doi:10.1017/ajil.2021.2

# Rewarding in International Law

*By Anne van Aaken*[*] *and Betül Simsek*[**]

## Abstract

*Why states comply with international law has long been at the forefront of international law and international relations scholarship. The compliance discussion has largely focused on negative incentives. We argue that there is another, undertheorized mechanism: rewarding. We provide a typology and illustrations of how rewards can be applied. Furthermore, we explore the rationale, potential, and limitations of rewarding, drawing on rationalist and psychological approaches. Both approaches provide ample justifications for making greater use of rewarding in international law.*

## I. Introduction

Within a state, compliance with legal rules can be secured through the courts and the police power. In the international arena, there is no analogous centralized enforcement system. Lacking an "international sheriff" to force states to obey international law, the international system relies on the traditional concept of self-help and countermeasures. The weakness of this decentralized system persists despite the proliferation of international courts and tribunals as well as the legally binding resolutions of the UN Security Council. Both international legal and international relations (IR) scholars have researched the many ways in which international law is given effect, including at the national level and specifically in national courts.[1]

* Alexander von Humboldt Professor for Law and Economics, Legal Theory, Public International Law and European Law. Director, Institute of Law and Economics, University of Hamburg. Corresponding author: anne.van.aaken@uni-hamburg.de.

** Research Associate, PhD Cand. Econ., Institute of Law and Economics, University of Hamburg. We would like to thank Tomer Broude, James W. Davis, Jeffrey L. Dunoff, Léa Marchal, Jerg Gutmann, Greg Shaffer, Ed Swaine, Joel P. Trachtman, Ingrid Wuerth, and five anonymous reviewers for immensely helpful comments. All remaining errors are entirely ours.

[1] The most popular accounts can be divided into the rational actor model and the normative/constructivist models, see, e.g., Ingrid Wuerth, *Compliance*, *in* Concepts for International Law: Contributions to Disciplinary Thought 117 (Jean d'Aspremont & Sahib Singh eds., 2019) (and sources cited therein). Under the rational actor model compliance emerges through self-interest, see, e.g., Andrew T. Guzman, *A Compliance-Based Theory of International Law*, 90 Cal. L. Rev. 1823 (2002); Andrew T. Guzman, How International Law Works: A Rational Choice Theory (2008) [hereinafter How International Law Works]; Tom Ginsburg & Richard H. McAdams, *Adjudicating in Anarchy: An Expressive Theory of International Dispute Resolution*, 45 Wm. & Mary L. Rev. 1229 (2004). For a rationalist treatment in the field of IR, see, e.g., George W. Downs, David M. Rocke & Peter N. Barsoom, *Is the Good News About Compliance Good News About Cooperation?*, 50 Int'l Org. 379 (1996). The normative approach assumes that states have a preference for cooperation; including the view that noncompliance is a problem of norm management, see, e.g., Abram Chayes & Antonia Handler Chayes, *On Compliance*, 47 Int'l Org. 175, 179 (1993); Abram Chayes &

Holding states to legal obligations poses conceptual as well as practical problems, making the questions of enforcement and compliance central to the study of international law.[2] Compliance theory provides an understanding of why states fulfill their international obligations, focusing on the design and operation of possible enforcement mechanisms.[3] International legal and IR scholarship have predominantly focused on negative incentives to comply.[4] Although the costliness of penalties and their ineffectiveness have been thoroughly discussed, they remain at the forefront of academic and policy discussions. This focus overlooks a potentially effective mechanism—rewarding—to induce compliance with international law.

Rewarding is indispensably linked to compliance theory but is undertheorized. Because of the horizontal nature of international law with its weak enforcement, all means of fostering compliance should be considered.[5] Rewards, as understood here, are improvements in a target's value position relative to a baseline of expectations.[6] Rewards are transfers of positively valued material or immaterial goods, such as opportunities for and benefits of cooperation, money, technology, or social approval/good reputation. Penalties are deprivations relative to the same baseline.[7] Once a penalty has become part of the baseline, then the removal of the penalty would have the effect of a reward. The same holds conversely: the removal of a reward can be a penalty (again assuming that the reward has become sufficiently foreseeable to be part of the baseline).

We discuss four types of rewards, based on two distinctions. The first distinction is between internal and external rewards. Internal rewards refer to the benefit of the treaty in question (base treaty) and are the benefits of cooperation and participation provided by that very treaty.[8] External rewards refer to a benefit *outside* the bargain of the base treaty. The other distinction is based on the time when the benefit accrues: entry rewards benefit the country at

ANTONIA HANDLER CHAYES, THE NEW SOVEREIGNTY (1995); Harold Hongju Koh, *Why Do Nations Obey International Law?*, 106 YALE L.J. 2599 (1997). For the importance of fairness and legitimacy, see THOMAS M. FRANCK, THE POWER OF LEGITIMACY AMONG NATIONS (1990).

[2] Compliance theories are also theories about "the nature and operation" of international law more generally. *See* Benedict Kingsbury, *The Concept of Compliance as a Function of Competing Conceptions of International Law*, 19 MICH. J. INT'L L. 345, 346 (1998).

[3] Compliance generally can be defined as "the interaction between rules and behavior." *See* Benjamin van Rooij & D. Daniel Sokol: *Compliance as the Interaction Between Rules and Behavior. Introduction to Cambridge Handbook of Compliance*, *in* CAMBRIDGE HANDBOOK OF COMPLIANCE 4 (Benjamin van Rooij & D. Daniel Sokol eds., forthcoming 2021), *available at* https://ssrn.com/abstract=3563295. Usually, compliance is distinguished from effectiveness of the norm, the latter indicating causality, the former correlation. We are also mindful of the distinction between treaty compliance and regime effectiveness, the latter focusing on reaching the objectives of the respective regime. We denote in the text if we refer to a specific meaning; otherwise we use compliance as an overall term.

[4] One notable exception is: Omri Ben-Shahar & Anu Bradford, *Efficient Enforcement in International Law*, 12 CHI. J. INT'L L. 375 (2012). Further important contributions come from IR scholars, see, e.g., David A. Baldwin, *Thinking about Threats*, 15 J. CONFL. RESOL. 71 (1971); David A. Baldwin, *The Power of Positive Sanctions*, 24 WORLD POL. 19 (1971) [hereinafter Baldwin, *The Power of Positive Sanctions*]; Thomas W. Milburn & Daniel J. Christie, *Rewarding in International Politics*, 10 POL. PSYCH. 625 (1989).

[5] As also recently discussed in domestic law, see Brian D. Galle, *The Economic Case for Rewards Over Imprisonment*, IND. L.J. (forthcoming), *available at* https://ssrn.com/abstract=3575031.

[6] The concept of using a baseline to distinguish threats from promises is drawn from PETER M. BLAU, EXCHANGE AND POWER IN SOCIAL LIFE (1965). Other scholars have used this concept in their work. *See, e.g.*, Baldwin, *The Power of Positive Sanctions*, *supra* note 4, or JAMES W. DAVIS, THREATS AND PROMISES: THE PURSUIT OF INTERNATIONAL INFLUENCE (2000).

[7] We use the term penalty, but in the literature this is often also referred to punishment or sanctioning.

[8] The base treaty is the treaty to be complied with.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

the time of entering the commitment; compliance rewards benefit the country when it complies with the commitment.

The rationale for rewarding is that cooperation can be encouraged through rewards offered by one party to offset the benefits that the other country draws from noncompliance.[9] In other words, compliance can be achieved if a reward outweighs the benefits from breaching international law. This rationale is linked to the Coase theorem, which states that if transaction costs are sufficiently low, it does not matter to which party one initially assigns a right.[10] The other party may (and if rational should and would) pay the right holder to relinquish it, so long as both would be better off. Thus, anyone considering penalties should be interested in rewards, both as a practical and theoretical matter. From a rational choice perspective, with few exceptions, rewards are pareto efficient because they make one country better off and no country worse off, while penalties are not pareto efficient since they make the target country always worse off.

