when trying to identify the conceptions of justice that are accepted in given community. Most fundamental, it is not in dispute that distributive justice is based on a principle of equality demanding that like cases be treated alike (Hart 1961: ch. 7). This formal dimension of distributive justice might also be stated in the following terms: only a *rule-governed* allocation of benefits and burdens is a just allocation (Bull 1977: 76). Obviously, *formal justice* is not sufficient to settle conflicts about the fair distribution of goods and bads, as any two persons are at the same time similar in countless regards and dissimilar in countless others. Formal justice therefore needs to be supplemented by an understanding of which similarities and dissimilarities, respectively, are morally relevant. In common morality we find a small number of *precepts of justice* which go some way toward filling this gap (Perelman 1990, Rawls 1971: sec. 47). These precepts are of two types: one precept requires institutions to conform to parity, the others demand some form of proportionality (H. P. Young 1994). *Parity* means that the shares that the parties receive are equal in size or equivalent in value; *proportionality* means that each receives in proportion to some morally significant variable. The most important proportionality precepts are the contribution principle ("To each according to his or her input") and the need principle ("To each according to his or her needs"). Social-psychologists describe these precepts and their variations in detail and use experimental designs to investigate the conditions under which individuals and groups tend to select one of these precepts rather than another (Deutsch 1975, Elster 1995, Hegtvedt 1992).

Recent work by Albin, Zartman and others (Albin 1993, Zartman 1995, Zartman *et al.* 1996) exemplifies the internalist approach to distributive justice in regimes. These studies are concerned with international negotiation but their results have important implications for regime analysis, since regimes for the most part emerge from a process of "institutional bargaining" in which agreement is reached on the constitutional contract of the regime (Young 1989, 1991). Zartman and his collaborators might be seen as meeting a challenge that Hurrell (1993: 68f) had posed several years ago, when he called on students of regimes to begin to view "notions of justice [...] as intrinsic to the process by which order is produced". They argue that notions of justice play an integral and irreducible part in states' attempts to settle or reduce their conflicts by way of negotiation. Most provocative is their claim that, in "the process of negotiating the exchange or division of the items contested between them, negotiators come to an agreement on the notion of justice which will govern this disposition; if they do not, negotiations will not be able to proceed to a conclusion" (Zartman *et al.* 1996: 80). In some cases, a ready-made principle of justice - usually some variant of parity - works as a "salient solution" or "focal point" (Schelling 1960) coordinating the expectations and the concession-making of the parties (Albin 1993: 230f). In other cases, the shared notion of justice which forms the basis of the agreement is the product of the negotiations themselves. Here, negotiation is a "search for justice" - the result of which is expressed in what Zartman calls a "formula" -, although the actors are selfish and power is not absent from the negotiating table (Zartman 1995: 892, 897f). [34]

This perspective invites the criticism that, by foregoing an external standard of justice, there is a danger of ending up with a definition of justice according to which just is whatever negotiators accept - in which case justice cannot explain acceptance and is redundant as an analytical category. Zartman and his collaborators, however, seem to avert this danger by advancing a (falsifiable) empirical hypothesis about the content of just (and therefore acceptable) formulas. They suggest that the formulas that are accepted are either issue-specific interpretations of one of a small list of precepts of justice (similar to the one above) or combine one or more of these precepts in a sort of compromise ("compound justice") (Zartman *et al.* 1996: 84). This specification of the range of possible negotiation outcomes is still fairly broad and inclusive; however, it appears to be sufficient to save the central claim these authors make - i.e. the claim that only formulas perceived as just are accepted - from tautology.

In the specific context of this article, however, their argument is problematic in yet another respect. According to Zartman and his collaborators, negotiations do not succeed and hence regimes are not created unless the actors select or form a common notion of justice which serves as the keystone of the regime ("territory for security", "unequal cuts in forces so as to restore parity", "embedded liberalism", etc.) The problem is that this is another necessary condition hypothesis, which seems to *entail* the justice hypothesis. To see this, consider the following: The hypothesis put forward by Zartman and his collaborators seems to imply that all regimes are just (from an internalist perspective), because only just formulas are endorsed and worked out in the first place. However, if all regimes are just, then all robust regimes must be just as well, which is another way of saying that only just regimes are robust (the justice hypothesis). Therefore, if Zartman and his collaborators are right, then the justice hypothesis is true *but it is also trivial* - given that we already know that regimes are fair as such.

However, we need not accept the conclusion that the justice hypothesis is trivial, even if we stick to the internalist approach to defining distributive justice. Once

more, we have *two options*: one is to reject Zartman's claim, the other is to show that the conclusion does not follow. We look at the two options in turn. The contention that all successful negotiations involve the acceptance by the parties of a just formula is likely to provoke grave doubts even in the minds of many internalists. Critics might point out that the internalism of Zartman and his collaborators is too narrowly focussed on the issue-area to which the negotiations pertain. The relevant internal standards of justice, it might be argued, are those of the international community as a whole (although, in some cases, regional communities may have further developed these principles or added new ones). As a consequence, principles of international distributive justice do not operate at the level of specific issue-areas (such as ozone depletion, conventional arms control in Europe, etc.) but at the level of what Keohane (1989b: 4) calls the fundamental "conventions" of international society or what Zartman (1995: 895, 899) himself refers to as "generalized formulas". Only principles which exist independently of the issues under negotiation have the *critical* potential that is implicit in the idea of justice and in morality more generally. Although these general principles of distributive justice influence the choice of the issue-specific formulas, the latter cannot be expected *generally* to reflect international justice conventions faithfully because of the intervention of power asymmetries and perhaps of other considerations besides justice such as efficiency. Therefore, the contingent outcomes of international negotiations are unreliable signposts when it comes to establishing the internal (i.e. community) standards of regime justice.

This argument, however, can be challenged on the grounds that it implicitly assumes that the states system is what Rawls (1971: 4f) calls a "well-ordered society". By this he means a society "effectively regulated by a public conception of justice" in which "everyone accepts and knows that the others accept the same principles of justice" and "the basic social institutions satisfy and are generally known to satisfy these principles". This assumption is a strong one, which is likely to recruit very sparse support in most IR quarters. Hoffmann (1981: 164) may be exaggerating somewhat when he remarks that in "the international milieu, there is a cacophony of standards", but he is more likely to be correct than those international communitarians who perceive a broad and substantial "overlapping consensus" on principles of international justice. Thus, it is questionable as to whether the concept of an international justice convention has any real-world referents or if it has, as to whether these ideas form a coherent whole. Keohane (1989b: 4) himself mentions only one international convention that seems to qualify as a principle of distributive justice: reciprocity. Reciprocity, however, seems too vague and flexible a criterion to underpin a distinction between just and unjust regimes. [35] Zartman (1995: 895, 899) mentions several "generalized formulas", but these are fairly heterogeneous and incoherent and will therefore often conflict in concrete cases ("first come, first served", "polluter pays", "self-determination only through referendum", etc.). Since there are no meta-level principles to decide which of these formulas trumps in conflict, it is again up to the negotiators to apply, to concretize or to ignore these overarching principles as they see fit.

If the first option we have when trying to avoid the conclusion that the justice hypothesis is trivial drops out, the second one is much more plausible and paves the way towards a workable concretization of the idea that (in the present context of study) the distributive justice of regimes should be understood in internalist terms. Here, the undesirable conclusion is avoided by questioning the logical soundness of the inference (rather than the truth of its premise), and it is indeed possible to accept without inconsistency Zartman's internalism and his hypothesis that agreed formulas reflect common notions of justice *and* to deny that all regimes are just. This is because the agreement on a formula or - in the language of regime theorists (Krasner 1983: 2, Müller 1993) - on principles and norms is not the end of the process of regime formation. In particular, there is no automatism that determines the shape, precision and degree of obligatoriness of the rules and decisionmaking procedures of the emerging regime once a set of principles and norms has been adopted. As a consequence, we can regard the principles and norms of the regime as embodying a common issue-related understanding of justice shared by the members of the regime *without* making a judgment on the fairness of the regime as such. *The regime is fair to the extent that the specific injunctions it requires its members to observe (i.e. the concrete rules of the regime) [36] reflect this normative consensus.* Just regimes involve a set of behavioral guide-lines that can be described as an unbiased translation of the principles and norms into verifiable obligations. Conversely, a regime in which only one item of the "trade" that produced the formula is represented at the level of the (comparatively) precise and stringent prescriptions and proscriptions is unfair. By implication, this perspective on regime justice, although radically internalist in that it does not look beyond the issue-area in question when defining what is just in the eyes of the actors, does not betray the notion that justice inheres a critical potential. Particular regimes may well be judged unfair - if only by their own standards. [37]

Along the lines of the argument discussed above, it might nevertheless be objected that this perspective dignifies regimes which are the product of power, coercion, and deception more than of a concern with international justice. But this objection rings somewhat hollow as long as there is no sufficiently firm and substantial consensus on what international justice means beyond individual issue-areas. Moreover, the role of power and coercion in the creation of regimes must not be overstated. As Young (1989: 362, 368f) as pointed out, structural or at least very common attributes of institutional bargaining, such as the rule of unanimity or the presence of a veil of uncertainty, narrowly circumscribe the influence of those factors. Finally, it should be added that the radically internalist perspective outlined in the previous paragraph has two advantages which make it particularly attractive in the context of a study which is concerned with the

requirements of an empirical test of the justice hypothesis.

1) It should not be too difficult to apply this criterion to individual cases and to decide whether or not a given regime is just. At a minimum, this approach should make it much easier to come to a judgment on the distributive justice of a regime than its alternative, which is burdened with the difficulty of justifying why one particular "justice convention" (or combination of such conventions) is chosen rather than another as a standard of distributive justice for the regime under consideration.

2) This conception of regime justice is *minimalist* in the sense that a regime that fails to meet this standard is likely to be unjust according to every other reasonable criterion of justice (whether internal or external) as well. In fact, this criterion seems to be implied by two of the most fundamental and least controversial of all moral ideas: the idea that one is obliged to keep one's promises (or *pacta sunt servanda*) and the idea that discrimination (or the selective application of a general rule) is unjust (i.e. the idea of formal justice). The minimalist (non-demanding) nature of our criterion means that we are much more likely to err on the side of justice than on the side of injustice. This fact, however, is of considerable value in the light of the methodological argument we have made above. [In the second section](#) of this article we have advocated a falsificationist strategy for evaluating the justice hypothesis. When applied to this hypothesis, this strategy directs us to search for regimes which are both highly robust and *unjust*. Our minimalism with respect to the concept of distributive justice, therefore, has the effect of *reducing* the probability that we have committed a measurement error, should we identify a given regime as a counter-example to, and consequently as a reason to reject, the justice hypothesis.

[Back to the top](#)

4. The Nuclear Non-Proliferation Regime - a Counterexample to the Justice Hypothesis?

In this section we argue that the nuclear non-proliferation regime (NPR) deserves the attention of scholars who wish to assess the empirical standing of the justice hypothesis, because this regime seems to combine two properties, which, according to this hypothesis, are mutually exclusive: a high level of robustness and a distribution of benefits and burdens that fails to meet the criterion of justice. Since the following observations are not the product of a thorough study of the NPR and its history, we do not claim to demonstrate conclusively that this regime is inconsistent with the justice hypothesis. Our goal is the more limited one of highlighting some basic and well-known facts about this regime which, in our view, justify its closer examination given that, as we have argued [in section two](#), testing of a necessary condition hypothesis should proceed by identifying and determining the truth value of plausible potential falsifiers of the hypothesis.

The goal of the *nuclear non-proliferation regime* is to prevent or at least to slow down the spread of nuclear weapons. The regime rests on several pillars the most important of which is the Treaty on the Non-Proliferation of Nuclear Weapons (NPT), which was concluded in 1968. Other sources of principles, norms, rules and procedures of the regime include the London Suppliers' Guide-Lines, the Statute of the International Atomic Energy Agency (IAEA), the safeguard rules in INFCIRC/66 and INFCIRC/153, the Comprehensive Test-Ban Treaty (CTBT) of 1996, and the various treaties establishing regional nuclear weapon free zones (Parker 1998: ch. 3). The NPR pursues its goals by creating a set of rights and obligations that differentiate between various categories of states: between states that possess nuclear weapons (nuclear weapon states - NWS) [[38](#)] and states that do not (non-nuclear weapon states - NNWS), and between suppliers (only some of which are NWS) and recipients of nuclear technology. NWS have undertaken not to transfer nuclear weapons to NNWS, not to threaten the use of nuclear weapons against such states, and to take steps towards disarmament. NNWS have pledged not to produce or acquire nuclear weapons and to allow the IAEA to monitor their compliance with this norm. All regime members retain the right to develop peaceful nuclear industries, and technologically advanced states are obligated to assist and to cooperate with less advanced states should the latter desire to use atomic energy for civilian purposes. Trade in nuclear technology is therefore admissible and encouraged. Suppliers, however, must not trade with states that have not accepted IAEA controls ("safeguards") on the use of their nuclear material and facilities. The IAEA has the right to report treaty violations to the Security Council which may take action if it comes to the conclusion that the member's non-cooperation constitutes a threat to international peace and security.