While the Coase theorem shows that we should pay attention to rewards, it incorrectly suggests a type of equivalence between rewards and penalties. However, psychological research, including in IR scholarship, shows distinct differences of individual and state perceptions and responses to rewards and penalties.[11] Perceived losses and gains provoke different behavior in many respects. The baseline matters because actors evaluate gains and losses from a reference point.[12] This Article thus goes beyond the Coasean insight by arguing that rewards can produce better results than penalties, on both efficiency grounds (pareto optimality) and psychological grounds (behavioral differences). It is commonly noted that states commit to and comply with treaties because of the benefits (rewards) they receive from doing so. The exchange of benefits in treaty negotiation is routine and widely documented. Indeed, a lot of compliance literature has focused on what we call internal rewards, namely the assumption that enjoying the benefits of cooperation is often a sufficient incentive to comply with a treaty.[13] The threat of withdrawal of benefits is conceptualized as outcasting.[14] Thus, some types of rewards are addressed in the compliance literature, but rewards have not been

[9] The target country is the actor whose compliance should be induced. Receiving country, rewarded country, as well as target country are used interchangeably. Similarly, we use the terms enforcing country, sender, and rewarding country interchangeably. That is the actor who is inducing the target country to comply. Rewards can also be used for nonstate actors, but we confine the analysis to states for space and simplicity reasons.

[10] Ronald H. Coase, *The Problem of Social Cost*, 3 J. L. & ECON. 1 (1960). The term "Coase theorem" was introduced in GEORGE J. STIGLER, THE THEORY OF PRICE (1966).

[11] We use behavioral economics and psychology synonymously here, although behavioral economics looks at revealed preferences through choice patterns that might contradict the axioms of rational choice theory (cognitive and motivational), whereas psychology looks at internal decision-making processes.

[12] People evaluate the utility of prospective outcomes against a reference point that is regarded as neutral. The framing of prospective outcomes in terms of a reference point establishes a psychological domain of gains (all outcomes above the reference point) and losses (those outcomes that fall below the reference point). *See, e.g.*, Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Risk*, 47 ECONOMETRICA 263 (1979); Amos Tversky & Daniel Kahneman, *The Framing of Decisions and the Psychology of Choice*, 211 SCI. 453 (1981); Martin Dufwenberg, Simon Gächter & Heike Hennig-Schmidt, *The Framing of Games and the Psychology of Play*, 73 GAMES & ECON. BEH. 459 (2011).

[13] *See, e.g.,* Oona A. Hathaway, *Between Power and Principle: An Integrated Theory of International Law*, 72 U. CHI. L. REV. 469, 479 (2005); Robert O. Keohane, *The Demand for International Regimes*, 36 INT'L ORG. 325, 331(1982) ("In general, we expect states to join those regimes in which they expect the benefits of membership to outweigh the costs.").

[14] Oona Hathaway & Scott J. Shapiro, *Outcasting: Enforcement in Domestic and International Law*, 121 YALE L.J. 252 (2011).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

explicitly analyzed and conceptualized within a typology that captures *all* positive inducements for cooperation. As an example, outcasting can be an effective penalty but is especially effective when coupled with the possibility for readmission (redemption), a reward.[15]

The psychological effects of rewards have been ignored in international law scholarship. However, because framing affects our thinking and decisions, it makes a difference how people (and states) look at international cooperation—positively in terms of rewards or negatively in terms of penalties.[16] Penalizing as a mechanism focuses on the ways that the *costs* of noncompliance can be increased, e.g., retaliation or outcasting. Rewarding as a mechanism concentrates on the ways that the *benefits* of compliance can be increased, e.g., through the transfer of valuable goods or naming and praising. For example, framing concessions in treaty negotiations as part of a benefit instead of a loss, namely the cooperative gain of the treaty, may well make a difference in the success of negotiations.[17] Also, changing the narrative from violation to compliance may induce more cooperation.

This Article makes two main contributions. First, our typology captures *all* positive incentives for states to cooperate, be it by joining a treaty or by complying with a treaty, and allows for the use of those different kinds of positive incentives for strategic options in treaty design or in policy. The precise way in which the different kinds of rewards are distinguished is novel. We thereby provide a "toolbox" for enhanced cooperation, and we discuss the conditions under which different tools may work. Second, we add behavioral insights to the rationalist analysis—insights that underscore the usefulness of rewarding. Enforcing international law through positive inducements alters the frames within which we think about international cooperation. Looking at these compliance mechanisms helps not only to better understand the mechanisms as already discussed in the literature but also shifts the focus from a penalty-oriented system to governance mechanisms between states. We argue that reframing the debate on compliance by considering rewarding may enhance cooperation overall. To be sure, rewarding is no panacea and may work best if combined with penalties, but it adds a dimension to compliance mechanisms in international law that has largely been overlooked and that can also be used in treaty drafting.[18]

Historically, the literature on rational design (why and how law is made; international law as *explanandum*) is different from that on compliance (the effects of law on behavior of states; international law as *explanans*).[19] These literatures also refer to different points in time: first, the time of treaty negotiation and making, reflecting the interest of states in the treaty, and, second, a given point of time when the interest in compliance with the treaty may not be

[15] Real exclusion from a treaty in the sense of expulsion is not common. Rather, outcasting within international law mainly means that the violating state is suspended from benefits or voting rights.

[16] For international law, see Anne van Aaken & Jan-Philip Elm: *Framing in and Through International Law*, *in* International Law's Invisible Frames – Social Cognition and Knowledge Production in International Legal Processes (Andrea Bianchi & Moshe Hirsch eds., forthcoming 2021), *available at* https://ssrn.com/abstract=3678551.

[17] Seminal on that topic: Kenneth J. Arrow, Robert H. Mnookin, Lee Ross, Amos Tversky & Robert B. Wilson, Barriers to Conflict Resolution (1995); and Carsten K.W. de Dreu, Peter J. D. Carnevale, Ben J. M. Emans & Evert van de Vliert, *Effects of Gain-Loss Frames in Negotiation: Loss Aversion, Mismatching, and Frame Adoption*, 60 Org. Beh. Hum. Dec. Processes 90 (1994).

[18] We deal extensively with the limitations of rewarding in Section VI.1 *infra*.

[19] Barbara Koremenos, Charles Lipson & Duncan Snidal, *The Rational Design of International Institutions*, 55 Int'l Org. 761 (2001); Kenneth W. Abbott & Duncan Snidal, *Hard and Soft Law in International Governance*, 54 Int'l Org. 421 (2000).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

present. Yet, this distinction has never been watertight.[20] Compliance mechanisms are closely related to treaty design since, when designing treaties, states must already have some ideas about how compliance with the commitments can be enhanced (backward induction). Although those questions should be conceptually distinguished (law is either on the left or right side of the equation), they are not independent of each other if the whole life cycle of international law is to be analyzed.[21] This Article is a contribution to both the treaty design and compliance literatures, but with a focus on the latter (or on the subset of rational design literature dealing with compliance), which in turn feeds into treaty design and thus the possibilities of cooperation.[22]

We proceed as follows. Part II surveys compliance mechanisms as commonly discussed in the literature. We then introduce the definition and typology of rewarding, connecting, and differentiating it from the literature discussed. Part III then illustrates our typology of rewarding with examples. In Part IV, we discuss the differences between rewarding and penalties on a rationalist basis. We then turn in Part V to a behavioral analysis of rewarding. Part VI discusses the limits of rewarding and the conditions under which it can be expected to work, including its combination with penalties. Part VII concludes.

## II. Compliance Mechanisms in the Literature

Compliance theories in international law rely on a variety of mechanisms, including norm spirals,[23] focal points, expressive law theories,[24] and international courts.[25] They either assume a unitary state or break up the black box of the state to explain compliance—for example, through national political processes[26] or national courts.[27] We acknowledge the merits of these theories, but our focus is on rationalist theories that predominantly use a unitary actor model, simply to make our point clearer.

The advantage of rational choice theory is that it provides falsifiable explanations about when countries will choose to violate international law.[28] This theory reduces the complexity

[20] Barbara Koremenos, The Continent of International Law (2016) (although situating herself in the design literature (at 2), she includes a chapter on penalty provisions where she also briefly discusses rewards (chapter 8)). Chayes & Chayes, *supra* note 1, at 183 ("[I]f the agreement is well-designed . . . compliance problems and enforcement issues are likely to be manageable.").

[21] For details, see Anne van Aaken & Joel P. Trachtman: *Political Economy of International Law: Towards a Holistic Model of State Behavior*, *in* Political Economy of International Law: A European Perspective 9 (Alberta Fabricotti ed., 2016).

[22] Kal Raustiala & Anne-Marie Slaughter: *International Law, International Relations and Compliance*, *in* Handbook of International Relations 538 (Walter Carlsnaes, Thomas Risse & Beth A. Simmons eds., 2001); Kal Raustiala, *Compliance & Effectiveness in International Regulatory Cooperation*, 32 Case W. Res. J. Int'l L. 387 (2000).

[23] Martha Finnemore & Kathryn Sikkink, *International Norm Dynamics and Political Change*, 52 Int'l Org. 887 (1998).

[24] Alex Geisinger & Michael Ashley Stein, *A Theory of Expressive International Law*, 60 Vand. L. Rev. 77 (2007); Ginsburg & McAdams, *supra* note 1.

[25] Karen J. Alter, *Do International Courts Enhance Compliance with International Law?*, 25 Rev. Asian & Pac. Stud. 51 (2003).

[26] Harold Hongju Koh, *Why Do Nations Obey International Law?*, 106 Yale L.J. 2599 (1997).

[27] Anthea Roberts, *Comparative International Law? The Role of National Courts in Creating and Enforcing International Law*, 60 Int'l & Comp. L. Q. 57 (2011).

[28] Guzman, *supra* note 1, at 1860.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

of the real world into clear parameters: states are assumed to be rational, self-interested, and utility-maximizing.[29]

In principle, all rational choice approaches to international law are based on the comparison of (objectively viewed) benefits and costs. From that perspective, states enter a treaty when the benefits arising from the treaty are higher than the costs of entering.[30] Likewise, states later comply with the treaty when the benefits of compliance are higher than the costs of compliance (where the costs of compliance also include the forgone benefits from noncompliance). In the literature, several mechanisms are discussed.

Most importantly, reciprocity is viewed as a basic mechanism for compliance. It exists in two forms: reciprocity as a practice of exchanging things with others for mutual benefit and reciprocity as the benefit from the act of compliance.[31] Reciprocity is deemed to be a crucial building block of human societies[32] as well as of international law[33] and is also reflected in Article 60(1–3) of the Vienna Convention on the Law of Treaties (VCLT).[34] It can be positive or negative. Positive reciprocity defines the benefit from the practice of exchanging things (e.g., rights, gains, and privileges), while negative reciprocity is defined by the withdrawal of beneficial exchanges. Reciprocal benefits are usually understood to be benefits from the treaty obtained through the compliance of the other party (or parties).

Often, states comply with treaties when they fear that their own noncompliance can trigger reciprocal noncompliance by other states. This works best when mutual cooperation is preferred to mutual violation.[35] If mutual cooperation is preferred, reciprocity has the capacity to induce compliance (as defection might otherwise end future cooperation).

A violation by one side is likely to provoke reciprocal withdrawal by the other side, at least in bilateral arrangements. But negative reciprocity, and the threat thereof, may be inefficient in treaties with public good aspects, since it can provoke a complete breakdown of cooperation, e.g., in arms control treaties.[36] Negative reciprocity is also unlawful under VCLT Article 60(5) for provisions relating to the protection of the human person contained in treaties of a

[29] We recognize that there are various forms of rationality, and that these have different implications for law. *See, e.g.*, JON ELSTER, ULYSSES AND THE SIRENS: STUDIES IN RATIONALITY AND IRRATIONALITY (1979). If states were perfectly rational and treaties were well-designed, then maybe there would be less need for compliance-enhancing mechanisms like penalties, as states would voluntarily comply to advance their own long-term interests. But since they are often imperfectly rational, (or captured by special interests), we need compliance-inducing tools.

[30] *E.g.*, ERIC POSNER & ALAN O. SYKES, ECONOMIC FOUNDATIONS OF INTERNATIONAL LAW 21 (2013).

[31] See also for a discussion of forms of reciprocity, Robert O. Keohane, *Reciprocity in International Relations*, 40 INT'L ORG. 1 (1986).

[32] Extensively, SAMUEL BOWLES & HERBERT GINTIS, A COOPERATIVE SPECIES: HUMAN RECIPROCITY AND ITS EVOLUTION (2011).

[33] *See* Bruno Simma: *Reciprocity*, *in* MAX PLANCK ENCYCLOPEDIA OF PUBLIC INTERNATIONAL LAW (Rüdiger Wolfrum ed., 2008), *at* https://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e1461#law-9780199231690-e1461-div1-4. From a rationalist perspective, see Keohane, *supra* note 13; Francesco Parisi & Nita Ghei, *The Role of Reciprocity in International Law*, 36 CORNELL INT'L L.J. 93 (2003). JUTTA BRUNNÉE & STEPHEN J. TOOPE, LEGITIMACY AND LEGALITY IN INTERNATIONAL LAW: AN INTERACTIONAL ACCOUNT 37–42 (2010) (using reciprocity in a noninstrumental Fullerian sense to explain obligation in international law).

[34] Vienna Convention on the Law of Treaties, Jan. 27, 1980, 1155 UNTS 331, 8 ILM 679.

[35] Alan O'Neil Sykes & Andrew Guzman: *Economics of International Law*, *in* THE OXFORD HANDBOOK OF LAW AND ECONOMICS: VOLUME 3: PUBLIC LAW AND LEGAL INSTITUTIONS 439, 442 (Francesco Parisi ed., 2017).

[36] In multilateral constellations, reciprocity may not work well or is undesirable, but outcasting may. *Cf.* GUZMAN, HOW INTERNATIONAL LAW WORKS, *supra* note 1, at 174–76; KOREMENOS, *supra* note 20, at 233. Those treaties may therefore contain other penalty provisions—or rewarding mechanisms.

humanitarian character. Notwithstanding the legal norm, violating human rights obligations by one state will not induce another country to treat its citizens in the same way.

Scholars have also highlighted the importance of reputation. States comply because they want to be able to make credible commitments in the future. By complying with promises, each country enhances its reputation as a state that honors its commitments and, therefore, its ability to reap cooperative benefits *in the future.* This allows a state to find more partners and extract more generous concessions.[37] Reputational benefits can ensue when entering a treaty (reputation for normative commitments) as well as when a state decides to comply with its international obligations (reputation for being a reliable partner). Reputation, positive or negative, can play alongside the other compliance mechanisms or can stand alone if the others do not work or are undesirable (e.g., in human rights treaties). Reputation as a compliance mechanism has been extensively discussed and reviewed, but many open questions remain.[38] It clearly depends on perception and may differ between audiences[39] as well as on the availability of information about behavior and its salience. Is the reputation attributable to the state or to the government?[40] Can reputation be compartmentalized, that is, does it matter whether a state (or government) has, for example, a good human rights record but a bad one in regard to investment?

A further compliance mechanism is retaliation. Retaliation represents a mostly costly action by one or more states with the intention to punish another state for violation of a commitment.[41] Retaliatory actions include retorsions and counter-measures, such as economic, diplomatic, or military sanctions outside the base treaty. Retaliation is rational when it influences the future action of the violating state, i.e., when retaliation is used to persuade the violator to comply in order to avoid further sanctions. However, retaliation is costly for both the noncomplying state and the retaliating state. This costliness encourages sanctions free-riding, which creates a sanctioning dilemma (a second order prisoners' dilemma).[42] Thus, retaliation generally works best in bilateral relations and in response to ongoing violations.

[37] Sykes & Guzman, *supra* note 35, at 445.

[38] GUZMAN, HOW INTERNATIONAL LAW WORKS, *supra* note 1, at 71–117; Rachel Brewster: *Reputation in International Relations and International Law Theory*, *in* INTERDISCIPLINARY PERSPECTIVES ON INTERNATIONAL LAW AND INTERNATIONAL RELATIONS: THE STATE OF THE ART 524 (Jeffrey L. Dunoff & Mark A. Pollack eds., 2012); George W. Downs & Michael A. Jones, *Reputation, Compliance, and International Law*, 31 J. LEGAL STUD. 95 (2002).

[39] There is also the question whether all observers draw the same reputational inference when observing a specific behavior. *See* JONATHAN MERCER, REPUTATION AND INTERNATIONAL POLITICS (1995); Daniel W. Drezner, *The Trouble with Carrots: Transaction Costs, Conflict Expectations, and Economic Inducements*, 9 SEC. STUD. 188 (1999). The most recent debate is between KEREN YARHI-MILO, KNOWING THE ADVERSARY LEADERS, INTELLIGENCE, AND ASSESSMENT OF INTENTIONS IN INTERNATIONAL RELATIONS (2014) and DANIELLE L. LUPTON, REPUTATION FOR RESOLVE: HOW LEADERS SIGNAL DETERMINATION IN INTERNATIONAL POLITICS (2020).

[40] Scholars suggest that "regime types" are important and that democracies, largely because of "audience costs," can make more credible commitments than authoritarian regimes. The *locus classicus* is James D. Fearon, *Domestic Political Audiences and the Escalation of International Disputes*, 88 AM. POL. SCI. REV. 577 (1994), but there are many security and international political economy spinoffs and some empirical critiques of the theoretical argument. *See, e.g.*, Marc Trachtenberg, *Audience Costs: An Historical Analysis*, 21 SEC. STUD. 3 (2012); Jonathan Mercer, *Audience Costs Are Toys*, 21 SEC. STUD. 398 (2012).