Highly robust...

Throughout its existence the NPR has been accompanied by doubts about its efficacy and durability. At the end of the cold war, however, Nye (1987: 401)

expected the NPR to prove "more durable" than other security regimes, and a few years later Ruggie (1992: 563f) could count the regime among those multilateral institutions that had shown to be markedly robust in the face of the dramatic upheavals of the late 1980s and early 1990s. In the course of the thirty years that the regime has been in place it has passed through several crises. But this is no indication of its low robustness. On the contrary, the challenges together with the fact that they could not destroy the regime testify to its staying power. Indeed, a glance at the history of the regime suggests that *it meets the criteria for a high level of robustness (as developed in section 3.1.) reasonably well.*

A minimum requirement of robustness is that the rules of the regime are generally *complied* with. This is because when gauging a regime's robustness we compare *inter alia* its pre- and post-shock levels of compliance. Without this precondition regimes which are not respected by their members would be robust almost by definition. Despite some spectacular cases of defection (see below), the NPR seems to meet this requirement. According to Parker (1998: 112f), the regime has proven significant in more than one respect: **1)** it has achieved an almost universal coverage (only three states - India, Pakistan, and Israel - have failed to accept NPT- or NPT-equivalent obligations); **2)** it seems to have induced change in the behavior of states (the diffusion of nuclear weapons has been much lower than experts had predicted in the 1960s; there are successful cases of "nuclear roll-back" such as South Africa); and **3)** compliance with its provisions has been relatively high (although there have been serious violations, these can be regarded as exceptions given the almost global membership). Over the years, the *substance of the regime* has not been static, but the changes that have occurred are changes within the regime rather than changes of the regime. The rights and obligations that form the normative core of the regime have been there from the beginning; only they have been concretised and elaborated over the years. For example, the nuclear disarmament norm is part of the NPT, although the progress that its addressees have made in this respect has been painfully slow and the CTBT has been signed only a few years ago.

Finally and most importantly, there have been *serious exogenous challenges* and they do not seem to have undermined the loyalty of the (vast majority of the) regime members. If anything they have fostered the willingness of the international community to strengthen and defend this important security regime. Thus, although several of such challenges occurred in the first half of the 1990s (the dissolution of the Soviet Union which created three more nuclear powers, locked in a severe security dilemma [Schimmelfennig 1994]; the detection of Iraq's secret nuclear armament program; North Korea's announcement to withdraw from the NPT), the other members of the regime did not reduce their commitment in order to regain unilateral leeway, but instead agreed to the NPT's indefinite extension in 1995. Similarly, the Indian and Pakistani nuclear tests in 1998 provoked international responses that evidenced at least as much the strength of the norm against nuclear testing as its weakness (Frank/Schaper 1999). None of this guarantees that the NPR will not fall into decay sometime in the near future, but for the time being it can be said to have proved remarkably robust by the criteria outlined in the preceding we have when trying to avoid the conclusion that the justice hypothesis is trivial section (see also Parker 1998: 116f). [39]

...even though unjust?

It is certainly not impossible to think of a moral defense of the NPR given that it seemed to have worked as an important constraint on the spread of nuclear weapons and thus reduced the danger of immensely devastating nuclear wars. But few have attempted a moral defense of the regime that specifically builds on the idea of distributive justice. Indeed, for Bull (1984: 10f), the NPR exemplifies the general truth that "[j]ustice is a particular virtue, and sometimes conflicts with other goals; in international relations order, or peace and security, is sometimes regarded as a higher good than justice". This reading of the regime is confirmed in the light of our minimalist criterion of distributive justice.

Use of this criterion involves forgoing the application of an external standard and exempting the formula that the regime is based upon from a critique in terms of justice. In the preceding section we have identified this formula with the normative superstructure of the regime or its principles and norms. A regime is unjust not because the bargain it has sprung from was unjust but because it has been implemented in a biased and discriminatory fashion. At the beginning of this section we have already outlined the "nuclear bargain" (Smith 1987: 257) without which there would have been no NPR or at least not the regime that we know. Basically, the NNWS traded their sovereign right to acquire nuclear weapons and to conduct their nuclear activities without outside supervision and interference for the NWS's pledge to give them access to the peaceful use of atomic energy, not to supply other states with nuclear weapons, and to work towards becoming NNWS themselves through the conclusion of disarmament treaties. This bargain may or may not satisfy a reasonable external standard of distributive justice. What makes the NPR an unjust regime by our criteria, however, is the selective and partial manner in which this bargain was worked out in the subsequent construction of the regime.

The bias that renders the regime unjust becomes manifest when we compare the - general and therefore less immediately obligatory and constraining - *norms* with

the - specific and therefore more immediately obligatory and constraining - *rules* of the regime (Müller 1989: 283-287). *The general pattern that shows is that those components of the bargain that were exclusively or primarily in the interest of the NWS had a much better chance of giving rise to concrete and binding rules and regulations than those terms of the trade that were the main concern of the NWS.* Thus, while the rules that specify the verification and export control norms have been steadily improved, refined and hardened, the norms that require cooperation in the development of a civilian nuclear sector and particularly the norms requiring the NWS to engage in a serious process of nuclear disarmament have found very little (if any) concretization. As a result, it is difficult to avoid the conclusion that the NPR, viewed as a whole, is (by its own standards) discriminatory and therefore unjust. [40]

It seems therefore that, at least in the case of the NPR, injustice *has been* compatible with a high degree of robustness. This is not to say that the elements of discrimination and injustice present in the NPR do not act as a strain on the regime and one that may well threaten its success and its existence in the long run. Indeed, the justice issue has been implicated in most of the shocks and crises that have plagued the regime in the last thirty years. Still, it cannot be overlooked that, so far, the regime has mastered these crises despite the double-standards it tends to apply when it comes to demanding on its members to make their actions suit to the words.