[41] GUZMAN, HOW INTERNATIONAL LAW WORKS, *supra* note 1, at 34 (defines retaliation as a costly action with the intention to punish). This is indeed mostly the case, but need not, e.g., if development aid is withdrawn. Guzman distinguishes it from reciprocity (at 47) because of its costliness.

[42] Alexander Thompson, *The Rational Enforcement of International Law: Solving the Sanctioners' Dilemma*, 1 INT'L THEORY 307, 311 (2009) ("The second problem comes in the multilateral context, where free-rider incentives make individual states even less likely to bear the burden of enforcement.").

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

Yet another compliance mechanism discussed is nonviolent outcasting, defined as the use of techniques to deny noncompliant states the benefits of social cooperation and membership or use of markets.[43] Outcasting penalizes by shutting the violating state or its economic operators out of the "club" or suspending them temporarily, depriving them of the benefits of cooperation, which has damaging consequences.[44] The use of exclusion as penalty for noncooperation or violation of international law converts public goods to excludable nonrivalrous goods in terms of consumption—that is, club goods.[45] Even more important, enforcement can also be carried out by nonstate actors, such as private banks in the Financial Action Task Force (FATF) mechanism. Outcasting is not a particular form of retaliation since it occurs solely *within* the treaty framework and is usually rather cheap for the states who outcast—it is rather a form of collective negative reciprocity, that is, complying members collectively withdraw their promises to the violating member.[46] Outcasting as a penalty is pervasive (e.g., in Article 4 of the Montreal Protocol banning the import of the substances listed in the Annexes from nonparties,[47] the Convention on International Trade in Endangered Species of Wild Fauna and Flora,[48] or the Basel Convention[49]) and soft law (e.g., FATF for money laundering and terrorism financing or the Kimberly Process of conflict diamonds). Many regional organizations like the African Union, the Organization of American States, the Council of Europe, and the EU have some sort of outcasting device for members breaking their rules or principles (e.g., by revoking voting rights).

The managerial approach to compliance by Chayes and Chayes, which may be closest to our approach, drew attention to how the provision of incentives, positive assistance, and constructive engagement work to bring about compliance with international law, arguing that this approach is often more effective than sanctions or costly coercive enforcement.[50] They also called attention to the problem of incapacity as one reason for noncompliance with a treaty—a point we take up below.[51] In this Article, we supplement valuable insights from the managerial school, most notably the problem of incapacity, with an overall analytical treatment and typology of positive inducements. It helps us to understand when compliance

[43] Hathaway & Shapiro, *supra* note 14.

[44] Anne van Aaken, *Effectuating Public International Law Through Market Mechanisms*, 165 J. INST. THEOR. ECON. 33 (2009); Anne van Aaken, *Trust, Verify, or Incentivize? Effectuating Public International Law Regulating Public Goods Through Market Mechanisms*, 104 ASIL PROC. 153 (2011).

[45] On the definition of club goods, see James M. Buchanan, *An Economic Theory of Clubs*, 32 ECONOMICA 1 (1965).

[46] Outcasting is costly for the enforcing countries only if the outcast country is so important that cooperation (and its benefits) diminishes substantially or breaks down completely.

[47] Montreal Protocol on Substances that Deplete the Ozone Layer, Sept. 16, 1987, S. Treaty Doc. No 100-10, 1522 UNTS 29. The Montreal Protocol is a protocol to the Vienna Convention for the Protection of the Ozone Layer, March 22, 1985, TIAS No. 11, 097, 1513 UNTS 324.

[48] Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), Arts. III–V, Mar. 3, 1973, 993 UNTS 243.

[49] Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and Their Disposal, Mar. 22, 1989, 1673 UNTS 126, *reprinted in* 28 ILM 657 (1989).

[50] *See* Chayes & Chayes, *On Compliance*, as well as CHAYES & CHAYES, THE NEW SOVEREIGNTY, both *supra* note 1.

[51] Previous scholarship has expressed some concerns with the managerial school. *See* GUZMAN, HOW INTERNATIONAL LAW WORKS, *supra* note 1, at 16 (arguing that it "does not offer any underlying theory or explanation of why states prefer to comply with international law. Nor does it help us to understand when this preference for compliance will trump other concerns and when it will not prevail.").

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

can be expected and how different forms of rewards can overcome incapacity issues. We also go beyond the managerial school by explaining the underlying mechanisms of rewarding through a behavioral approach.

The first three mechanisms (reciprocity, reputation, retaliation), in their different forms, are what have been historically utilized to "promote compliance with international legal rules."[52] We submit that even in their positive form (positive reciprocity and reputation), these mechanisms do not in fact capture the universe of compliance mechanisms, and the discussion fails to distinguish this universe analytically.

## III. Rewarding: The Phenomenon

We argue that rewarding is indispensably linked to compliance theory but is undertheorized. Reward and penalty (or carrot and stick) are often seen as the opposite sides of the same coin and thus rewards have not received special attention.

The focus of our analysis is on treaty law, although it can be applied to all sources of international law. We do not exclude soft law agreements since they may contain compliance mechanisms similar to treaty law.[53] Bilateral constellations as well as multilateral constellations are considered and we cover all stages of the life cycle of international law. After developing a typology of rewarding, we turn to illustrations based on that typology, contributing to a better understanding of compliance mechanisms within those examples.

### *A. Typology of Rewarding*

To distinguish rewards from penalties, one must establish the target's baseline of expectations at the time the sender's influence attempt begins.[54] The baseline depends on rational expectations about the future, including the likelihood of reward/penalty. Rewards are improvements in a target's value position relative to its baseline of expectations; penalties are deprivations relative to the same baseline. For instance, giving a hundred dollar bonus to a man who expected a bonus of two hundred dollars may take the form of a reward but it is not perceived as such; similarly, cutting the salary by a hundred dollars while the man expected a two hundred dollar fine may take the form of a penalty but it is not perceived as such.[55] A conditional commitment not to reward if the target fails to comply is not necessarily a threat if the target had no prior expectation of receiving the reward.[56] Thus, withdrawing a

[52] Sykes & Guzman, *supra* note 35, at 444.

[53] We are following the traditional legal definition of hard and soft law. Daniel Thürer: *Soft Law*, *in* Encyclopedia of Public International Law (Rüdiger Wolfrum ed., 2009), *at* https://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e1469?prd=EPIL. See also the soft law literature on compliance: Andrew Guzman & Timothy Meyer: *Soft Law*, *in* Economic Analysis of International Law 123 (Eugene Kontorovich & Francesco Parisi eds., 2016); Tomer Broude & Yahli Shereshevsky, *Explaining the Practical Purchase of Soft Law: Competing and Complementary Behavioral Hypotheses*, *in* International Law as Behavior 98 (Harlan Grant Cohen & Timothy Meyer eds., 2021). For an empirical paper on soft law from an economic perspective, see Stefan Voigt: *The Economics of Informal International Law – An Empirical Assessment*, *in* Informal International Law-Making 81 (Joost Pauwelyn, Ramses Wessel & Jan Wouters eds., 2012).

[54] *See* references in note 6 *supra*. For a detailed discussion about the concept of baseline and possible pitfalls when distinguishing penalties from rewards, see Baldwin, *The Power of Positive Sanctions*, *supra* note 4, at 23–27, Section II: The Concept of Positive Sanctions.

[55] Baldwin, *The Power of Positive Sanctions*, *supra* note 4, at 23.

[56] *Id.* at 26.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

reward is only considered a punishment when the reward was expected. The baseline of expectations can shift over time. Once a penalty has become part of the baseline, then the removal of the penalty would have the effect of a reward. The converse situation also applies to rewards: the removal of a reward can be a penalty. To summarize, the reward/penalty distinction depends on the baseline, which in turn depends on expectations about the future, including the likelihood of reward/penalty. This means that defining the baseline and the framing of expectations is critically important for the effectiveness of either mechanism.[57]

The most important distinction to be drawn is between the benefits of the agreed upon bargain of a treaty (e.g., the exchange of goods or the provision of public goods) and rewarding outside the (base) treaty to be complied with (e.g., the payment of money that was not part of the treaty). In other words, the distinction is made between rewards as cooperative benefits accruing to a party of a treaty (internal rewarding) and rewards external to or "on top" of the cooperative benefit (external rewarding). We acknowledge that this distinction is made for analytical purposes, and there are examples, e.g., the Montreal Protocol discussed below, which show that the distinction depends on how the treaty is drafted and how it has been handled in practice.[58] But the distinction is crucial from a legal perspective, even if less so from a social science perspective. Another important distinction between internal and external rewards is their predictability and flexibility. Given that internal rewards are within the treaty, they are commonly more predictable but less flexible, whereas with external rewards it is the opposite.