[Back to the top](#)

5. Conclusion

In this article we have looked at the hypothesis that only regimes that secure a fair distribution of the benefits and burdens of cooperation among their members prove highly resilient when confronted with exogenous challenges to their existence and effectiveness. The bulk of the article has been concerned with clarifying the methodological issues that are involved in empirically testing such a hypothesis. We have argued that a necessary condition hypothesis such as the justice hypothesis lends itself well to the application of a falsificationist strategy, which involves identifying and examining plausible potential falsifiers of the hypothesis under consideration. Although this methodological choice would seem to make testing the hypothesis a straightforward business, this has turned out not to be the case. What renders the empirical evaluation of this proposition difficult and its results uncertain are the problems that arise when it comes to conceptualizing and operationalizing its variables. Since robustness presupposes the notion of an exogenous shock that puts the regime to the test, it is necessary to specify what processes or events qualify as *prima facie* exogenous challenges to a regime and precisely how a regime reveals its degree of resilience when encountering such a challenge. This raises severe problems of commensurability, although not all of them need to be solved in a test of this kind. As to distributive justice, we have advocated a radically internalist approach to defining outcome fairness in international regimes. This approach seems to be the preferred one given that our purpose has not been one of normative evaluation (which may require an external standard) and that, at present, the international society seems to lack a firm and substantial consensus on general principles of distributive justice. The article concludes with a brief examination of a particular regime, viz. the nuclear non-proliferation regime, which we contend is a *plausible potential falsifier* of the justice hypothesis. While our analysis of the regime is rather sketchy and falls short of the requirements of a potentially falsifying case study, it suggests that the NPR may well be more than that and amount to an *actual falsifier* of the justice hypothesis; at a minimum, it points to an apparent anomaly that proponents of this hypothesis must address successfully if they wish to uphold their proposition without either qualification or modification.

[Back to the top](#)

References

- Albin, Cecilia 1993: The Role of Fairness in Negotiations, in: Negotiation Journal 9: 3, 223-244.
- Baldwin, David A. 1993: Neoliberalism, Neorealism, and World Politics, in: Baldwin, David A. (ed.): Neorealism and Neoliberalism. The Contemporary Debate, New York: Columbia University Press, 3-25.

Boulding, Kenneth E. 1978: *Stable Peace*, Austin: University of Texas Press.

Brown, Michael E./Lynn-Jones, Sean/Miller, Steven E. (eds.) 1996: *Debating the Democratic Peace*, London: MIT Press.

Bull, Hedley 1977: *The Anarchical Society. A Study of Order in World Politics*, Basingstoke: Macmillan.

Bull, Hedley 1984: *Justice in International Relations (Hagey Lectures)*, Waterloo, Ont.: University of Waterloo Press.

Collier, David 1995: *Translating Quantitative Methods for Qualitative Researchers: The Case of Selection Bias*, in: *American Political Science Review* 89: 2, 461-466.

Dessler, David 1992: *The Architecture of Causal Analysis*. Paper presented to the Center for International Affairs, Harvard University (unpub. ms.).

Deutsch, Morton 1975: *Equity, Equality, and Need: What Determines Which Value Will Be Used as the Basis of Distributive Justice?*, in: *Journal of Social Issues* 31: 3, 137-150.

Donnelly, Jack 1992: *Twentieth-Century Realism*, in: Nardin, Terry/Mapel, David R. (eds.): *Traditions of International Ethics*, Cambridge: Cambridge University Press, 85-111.

Doyle, Michael W. 1983: *Kant, Liberal Legacies and Foreign Affairs*, in: *Philosophy and Public Affairs* 12, 205-235 (part one), 323-353 (part two).

Dray, William H. 1957: *Laws and Explanation in History*, Oxford: Clarendon Press.

Efinger, Manfred/Mayer, Peter/Schwarzer, Gudrun 1993: *Integrating and Contextualizing Hypotheses. Alternative Paths to Better Explanations of Regime Formation?*, in: Rittberger, Volker (ed.): *Regime Theory and International Relations*, Oxford: Clarendon Press, 252-282.

Elster, Jon 1991: *The Possibility of Rational Politics*, in: Held, David (ed.): *Political Theory Today*, Cambridge: Polity Press, 114-142.

Elster, Jon 1995: *The Empirical Study of Justice*, in: Miller, David/Walzer, Michael (eds.): *Pluralism, Justice and Equality*, Oxford: Oxford University Press, 81-98.

Franck, Thomas M. 1995: *Fairness in International Law and Institutions*, Oxford: Oxford University Press.

Frank, Katja/Schaper, Annette 1999: *Das Nichtverbreitungsregime in der Krise? Massenvernichtungswaffen in Indien, Pakistan, Nordkorea und Irak*, in: HSFK/FEST/IFSH (eds.): *Friedensgutachten 1999*, Münster: Lit, 228-237.

George, Alexander L./Farley, Philip J./Dallin, Alexander (eds.) 1988: *U.S.-Soviet Security Cooperation: Achievements, Failures, Lessons*, Oxford: Oxford University Press.

Grieco, Joseph M. 1990: *Cooperation among Nations. Europe, America, and Non-Tariff Barriers to Trade*, Ithaca, NY: Cornell University Press.

Hart, H. L. A. 1961: *The Concept of Law*, Oxford: Clarendon Press.

Hasenclever, Andreas/Mayer, Peter/Rittberger, Volker 1996: *Justice, Equality, and the Robustness of International Regimes. A Research Design* (Tübinger Arbeitspapiere zur internationalen Politik und Friedensforschung 25), Tübingen.

<http://www.uni-tuebingen.de/uni/spi/taps/tap25.htm>

Hasenclever, Andreas/Mayer, Peter/Rittberger, Volker 1997: *Theories of International Regimes*, Cambridge: Cambridge University Press.

Hasenclever, Andreas/Mayer, Peter/Rittberger, Volker 1998: *Fair Burden-Sharing and the Robustness of International Regimes: The Case of Food Aid*, (Tübinger Arbeitspapiere zur internationalen Politik und Friedensforschung 31), Tübingen.

- Hasenclever, Andreas/Mayer, Peter/Rittberger, Volker 2000: Integrating Theories of International Regimes, in: *Review of International Studies* 26: 1, 3-33.
- Hausmann, Thomas 1991: *Erklären und Verstehen. Zur Theorie und Pragmatik der Geschichtswissenschaft*, Frankfurt a.M.: Suhrkamp.
- Hegtvædt, Karen A. 1992: When Is a Distribution Rule Just?, in: *Rationality and Society* 4: 3, 308-331.
- Hempel, Carl 1965: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press.
- Hobbes, Thomas 1992 [1651]: *Leviathan*, Cambridge: Cambridge University Press.
- Hoffmann, Stanley 1981: *Duties Beyond Borders: On the Limits and Possibilities of Ethical International Politics*, Syracuse, NY: Syracuse University Press.
- Holsti, Kalevi J. 1991: *Peace and War: Armed Conflicts and International Order*, Cambridge: Cambridge University Press.
- Hurrell, Andrew 1993: International Society and the Study of Regimes, in: Rittberger, Volker (ed.): *Regime Theory and International Relations*, Oxford: Clarendon Press, 49-72.
- Kant, Immanuel 1970 [1795]: Perpetual Peace, in: Forsyth, Murray G. *et al.* (ed.): *The Theory of International Relations*, London: Allen & Unwin, 20-45.
- Keohane, Robert O. 1984: *After Hegemony. Cooperation and Discord in the World Political Economy*, Princeton, NJ: Princeton University Press.
- Keohane, Robert O. 1989a: Closing the Fairness-Practice Gap, in: *Ethics and International Affairs* 3, 101-116.
- Keohane, Robert O. 1989b: Neoliberal Institutionalism. A Perspective on World Politics, in: *International Institutions and State Power. Essays in International Relations Theory*, Boulder, Colo.: Westview Press, 1-20.
- Keohane, Robert O. 1990: International Liberalism Reconsidered, in: Dunn, John (ed.): *The Economic Limits to Modern Politics (Murphy Institute Studies in Political Economy)*, Cambridge: Cambridge University Press, 165-194.
- King, Gary/Keohane, Robert O./Verba, Sidney 1994: *Designing Social Inquiry. Scientific Inference in Qualitative Research*, Princeton, NJ: Princeton University Press.
- Krasner, Stephen D. 1983: Structural Causes and Regime Consequences: Regimes as Intervening Variables, in: Krasner, Stephen D. (ed.): *International Regimes*, Ithaca, NY: Cornell University Press, 1-21.
- Kratochwil, Friedrich V./Ruggie, John Gerard 1986: International Organization. A State of the Art on an Art of the State, in: *International Organization* 40: 4, 753-775.
- Little, Daniel 1991: *Varieties of Social Explanation. An Introduction to the Philosophy of Social Science*, Boulder/San Francisco/Oxford: Westview Press.
- Martin, Lisa 1992: Interests, Power, and Multilateralism, in: *International Organization* 46: 4, 765-792.
- Martin, Lisa L./Simmons, Beth 1999: Theories and Empirical Studies of International Institutions, in: Katzenstein, Peter J./Keohane, Robert O./Krasner, Stephen D. (eds.): *Exploration and Contestation in the Study of World Politics*, Cambridge, Mass.: MIT Press, 89-117.
- Mill, John Stuart 1988 [1848]: *A System of Logic*, London: Duckworth.
- Moravcsik, Andrew 1992: *Liberalism and International Relations Theory (Working Paper Series 92-6)*, Cambridge: Harvard University, The Center for International Affairs.

- Müller, Harald 1989: Regimeanalyse und Sicherheitspolitik. Das Beispiel Nonproliferation, in: Kohler-Koch, Beate (ed.): Regime in den internationalen Beziehungen, Baden-Baden: Nomos, 277-313.
- Müller, Harald 1993: Die Chance der Kooperation. Regime in den internationalen Beziehungen, Darmstadt: Wissenschaftliche Buchgesellschaft.
- Müller, Harald 1994: Institutionen und internationale Ordnung, in: Krell, Gert/Müller, Harald (eds.): Frieden und Konflikt in den internationalen Beziehungen. Festschrift für Ernst-Otto Czempiel, Frankfurt/New York: Campus, 223-253.
- Nagel, Ernest 1961: The Structure of Science: Problems in the Logic of Scientific Explanation, New York: Hartcourt, Brace, and World.
- Nye, Joseph S., Jr. 1987: Nuclear Learning and U.S.? Soviet Security Regimes, in: International Organization 41: 3, 371-402.
- Opp, Karl-Dieter 1995: Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theoriebildung und praktischen Anwendung, 3rd ed., Opladen: Westdeutscher Verlag. Parker, Charles 1998: Security Regime Significance in the Post Cold War Context: The Case of the Nuclear Non-Proliferation Regime, Stockholm: The Swedish Institute of International Affairs.
- Perelman, Chaïm 1990: De la Justice, in: Oeuvres III: Ethique et Droit, Bruxelles: Editions de l'Université de Bruxelles, 13-86.
- Popper, Karl R. 1968: The Logic of Scientific Discovery, New York: Harper & Row.
- Puchala, Donald J./Hopkins, Raymond F. 1983: International Regimes: Lessons from Inductive Analysis, in: Krasner, Stephen D. (ed.): International Regimes, Ithaca/London: Cornell University Press, 61-92.
- Quine, Willard Van Orman 1982: Methods of Logic, 4th ed., Cambridge, Mass.: Harvard University Press.
- Ragin, Charles C. 1987: The Comparative Method. Moving Beyond Qualitative and Quantitative Strategies, Berkeley/Los Angeles/London: University of California Press.
- Rawls, John 1971: A Theory of Justice, Cambridge, Mass.: Harvard University Press.
- Rousseau, Jean-Jacques 1964 [1782]: Jugement sur le projet de paix perpétuelle, in: Oeuvres Complètes, vol. 3, Paris: Gallimard, 591-600.
- Ruggie, John Gerard 1992: Multilateralism. The Anatomy of an Institution, in: International Organization 46: 3, 561-598.
- Salmon, Wesley C. 1984: Scientific Explanation and the Causal Structure of the World, Princeton, NJ: Princeton University Press.
- Schelling, Thomas C. 1960: The Strategy of Conflict, Cambridge, Mass.: Harvard University Press.
- Schimmelfennig, Frank 1994: Arms Control Regimes and the Dissolution of the Soviet Union. Realism, Institutionalism and Regime Robustness, in: Cooperation and Conflict 29: 2, 115-148.
- Smith, Michael Joseph 1986: Realist Thought from Weber to Kissinger, Baton Rouge: Louisiana State University Press.
- Smith, Roger K. 1987: Explaining the Non-Proliferation Regime. Anomalies for Contemporary International Relations Theory, in: International Organization 41: 2, 253-282.
- Stein, Arthur A. 1983: Coordination and Collaboration. Regimes in an Anarchic World, in: Krasner, Stephen D. (ed.): International Regimes, Ithaca, NY: Cornell University Press, 115-140.
- Suppes, Patrick 1970: A Probabilistic Theory of Causality, Amsterdam: North-Holland.
- Thompson, Janna 1992: Justice and World Order: A Philosophical Inquiry, London: Routledge.