The second distinction concerns the point in time and is made between rewards-for-entering and rewards-for-complying with a treaty. This distinction is necessary since a state's incentives at the treaty-negotiating stage may be different from those it faces when the time for compliance arrives. The entry reward is often the cooperative benefit of the respective treaty (internal reward, positive reciprocity) but can go further than that—there may be rewards given on top of the cooperative benefit, such as external rewards (e.g., promises of tariff concessions if International Labour Organization Conventions are ratified) and positive reputation. Any concessions given during treaty negotiations can be considered as internal rewards if they are included in the treaty text. If the bargain of the treaty is insufficient or inclusion of the reward in the treaty would be inappropriate, states may resort to external rewards at the negotiation stage.[59]

We can thus define a typology of rewards and penalties, which can be tangible or intangible. First, external penalties, e.g. retaliation, are a form of sanctions outside the base treaty. Similarly, fines can be considered as external penalties, as they have to be paid by the violating state "on top" of the treaty (the state is not subject to outcasting or to withdrawals of benefits of the base treaty).[60] Terminating a linked treaty is a form of external penalty. Reputation is

[57] Note that in a multilateral context, the baseline can also be influenced by how other states in similar situations were or will be treated.

[58] *See* note 47 *supra*, et seq. and corresponding text.

[59] Both types of rewards, internal and external, contribute to expanding the zone of potential agreement (ZOPA) in treaty negotiations and relate to the cost-benefit calculus in treaty making.

[60] A prominent example stems from Article 260 of the Treaty on the Functioning of the European Union (TFEU), June 7, 2016, OJ C 202; ex Article 228 TEC). It is a special judicial procedure for the enforcement of judgments that provides for the imposition of penalty payments or lump sums by the Court of Justice of the European Union (CJEU) on a member state that fails to comply with an earlier judgment of the CJEU

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

widely considered relevant for the future and not the base treaty, and thus we qualify a bad reputation as an external penalty.

Second, there are internal penalties, in the form of withholding cooperative or other benefits within the treaty limited to treaty parties nonperforming their obligations or to states not entering the treaty in the first place; this is negative reciprocity as well as outcasting. Internal penalty is the withdrawal and nonenjoyment of the benefit of cooperation of the base treaty by the violator. The formation of club goods and outcasting have a rationalist basis as explanation and experiments confirm their effectiveness. Experimentally, it has been shown that excluding defectors is a cheap and powerful sanctioning device.[61]

Third, there are internal rewards, allowing states to gain (when acceding a treaty) or retain the cooperative benefits of the (base) treaty when complying. Reciprocity is the practice of exchanging things with others for mutual benefit. In other words, (positive) reciprocity is the practice of rewarding cooperation through the exchange of rights, gains, privileges, and assistance within a treaty. Here, we also situate readmission or redemption after outcasting. Where exclusion is reversible, i.e., inclusion after an individual has been excluded ("redemption" in experimental terms), it is possible to achieve even larger contributions to the public good, possibly due to the endowment effect.[62]

Fourth, there are external rewards that are reputational as well as direct benefits (e.g., side payments, other advantages through linked treaties, or intangible rewards like state visits) that follow the entry or compliance with a treaty and are not captured by the cooperative benefit of the base treaty itself. A good reputation is an external reward in our framework, since it is a benefit outside the bargain of the base treaty. To be precise, while naming and praising is the rewarding act implemented by the sender, a good reputation is the result of this act for the receiver. We simplify reputation as rewarding here since the described differentiation does not matter for our argument. Reputation can accrue from joining a treaty as well as from complying with international legal obligations. The former is a reputation for normative commitment whereas the latter relates to being a trustworthy partner.

External *rewards* can become necessary when the benefit of the bargain or treaty (i.e., the internal reward) is insufficient to induce accession or compliance (for some states and/or at a certain point in time).[63] So, the question arises, if ensuring compliance with treaties needs external rewards—is the internal bargain sufficient? If not, is it desirable that those treaties

that a member state is in breach of its obligations under EU law. The EU Commission clarified the rules in a Memo in 2005, available at https://ec.europa.eu/commission/presscorner/detail/en/MEMO_05_482.

[61] Matthias Cinyabuguma, Talbot Page & Louis Putterman, *Cooperation Under the Threat of Expulsion in a Public Goods Experiment*, 89 J. Pub. Econ. 1421,1421 (2005) (They show "that contributions rose to nearly 100% of endowments with significantly higher efficiency compared with a no-expulsion baseline."). On the efficacy of this device, see Gary Charness & Chun-Lei Yang, *Endogenous Group Formation and Public Goods Provision: Exclusion, Exit, Mergers, and Redemption* (Working Paper, 2008), *available at* https://ssrn.com/abstract=932251. For a survey on different variations in experiments for enhancing cooperation in public good games, see Ananish Chaudhuri, *Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature*, 14 Exp. Econ. 47 (2011).

[62] *Id.* This makes regaining a privilege different from initially receiving. *See also* Alice Solda & Marie Claire Villeval, *Exclusion and Reintegration in a Social Dilemma*, 58 Econ. Inquiry 120 (2020).

[63] This is similar to an efficient breach situation. *See* Richard Morrison, *Efficient Breach of International Agreements*, 23 Denv. J. Int'l L. & Pol'y 183 (1994); Eric A. Posner & Alan O. Sykes, *Efficient Breach of International Law: Optimal Remedies, Legalized Noncompliance, and Related Issues*, 110 Mich. L. Rev. 243 (2011).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

TABLE 1.
TYPOLOGY OF PENALTIES AND REWARDS

| | Internal | External |
|---|---|---|
| **Rewards** | Entry rewards<br>• access to finance and assistance within the objectives of the treaty<br>• cooperative gains | Entry rewards<br>• linkage between treaties or promise of benefits outside bargain (e.g. development aid or trade treaty)<br>• positive reputation for entering a treaty generating future gains of cooperation |
| | Compliance rewards<br>• redemption (prospect of inclusion after previous exclusion)<br>• use of mechanisms within the treaty (e.g., Kyoto Protocol: use of flexibility mechanisms) | Compliance rewards<br>• access to finance and assistance promoting objectives outside the treaty<br>• positive reputation for compliance with the treaty generating future gains of cooperation |
| **Penalties** | Outcasting and negative reciprocity: withholding cooperative or other benefits within the treaty limited to treaty parties nonperforming their obligations | Retaliation, fines, and negative reputation, termination/suspension of linked treaties |

are concluded in the first place? The answer is yes, if the treaty provides a global public good, like environmental treaties or has third party beneficiaries, like human rights treaties.

We illustrate these types of rewarding in the following sections, since we submit that all of those mechanisms are used in international law to change states' behavior. The classical compliance framework is too narrow to understand how states behave. Therefore, we try to assess what drives states to interact cooperatively, which is part of a broader governance discussion. Showing the array of possibilities also informs states about rational design and how to achieve regulatory goals at different stages (treaty making as well as compliance).

### *B. Illustrations of Rewarding*

Although international law already uses rewarding, formal rewarding in the compliance stage is rather uncommon.[64] Although many compliance mechanisms are designed within a treaty, states can also add external rewards (which are mostly intangible and reputational, but need not be, as will be shown below). Often, treaties contain all four kinds of rewards—internal and external rewards as well as entry rewards and compliance rewards. A combination of rewards is especially common in treaties providing global public goods, such as environmental or disarmament treaties. Understanding how rewards are used in different issue areas of international law helps to shed light on the different tools applied (or that could be applied) to effectuate international law. We provide examples to illustrate the different constellations of internal rewards, followed by examples for external rewards.