Van Evera, Stephen 1997: *Guide to Methods for Students of Political Science*, Ithaca, NY: Cornell University Press.

Walt, Stephen M. 1997: *Why Alliances Endure or Collapse*, in: *Survival* 39: 1, 156-179. Walzer, Michael 1987: *Interpretation and Social Criticism*, Cambridge, Mass.: Harvard University Press.

Welch, David A. 1993: *Justice and the Genesis of War*, Cambridge: Cambridge University Press.

Young, H. Peyton 1994: *Equity in Theory and Practice*, Princeton, N.J.: Princeton University Press.

Young, Oran R. 1989: *The Politics of International Regime Formation. Managing Natural Resources and the Environment*, in: *International Organization* 43: 3, 349-376.

Young, Oran R. 1991: *Political Leadership and Regime Formation. On the Development of Institutions in International Society*, in: *International Organization* 45: 3, 281-308.

Young, Oran R. 1994: *International Governance. Protecting the Environment in a Stateless Society*, Ithaca, NY: Cornell University Press.

Young, Oran R./Osherenko, Gail 1993: *Testing Theories of Regime Formation. Findings from a Large Collaborative Research Project*, in: Rittberger, Volker (ed.): *Regime Theory and International Relations*, Oxford: Clarendon Press, 223-250.

Zacher, Mark W. 1987: *Trade Gaps, Analytical Gaps. Regime Analysis and International Commodity Trade Regulation*, in: *International Organization* 41: 2, 173-202.

Zacher, Mark W. 1992: *The Decaying Pillars of the Westphalian Temple: Implications for International Order and Governance*, in: Rosenau, James N./Czempiel, Ernst-Otto (eds.): *Governance Without Government: Order and Change in World Politics*, Cambridge: Cambridge University Press, 58-108.

Zartman, I. William 1995: *The Role of Justice in Global Security Negotiations*, in: *American Behavioral Scientist* 38: 6, 889-903.

Zartman, I. William/Druckman, Daniel/Jensen, Lloyd/Pruitt, Dean G./Young, H. Peyton 1996: *Negotiation as a Search for Justice*, in: *International Negotiation* 1: 1, 79-98.

Zürn, Michael 1987: *Gerechte internationale Regime. Bedingungen und Restriktionen der Entstehung nicht-hegemonialer internationaler Regime untersucht am Beispiel der Weltkommunikationsordnung*, Frankfurt am Main: Haag und Herchen.

Zürn, Michael 1992: *Interessen und Institutionen in der internationalen Politik. Grundlegung und Anwendung des situationsstrukturellen Ansatzes*, Opladen: Leske + Budrich.

[Back to the top](#)

Footnotes

[1]

An earlier version of this article was presented at the 41st Annual International Studies Association Convention, Los Angeles, 14-18 March 2000. We wish to thank Steinar Andresen, Martin Beck, Andrew Bennett, Gary Goertz, Olav Schram Stokke, and Michael Zürn for helpful comments and discussion. Part of the research done for this paper was supported by a grant we gratefully received from the Deutsche Forschungsgemeinschaft.

[2]

For a more restrictive interpretation of the hard core of the liberal research program which excludes "regulatory liberalism" (i.e. institutionalism) see Moravcsik (1992).

[3]

In the course of his argument Keohane (1984: ch. 7) relaxed some of these assumptions including the premise that states are incapable of altruistic behavior. Moreover, he embarked on a discussion of the extent to which contemporary international regimes meet the requirements of distributive justice (Keohane 1984: ch. 11, 1989a). But these variations and additions were not further developed and by and large went unnoticed by other scholars contributing to the neoliberal theory of international regimes.

[4]

In this article we use the words "justice", "fairness", and "equity" interchangeably.

[5]

Similar views are expressed by Albin (1993: 240), Boulding (1978: x, 70f), Franck (1995: 7-9, 13, 25 n. 3, 32f), Grieco (1990: 47f.), Holsti (1991: 337f), Kratochwil/Ruggie (1986: 773), Martin/Simmons (1999: 105), Müller (1994: 232), Walt (1997: 161), and Zacher (1987: 177f). Some of these scholars do not explicitly refer to justice, equity, or fairness, but use broader normative concepts such as legitimacy or, as in Ruggie's case, more narrow ones such as equality or equivalence in the distribution of benefits. (As we shall see below, parity, i.e. equal distribution, *is* a justice principle, though only one among several.)

[6]

Another actor model which explains the instability of biased regimes is Grieco's (1990: 28, 37-40) "defensive positionalist". Here the causal mechanism that produces the relationship between "balanced gains" and robustness is fear and mistrust rather than a concern for justice. Grieco's model, too, is not rationalistic in the classical sense, as it assumes satisficing rather than maximizing behavior (Hasenclever/Mayer/Rittberger 1997: 131f). Finally, Walt (1997: 161) mentions domestic political competition as a transmission belt linking unfair burden-sharing with low levels of alliance stability.

[7]

To avoid unnecessary complication we concentrate henceforth on directly proportional correlation hypotheses leaving aside the inversely proportional (the more A the less B) variant.

[8]

Dessler (1992: 3, n. 4) describes these hypotheses as "generalizations that relate the variation in some outcome across a population of cases to variation in some independent variable".