[64] KOREMENOS, *supra* note 20, at 230 (finds in her empirical study of 234 treaties that only 11% of treaties have rewards and only in the sub-issue area of disarmament whereas many more have penalty provisions (except in the areas of finance and monetary treaties) – 32%). No definition of rewards is provided.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

### 1. *Illustrations of Internal Rewarding*

Internal rewards are derived from membership. Benefits from participating in economic, political, and legal ties with one another can generate the necessary incentives to enter and comply with a commitment. Benefits may be conditioned upon compliance either with (monetary) contributions or other substantive norms of the treaty. For instance, member states of the World Health Organization (WHO) benefit from a vast array of international public health programs (internal reward at entry stage), but voting privileges and services to which a member is entitled may be lost if the member state does not comply with its mandatory budget contributions (internal penalty at the compliance stage).[65] Not fulfilling the mandatory contributions may also lead to a loss in reputation and exiting the treaty during a pandemic may spoil the reputation of a country even more.[66] This external penalty enhances the internal reward mechanism.

The prospect of redemption or readmission for previously excluded countries is the reverse of outcasting and another constellation of internal rewards at the compliance stage that is found in many treaties.[67] Thus, members can regain (voting) rights once they fulfill their obligations. Readmission has been experimentally shown to enhance cooperation even more.[68] Enjoying benefits creates the so-called endowment effect, a well-researched behavioral bias finding that people are more likely to retain an object they own than acquire that same object when they do not own it.[69] It is connected to Prospect Theory and loss aversion.[70] Thus, from a behavioral perspective, after entering a treaty and enjoying the benefits, treaty compliance might be improved if those benefits can be lost by outcasting.

In cases of incapacity,[71] a penalty (internal or external) is unable to deter a country from violating international law because of the violator's inability to comply.[72] Lack of the necessary financial, administrative, or technological resources is perceived as one important reason why states do not follow international law, especially multilateral environmental agreements (MEAs).[73] Penalties can only be effective if the receiving country is capable of changing its

[65] Constitution of the World Health Organization, Art. 7, July 22, 1946, 14 UNTS 185; Hathaway & Shapiro, *supra* note 14, at 305–06.

[66] Amy Maxmen, *What a US Exit from the WHO Means for COVID-19 and Global Health*, 582 NATURE 17 (2020).

[67] Text at notes 47–49 *supra*.

[68] Text at notes 61–62 *supra*.

[69] Daniel Kahneman, Jack L. Knetsch & Richard H. Thaler, *Experimental Tests of the Endowment Effect and the Coase Theorem*, 98 J. POL. ECON. 1325 (1990); Daniel Kahneman, Jack L. Knetsch & Richard H. Thaler, *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, 5 J. ECON. PERSPECT. 193 (1991); Carey K. Morewedge & Colleen E. Giblin, *Explanations of the Endowment Effect: An Integrative Review*, 19 TRENDS COGNITIVE SCI. 339 (2015).

[70] Prospect Theory is a psychology Theory that describes how people make decisions when presented with alternatives that involve risk, probability, and uncertainty. It holds that people make decisions based on perceived losses or gains and are thus reference point dependent. People are usually averse to the possibility of losing, such that they would rather avoid a loss rather than take a risk to make an equivalent gain. See note 12 *supra*.

[71] Chayes & Chayes, *supra* note 1, at 188.

[72] In cases of impracticability, there is a question as to who should bear the burden of the high costs of performance, given that the obligee contracted to carry out its obligations.

[73] RONALD B. MITCHELL, INTERNATIONAL POLITICS AND THE ENVIRONMENT 162 (2010) ("When developing countries fail to meet their environmental commitments, it often reflects a lack of adequate or appropriate resources (and the presence of more pressing concerns) rather than a calculation that non-compliance better fits their interests."). Capacity building is needed: Lothar Gündling, *Compliance Assistance in International*

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

behavior in the direction requested.[74] Negative reputation is not likely to play a part in the case of incapacity and negative reciprocity is undesirable in public good treaties. If an agreement has been acceded to but cannot be complied with because of incapacity, internal rewards help to overcome (material) constraints by providing the prospective violator with the necessary resources.[75] This is usually employed when the bargain of the treaty is a global public good and broad membership is desired even if noncompliance due to incapacity is likely when acceding to the treaty (e.g., the Paris Agreement on Climate Change[76]).

Recognizing how much noncompliance arises from incapacity, MEAs shifted from penalizing violations to facilitating compliance.[77] Potential donors have incentives to contribute to internal rewards, because if not, the target actor will continue engaging in the environmentally harmful behavior. States that view themselves as sufficiently harmed by another state's activity should be more willing to offer internal rewards large enough to convince the targeted actor to discontinue that behavior.

A very prominent example of an environmental treaty providing global public goods (or preventing public bads) using a mix of internal and external rewards at the compliance stage is the Montreal Protocol on Substances that Deplete the Ozone Layer.[78] Applying our typology shows the different ways in which rewards are used for that treaty. The Protocol aims to ban the global production and use of ozone-damaging chemicals including Chlorofluorocarbons (CFCs), Hydrochlorofluorocarbons (HCFCs), and halon as used in air conditioning and refrigeration systems, placing limits on the amount of ozone-depleting substances each member state may produce and consume. From a rationalist perspective, public goods are expected to be underprovided. Yet, it has been dubbed "one of the most successful and effective environmental treaties ever negotiated and implemented."[79] What were the causes of its success? One element that encouraged countries to ratify the Montreal Protocol was the trade provisions. These limited member states to trade only with other member states, thus creating a club good. Once the main producing countries signed up, it was only a matter of time before all countries had to sign up or risk not having access to increasingly limited supplies of CFCs and other ozone-depleting substances (ODS). The treaty thus achieved universal ratification. In return for agreeing to observe these limits of production, states parties receive access to trading privileges denied to nonparties; they are thus rewarded. Because the reward is part of the bargain, this is an example of a classical entry internal reward, the baseline being nonmembership.[80] How was compliance achieved? The above-mentioned mutual gains from trade within the club make compliance with the Protocol attractive to target states. Given that negative reciprocity is inefficient since the treaty provides a global public good, the treaty creates the possibility of withholding certain rights and benefits that a party receives from the

*Environmental Law: Capacity-Building Through Financial and Technology Transfer*, 56 HEIDELB. J. INT'L L. 796 (1996).

[74] Ben-Shahar & Bradford, *supra* note 4, at 383.

[75] *Id.* at 383–84.

[76] Paris Agreement on Climate Change, Dec. 12, 2015, UN Doc. FCCC/CP/2015/L.9/Rev.1.

[77] Chayes & Chayes, *supra* note 1.

[78] *See* note 47 *supra*.

[79] Ian Rae, *Saving the Ozone Layer: Why the Montreal Protocol Worked*, THE CONVERSATION (Sept. 9, 2012), *at* https://theconversation.com/saving-the-ozone-layer-why-the-montreal-protocol-worked-9249.

[80] Hathaway & Shapiro, *supra* note 14.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

treaty. It thus uses (partial) outcasting as an internal penalty and redemption as an internal reward as compliance mechanisms.

Yet, additional internal rewards are considered crucial for the success of this treaty: the Multilateral Fund (MF) of Article 10, which provides incremental funding for developing countries (so-called Article 5 countries) to help them meet their compliance targets, thus deals with incapacity as described above.[81] A crucial element for the success was the recognition that nonreporting countries[82] were to a great majority developing countries unable to comply without technical assistance.[83] By recognizing this, the "Montreal Protocol [became] the first treaty under which the parties undertake to provide significant financial assistance to defray the incremental costs of compliance for developing countries."[84] Internal rewards in the form of technical and financial assistance helped to overcome material constraints of compliance. Significantly, the treaty has also provided institutional support. This helped countries to build capacity within their governments to implement phase-out activities and establish regional networks, so they can share experiences and learn from each other. As of December 2019, the contributions received by the MF from developed countries, or non-Article 5 countries, totaled over US$ 4.07 billion. The MF has also received additional voluntary contributions amounting to US$ 25.5 million from a group of donor countries to finance fast-start activities for the implementation of the HCFC phase-down. To facilitate phase-out of Article 5 countries, the Executive Committee has approved 144 country programs, 144 HCFC phase-out management plans, and has funded the establishment and the operating costs of ozone offices in 145 Article 5 countries.[85]

The drafting of Article 5 also provides an interesting example of the (analytical) distinction between internal and external rewards. The Soviet Union became a party to the Protocol in 1988. In the mid-1990s, some former Soviet republics faced potential noncompliance and asked for a grace period to meet the Protocol's provisions. Some states received funding from the MF as Article 5 states. But Russia was not an Article 5 state, and thus was not eligible for this funding. Eventually, Russia received funding from the Global Environment Facility (GEF; a fund helping states to meet the objectives of MEAs), the U.S. and Danish governments, and the World Bank. The GEF requested that continued funding was subject to the Protocol processes for noncompliance and payment depended on favorable reports by the Protocol's implementation committee. Thus, the funds for Russia came from "outside" the treaty and from bodies with no formal role in the treaty system, yet treaty bodies continued to play a major role in reviewing progress and addressing any compliance issues that arose.[86] Legally, the funds given to Russia were an external reward, but they would have

[81] Article 5 of the Montreal Protocol entitles developing countries to assistance from developed countries under the Multilateral Fund for the Implementation of the Montreal Protocol. Assistance takes the form of grants or concessional loans. *See* Multilateral Fund Secretariat, *at* http://www.multilateralfund.org/aboutmlf/fundsecretariat/default.aspx.