[9]

This terminology is not entirely satisfactory for at least two reasons. First, some authors use "conditional hypothesis (or generalization)" to refer to relationships

which are hypothesized to hold only under certain circumstances. Conditional (or contingent) hypotheses in this sense take the form: if C then H, where "C" denotes a certain set of circumstances and "H" stands for "if A, then B" or some other type of hypothesis. Second and more important, it can be shown that, in the final analysis, correlational hypotheses (as presumably all law-like generalizations [Nagel 1961: 47f]) are if-then statements no less than conditional (i.e. necessary or sufficient condition) hypotheses. For example, the logical structure of ordinal-scale level correlational hypotheses (assuming direct proportionality) may be analyzed as follows: (x) (y) ($[A(x) < A(y)] \→ [B(x) < B(y)]$), which is another way of saying: for any x and for any y (where x and y are non-identical units belonging to the range of application of the hypothesis), *if* x takes a smaller value on the variable A than y, *then* it also takes a smaller value on the variable B. Hence, by "conditional hypotheses" we mean *simple* if-then statements (if A then B) which allow for no further analysis along such lines. Conditional hypotheses are also referred to as "covering law hypotheses" (Dessler 1992: 1, 3-5, Dray 1957: ch. 1, Nagel 1961: 570).

[10]

Conditional hypotheses should be distinguished from "general factor-claims" (Mill 1988 [1848]). General factor-claims refer to causal tendencies which operate without exception, although, in particular cases, they may be outweighed by other tendencies. For example, justice may invariably work in favor of regime robustness, even though the outcome "high robustness" will sometimes (or even often) not be observable in just regimes owing to the presence of countervailing factors (such as a low degree of institutional success or an inhospitable distribution of power among regime members). By contrast, conditional hypotheses refer to outcomes rather than unobservable tendencies, for which only indirect evidence can be mustered (Dessler 1992: 6-8).

[11]

In other words, sufficient condition hypotheses are consistent with the view that in the social world there are many instances of "equifinality" (George/Farley/Dallin 1988: 11, 17 n. 31) or "multiple causation" (Ragin 1987: ch. 2).

[12]

The interdependence is not perfect, since there are hypotheses which can be used for explaining but not for predicting (e.g. evolutionary theory). Moreover, prediction but not explanation can rely on non-causal hypotheses, i.e. propositions simply stating a constant conjunction of certain event-types (e.g. the regular relationship between barometer readings and changes in the weather).

[13]

For good accounts of the main features of this model as well as some of the criticisms it has attracted see, e.g., Haussmann (1991: ch. 1.2.) and Little (1991: 5f).

[14]

This is, of course, on the assumption that this characterization of the distributional dimension of the regime is accurate and that the particular form of inequality to be found in this regime is indeed unjust. Moreover, the explanation is only satisfactory if the relationship that is specified by the hypothesis is a *causal* one. That this is the case is most convincingly demonstrated by identifying, and producing evidence of, the causal mechanism that links injustice with low levels of robustness (Little 1991: 24f).

[15]

This term stands for "the conditional probability of G given F".

[16]

F has positive statistical relevance for G if and only if the conditional probability of G given F is greater than the absolute probability of G, formally: $p(G \&zwjn; F) > p(G)$.

[17]

This can be demonstrated as follows: In any population in which "B only if A" is true there are no units which are both B and non-A. Nothing else follows from this statement with respect to the distribution of A and B. So suppose that there are a units which are both A and B, b units which are A and non-B and c units which are non-A and non-B (where a, b, c are natural numbers). In this general population, $p(B) = a / (a + b + c)$ and $p(B \&zwjn; A) = a / (a + b)$. Since for any a, b, c : $a / (a + b + c) < a / (a + b)$, it follows that $p(B \&zwjn; A) > p(B)$. (Note that if A were not a true condition with respect to this population and hence $c = 0$ it would follow that $p(B \&zwjn; A) = p(B)$.)

[18]

If we have doubts about this explanation it is not because the statistical generalization it involves does not meet the high-probability criterion but rather because we suspect that this might be an instance of "collateral causation", in which two seemingly causally related phenomena are in reality effects of a common cause (Little 1991: 21). Accordingly, Salmon (1984: 22) emphasizes that a complete explanation of a particular fact cannot rely on some set of statistical relevance relations alone, but must include an explanation of these relations "in terms of *causal* relations".

[19]

Popper's critics claim that he virtually makes no difference between an untested and a highly corroborated hypothesis (King/Keohane/Verba 1994: 100-103, Van Evera 1997: 43). But this does not well accord with the fact that Popper (1968: para. 82) makes the effort to develop a "positive theory of corroboration". Without this distinction theoretical knowledge could hardly aid in explaining and predicting events, and using theories for purposes of (social) engineering would be nothing short of irresponsible.

[20]

The falsificationist strategy as such can be understood as obeying to a principle of efficiency. By directing the researcher to focus on potential falsifiers, it calls on him or her not to waste his or her time on experiments and observations that cannot decide the fate of the hypothesis.

[21]

For convenience, we use the word "falsifier" both to refer to the singular statements which are inconsistent with the hypothesis under consideration and to characterize the objects (here: regimes) these statements describe.

[22]

Obviously, this case selection strategy does not conform to the methodological rule that cases must not be chosen on the basis of a particular value of the dependent variable (King/Keohane/Verba 1994: 129-132). This rule, however, does not apply when a deterministic conditional hypothesis such as the justice hypothesis is tested, where one looks for falsifying observations rather than associations between variables (see also Collier 1995: 464).

[23]

At the saturation point of a sequence of tests, further tests will add little to the degree of corroboration (which is mainly a function of the hardness and less so of the number of tests) because the hypothesis has survived virtually every conceivable hard test (cf. Popper 1968: para. 82). At this point the hypothesis may be provisionally accepted as true.

[24]

If we resort to this "contextualization strategy" (Efinger/Mayer/Schwarzer 1993, Hasenclever/Mayer/Rittberger 2000), we should not only continue to try to falsify the - now contingent - hypothesis. We should also think about why and how, in high-politics regimes (and only there), the causal mechanism that mediates between low justice and low robustness may sometimes be prevented from operating. For example, one might argue that only when the stakes are very high, powerful actors are prepared to use their might to impose and to enforce agreements that maximize their gains at the expense of the legitimate interests of weaker actors.

[25]

We use the terms "robust" and "resilient" interchangeably.

[26]

It might be argued that we attach too much importance to normative continuity, thus playing robustness off against flexibility when in fact a measure of flexibility may well be a precondition for robustness. However, in our operationalization of robustness the possibility of learning and adaptation is taken into account by allowing changes *within* the regime to take place without implying a lack of robustness. Note also that the situations we describe as exogenous shocks create incentives for individual actors to *reduce* their commitment to the regime. That is, we are *not* referring to the kind of exogenous shocks that have been discussed in the literature on international environmental cooperation. Shocks of the latter kind (for example, the discovery of the ozone hole) *enhance* the prospects of cooperation by bringing home to the actors the need to "act" (Young/Osherenko 1993: 234f). While this kind of exogenous shock may well prompt states to consider changes of the present regime as a means of improving (rather than reducing) their cooperation, this is much less likely in the case of the first kind of shock which is more prone to induce regressive changes of the regime (resulting in lower levels of cooperation) or mere adaptations to a new distribution of power among regime members.

[27]

The latter stipulation is particularly appropriate in the case that the justice hypothesis is reinterpreted as a general factor claim according to which (imposed or unnoticed) injustice undermines cooperation. If this hypothesis is true, many unjust regimes may not last long enough to experience a serious exogenous challenge. As a consequence, an evaluation of the factor claim would be biased against the hypothesis if these regimes were excluded from the universe of cases.

[28]

Maximin is a risk-averse strategy that calls on actors to evaluate their behavioral options in terms the worst possible results they may have (no matter how likely these results are) and to pick the option the worst possible consequence of which is least undesirable. A Nash equilibrium in a 2x2 game is a pair of strategies each of which is an optimal response to the other. In equilibrium no actor desires unilateral change. A Pareto optimum is an outcome which is efficient in the sense that the payoff of neither of the actors can be improved without the other being made worse off. Some but not all Pareto optima may be taken as representing cooperative outcomes of the game. Zürn (1992: 153f) proposed to call those Pareto efficient outcomes that qualify as cooperative in an intuitive sense "qualified Pareto optima". A Pareto optimum is "qualified" if it is at least as desirable to all players as the non-cooperative outcome that is likely to emerge from independent decisionmaking (individual utility maximizing behavior).

[[29](#)]

According to institutionalists, many if not all regimes assume a "life of their own" providing them with a measure of independence from the factors that help explain their creation. This proposition is not inconsistent with our definition of shocks, though. After all, robustness is mainly about having acquired such a degree of independence.

[[30](#)]

Precluding a statistical, large-n test, this fact does have important methodological implications for the correlational variant of the justice hypothesis ("The more nearly just a regime, the more it is robust"). Attempts to evaluate this hypothesis empirically, therefore, have to rely on the comparative method (Hasenclever/Mayer/Rittberger 1996, 1998).

[[31](#)]

Alternatively, a more theoretically-informed searching strategy might be applied. The strategy consists in inspecting regimes that are collective responses to situations resembling the Battle of the Sexes game. For this class of regimes should include regimes which are at least *prima facie* counter-examples to the justice hypothesis. This is because these regimes 1) are coordination regimes and, consequently, should tend to be very robust and 2) will often institutionalize an unequal distribution of gains (Zürn 1992: 191f). As the pay-off matrix in the text indicates, coordination regimes forming in Battle situations normally select one of several possible equilibria which are unequally satisfying to the actors: while one gets his or her first best outcome, the other must be content with his or her second preference. To be sure, neither are preferences identical with gains, nor is an unequal distribution of gains necessarily unjust. Still, it is reasonable to assume that, on closer inspection, regimes of this type may prove to constitute serious anomalies for the justice hypothesis.

[[32](#)]

Distributive justice is only one of a family of moral ideas which come under the concept of justice. Others include commutative (exchange) justice, procedural justice and justice as an individual virtue (e.g. Albin 1993, Bull 1977: 75-82). Throughout this article, when we talk of justice without further qualification, we mean distributive justice.

[[33](#)]

Strictly speaking, this is only true of moral philosophers who, in Walzer's (1987) terminology, proceed by way of either "discovering" or "inventing" moral truths. The third group of philosophers, whom Walzer calls "interpreters" and sides with himself, are internalists who regard ethics as an intellectual activity that essentially takes place within a moral tradition.

[[34](#)]

According to Albin (1993: 229-234), justice principles as focal points are characteristic of distributive negotiations, whereas negotiated justice is the hallmark of integrative bargaining. Note that even in the first case the standard of justice is not external. Only those distributive principles can coordinate expectations that are accepted as fair by the actors themselves.

[[35](#)]

Most of the formulas Zartman and his collaborators mention seem to be compatible with the idea of reciprocity. These formulas either correspond to the parity principle or represent a form of compound justice where different proportionality principles are traded so as to achieve a kind of second-order parity (Zartman *et*

[[36](#)]

Decisionmaking procedures are irrelevant here, as their immediate implications are for procedural rather than distributive justice.

[[37](#)]

A similar approach to international justice is taken by Welch (1993) except that he does not distinguish between general and specific provisions in regimes. This is because his interest is different from ours: he is looking for - internal - criteria for just and unjust behavior or legitimate and illegitimate entitlement claims rather than for just and unjust orders or regimes, and it is in regimes that he finds these criteria.

[[38](#)]

These are the five states that possessed nuclear weapons when the NPT was concluded: the U.S., the Soviet Union/Russia, China, Great Britain and France. Recently, India and Pakistan have unsuccessfully attempted to gain recognition as (official) NWS. This would necessitate a modification of the NPT because of Article IX.3, which defines NWS as those states which have manufactured and exploded a nuclear device before 1 January 1967.

[[39](#)]

Several of the abstract "types of exogenous shock" we have distinguished [in section 3.1.](#) seem to apply in the cases mentioned in the text. The most obvious one, of course, is the "emergence of new relevant actors".

[[40](#)]

Discrimination is not unjust because unequal treatment is unjust *per se*. Proportionality principles such as the need principle ("To each according to his or her needs") require differential treatment. But this distinction must be justified in terms of some value or goal that is accepted by the community in question. It is not part of the underlying bargain of the regime, however, that power is such a value or that the preservation of the international society as a class society is one of the shared goals.