[82] One requirement of the treaty was that member states had to report annual CFC consumption.

[83] Chayes & Chayes, *supra* note 1, at 194.

[84] *Id.*

[85] Multilateral Fund for the Implementation of the Montreal Protocol, *at* http://www.multilateralfund.org/default.aspx.

[86] Jacob Werksman, *Compliance and Transition: Russia's Non-compliance Tests the Ozone Regime*, 56 HEIDELB. J. INT'L L. 750 (1996).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

been internal rewards if Russia had been included as an Article 5 state in the first place—a historical contingency.

The noncompliance procedure[87] was designed as a nonpunitive and advisory procedure.[88] It prioritized helping countries back into compliance and is a "flexible means to ensure some degree of implementation without suggesting the automatic blameworthiness of all nonperformance."[89] To clarify the range of outcomes that parties may expect from the noncompliance procedure, the parties have adopted an "Indicative List of Measures that Might Be Taken by a Meeting of the Parties in Respect of Non-compliance with the Protocol."[90] These measures are a mix of, on the one hand, positive inducements such as appropriate assistance (including assistance for the collection and reporting of data), technical assistance, technology transfer, financial assistance, information transfer, and training. On the other hand, they include the suspension of benefits of the protocol, with or without time limits, including those concerned with industrial rationalization, production, consumption, trade, transfer of technology, financial mechanisms, and institutional arrangements.[91] Those can be suspended and reinstated (redemption). This mix of different rewards has been successful. It is telling that all 142 developing countries were able to meet the 100 percent phase-out mark for CFCs, halons, and other ODS in 2010.[92]

The Montreal Protocol reflects a tone shift in international law by using an encouragement-based approach that uses rewards. It has served as a model for other systems,[93] like, e.g. the United Nations Framework Convention on Climate Change (UNFCCC) and its protocols.[94] The Marrakesh Accords, adopting the compliance mechanism for the Kyoto Protocol, e.g., established a Compliance Committee, consisting of a Facilitative and an Enforcement Branch, thus accounting for the reasons of noncompliance.[95] The Facilitative Branch included advice, financial and technical assistance, and capacity building to achieve the objective of the base treaty (internal rewards at the compliance stage).[96]

[87] Non-compliance Procedure, Annex II: Non-compliance Procedure (1998) – Tenth Meeting of the Parties.

[88] For a broad discussion of the development and operation of the Non-Compliance Procedure, see Martti Koskenniemi, *Breach of Treaty or Non-compliance? Reflections on the Enforcement of the Montreal Protocol*, 3 Y.B. INT'L ENVTL. L. 123, 131 (1992) ("Many speakers expressed doubts about the character of the proposed measures as sanctions. The '"positive and conciliatory aspects'" of the Non-Compliance Procedure were stressed by developing countries who were afraid that they would be the first to become objects of such measures.").

[89] *Id.* at 147.

[90] Report of the Fourth Meeting of the Parties to the Montreal Protocol on Substances that Deplete the Ozone Layer, UN Doc. UNEP/OzL.Pro.4/15 (Nov. 25, 1992); Decision IV/5 and Annex V of the Report of MOP-4.

[91] For an example of how the procedure works under noncompliance, see Werksman, *supra* note 86.

[92] Rae, *supra* note 79.

[93] A list of bilateral and multilateral funds can be found here: https://unfccc.int/topics/climate-finance/resources/multilateral-and-bilateral-funding-sources.

[94] UN Framework Convention on Climate Change, May 9, 1992, 1771 UNTS 107.

[95] The reasons for the bifurcation of facilitation and enforcement are manifold, some have to do with the differences between parties and their respective commitments. So, by definition, non-Annex I parties (developing countries) could only come before the Facilitative Branch. Jutta Brunnée, *A Fine Balance: Facilitation and Enforcement in the Design of a Compliance Regime for the Kyoto Protocol*, 13 TUL. ENVTL. L.J. 223 (2000); Jutta Brunnée, *The Kyoto Protocol: A Testing Ground for Compliance Theories?*, 63 HEIDELB. J. INT'L L. 255 (2003).

[96] Ronald B. Mitchell, *Flexibility, Compliance and Norm Development in the Climate Regime*, *in* IMPLEMENTING THE CLIMATE REGIME: INTERNATIONAL COMPLIANCE 65, 73 et seq. (Olav Schram Stokke, Jon Hovi & Geir Ulfstein eds., 2005).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

We now turn to an illustration of reciprocity. Commonly, positive reciprocity is discussed in the entry phase (as the bargain of the treaty). Negative reciprocity is used as a means for compliance. If the state does not comply with the treaty, it loses the benefits of the treaty. But positive reciprocity can also be used in the compliance stage and this framing matters. Changing the narrative to positive reciprocity can be found within international humanitarian law (IHL) in order to induce more compliance; behavior of IHL relevant actors is thus "reframed." Because IHL is often violated, its significance is contested. This "credibility gap"[97] caused by the sole reliance on reports of violations triggers "the perception that . . . IHL is always violated and therefore useless."[98] This "negative and dismissive discourse renders violations banal and risks creating an environment where they may become more acceptable"[99] and based on the principle of negative reciprocity, inducing a violating spiral ("The other side does not respect it, so why should we?"[100]). Therefore, the International Committee of the Red Cross (ICRC) is currently undertaking the so-called "changing the narrative" project to reaffirm the relevance of IHL in contemporary armed conflicts by giving concrete examples of compliance with IHL.

Instead of only focusing on violations of IHL, the ICRC wants to change the narrative to good practices, and thus break the perceived violation dominance, turning this negative spiral of reciprocity into a positive one. Furthermore, by mentioning the compliance of the actors who adhere to IHL, they may enhance their reputation, thus producing additionally an external reward. One initiative is the recently launched "IHL in Action: Respect for the Law on the Battlefield" database.[101] It is a collection of case studies documenting compliance with IHL in modern warfare. Based on publicly available information, these cases have been assessed by academics as demonstrating positive application of IHL. They demonstrate the importance of complying with that body of law in order to minimize human suffering in armed conflicts. Recent psychological studies highlight that focusing on unfavorable outcomes is not an effective way to change attitudes while pointing out desired behaviors is more likely to generate change and set in motion a positive spiral of reciprocity.[102] The ICRC believes that a more positive focus on IHL reporting can engender further compliance with the law.

### 2. *Illustrations of External Rewarding*

External rewards are benefits outside the base treaty, i.e., "on top" of the treaty. They may be needed to induce entry/compliance if the cooperative gain of the treaty is insufficient or suffers from social dilemma problems, which may be especially the case when global public

[97] Marco Sassòli & Yvette Issar, *Challenges to International Humanitarian Law*, *in* 100 Years of Peace Through Law: Past and Future 181 (Andreas von Arnaud, Nele Matz-Lück & Kerstin Odendahl eds., 2015). The article proposes the positive narrative to end the vicious spiral.

[98] Juliane Garcia Ravel & Vincent Bernard, *Changing the Narrative on International Humanitarian Law,* Humanitarian L. & Pol'y (Nov. 24, 2017), *at* https://blogs.icrc.org/law-and-policy/2017/11/24/changing-the-narrative-on-international-humanitarian-law.

[99] *Id.*

[100] *Id.*

[101] IHL in Action: Respect for the Law on the Battlefield, *at* https://ihl-in-action.icrc.org.

[102] Morsella Ezequiel, John A. Bargh & Peter M. Gollwitzer, Oxford Handbook of Human Action (2009).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

goods or commons are at stake.[103] Like internal rewards, external rewards can be used to overcome missing contributions to public goods. This is especially important if (negative) reciprocity is undesired in the compliance stage to avoid reverting to the status quo ante.[104] Reciprocal nonperformance (internal penalty) is not only irrelevant, unlawful, or undesirable in the context of arms control commitments, but also in MEAs and in humanitarian or human rights treaties, treaties establishing rights in favor of third states, or setting up *jus cogens* obligations.[105] For arms control treaties, Bernauer and Ruloff explain:

> A provides aid, thereby compensating B for giving up weapons that B would otherwise acquire, even if this would violate an arms control agreement, because it would be in B's best interest to do so. If B rejects the offer by A the only welfare consequences for both actors are the transaction costs of making and negotiating the proposal. If the compensation offer is accepted, both actors improve their welfare, as compared to the status quo ante. Exchange is, therefore, a Pareto-efficient form of contingent action or linkage.[106]

If the treaty itself lacks incentives for state B to comply, that is, diverse performance reciprocity is insufficient, only an external reward makes compliance attractive for B.

External rewards can be found in linkage constellations, enhancing reputation, side payments, and rewards given within the treaty but on top of the bargain of the base treaty itself.

i. Reward via Linkage

A classical example for external rewards is linkage of treaties, be it at the entry stage or the compliance stage.[107] EU accession was used as an entry reward for cooperation with the International Criminal Tribunal for the former Yugoslavia (ICTY).[108] The cooperation and extradition requests by the ICTY were the base norms to be complied with and EU (and NATO) membership was the reward. The Republic of Croatia is a classic example. Croatia formally applied for EU membership on February 21, 2003. The European Council decided that accession negotiations with Croatia would open on March 17, 2005, provided that Croatia cooperated fully with the ICTY, in addition to the classical criteria for membership.[109] This meant that it had to take all necessary steps to ensure that the last remaining indictee was located and transferred to The Hague as soon as possible.[110]

[103] J. Samuel Barkin & Yuliya Rashchupkina, *Public Goods, Common Pool Resources, and International Law*, 111 AJIL 376 (2017).

[104] For concerns about moral hazard, see Section VI.A.1 *infra*.

[105] Koskenniemi, *supra* note 88, at 142.

[106] Thomas Bernauer & Dieter Ruloff, *Introduction and Analytical Framework*, *in* THE POLITICS OF POSITIVE INCENTIVES IN ARMS CONTROL 1, 3 (Thomas Bernauer & Dieter Ruloff eds., 1999).

[107] Paul Poast, *Issue Linkage and International Cooperation: An Empirical Investigation*, 30 CONFLICT MGMT. & PEACE SCI. 286 (2013). Note that issue linkage can also occur within a treaty, e.g. the WTO.

[108] *See* Nikola Brzica, *Croatia's Path to the EU Via the ICTY*, *available at* https://www.academia.edu/11787114/Croatias_Path_To_The_EU_Via_The_ICTY.

[109] Progress of accession talks is measured against a number of requirements, as laid down in 1993 in the "Copenhagen criteria" at that time (later enshrined in Article 6(1) of the Treaty on European Union and proclaimed in the Charter of Fundamental Rights), which examine the political and economic situation of the country as well as its ability to adopt the obligations of membership.

[110] European Commission Press Release, Croatia – One Step Closer to the EU, Provided There Is Full Cooperation with ICTY (Jan. 31, 2005), *at* https://ec.europa.eu/commission/presscorner/detail/en/IP_05_110;

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

Negotiations started only in October 2005, when the ICTY prosecution confirmed that Croatia was cooperating with the ICTY.[111] But even the arrest of General Ante Gotovina two months later did not fully improve relations with the ICTY, when ICTY Prosecutor Serge Brammertz stated that some requested documents were still missing.[112] It was thus only after the closing arguments of the Gotovina trial were held in September 2010 that tensions between the ICTY and Croatia ended and Croatia then became a member of the EU on July 1, 2013.

There are many more examples. Often treaties offer new commercial ties or the reduction of existing barriers to trade to create incentives for entering and complying with a (base) treaty. The Generalized System of Preferences (GSP) under the World Trade Organization (WTO) Agreements[113] serves as an example for external rewards to ensure entry and compliance with other international law, given that GSPs are not part of the reciprocity package of the WTO's bargain but remain unilateral and voluntary measures.[114] In the EU, the arrangement removes import duties from most products coming into the common market from designated GSP+ beneficiary countries. The GSP+ helps developing countries to alleviate poverty and create jobs, but at the same time requires the countries to respect core principles of an array of treaties including labor, human rights, and anti-corruption (the base treaties). The entry reward may be insufficient to induce compliance and thus the EU continuously monitors the beneficiary countries' effective implementation of their obligations and publishes a report on the implementation every two years. Both rewarding and rewarded countries can be better off compared to the status quo *ante* (seeing each country as a whole, notwithstanding winners and losers of trade within the country or third country effects). Here, cooperative benefits are the result of the respective linked treaty and accrue within that treaty, but they secure entry and compliance with a base treaty.

ii. Reputational Rewards

Moreover, external rewards can ease entry and compliance through a reputation channel. In our definition, a good reputation is an external reward for entering (normative

European Parliament, *Parliamentary Questions* (Sept. 20, 2005), *at* https://www.europarl.europa.eu/sides/getAllAnswers.do?reference=E-2005-1258&language=EN.

[111] Office of the Prosecutor Press Release, Assessment by the Prosecutor of the Co-operation Provided by Croatia, Press Release JP/MO/1009e (Oct. 3, 2005), *at* https://www.icty.org/sid/8535.

[112] United Nations Press Release, Prosecutor Brammertz's Address Before the Security Council, Press Release OK/OTP/1354e (June 18, 2010), *at* https://www.icty.org/sid/10423 ("the issue of the missing important documents related to Operation Storm in 1995 remains outstanding").

[113] Thirteen WTO members have notified their GSP schemes to the UN Conference on Trade and Development. For the EU: European Commission, *The EU's New Generalised Scheme of Preferences (GSP)*, (Dec. 2012), *available at* https://trade.ec.europa.eu/doclib/docs/2012/december/tradoc_150164.pdf; for the United States: Office of the United States Trade Representative, *Generalized System of Preferences (GSP)*, *at* https://ustr.gov/issue-areas/trade-development/preference-programs/generalized-system-preference-gsp; for Switzerland: Federal Customs Administration, *Developing Countries GSP (Generalized System of Preferences)*, *at* https://www.ezv.admin.ch/ezv/en/home/information-companies/exemptions–reliefs–preferential-tariffs-and-export-contributio/importation-into-switzerland/developing-countries-gsp–generalized-system-of-preferences-.html; Switzerland's GSP has no conditions and thus is not an external reward.

[114] *See* Kevin C. Kennedy, *The Generalized System of Preferences After Four Decades*, 20 MICH. ST. INT'L L. REV. 521, 528 (2012) ("As beneficiary countries cannot count on availability of preferences, the consequent uncertainty of market access is a major concern to the countries affected."). As alluded to before, rewards outside of treaties have more flexibility but less predictability.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

commitment) or complying (reliability). A good reputation will allow a state to find more partners and to extract more generous concessions in future transactions. It is especially important if internal rewards do not work. Rewarding can stimulate feelings of pride and positive self-image.[115] Social approval makes individuals happy and proud while disapproval causes embarrassment and shame and makes people unhappy.[116]

When praise for compliance, be it informally or in annual reports by international organizations, is attached to the base treaty as an instrument,[117] reputation is the benefit conveyed by it. While much of the literature has concentrated on naming and shaming,[118] literature on naming and praising in international law is rather scarce,[119] although it has been discussed (as negative and positive reputation) in the realm of Global Performance Indicators (GPIs), such as the World Bank's Ease of Doing Business Index and the Transparency Perception Index.[120] Declining in the rank of a GPI may damage reputation with various audiences (citizens, business, NGOs, other states), and climbing in a GPI may improve reputation.

Since reciprocity, retaliation, and outcasting do not work in the human rights sphere, reputation (next to assistance) is left as an *international* compliance mechanism. Shaming and its effect on reputation can also set in motion *national* processes through the activation of civil society and thus be one means of fostering compliance.[121] But focusing on bad reputation is not the only way to foster compliance—focusing on positive reputation may do so as well.

UN human rights treaty bodies adopt nonlegally binding "concluding observations" after consideration of periodic reports of states parties to the respective convention. They may include acknowledgment of positive steps taken by the state to achieve its obligations, identification of problematic areas that require further action by the state, discussion of practical steps that the state can take in order to improve its implementation of human rights standards,

[115] *See* Doron Teichman & Eyal Zamir, *Nudge Goes International*, 30 EUR. J. INT'L L. 1263, 1275 (2019) ("Credible publications of a country's international ranking can spur that country's elite decision-makers to action since a country's performance in international rankings can affect their self-esteem.").

[116] Ernst Fehr & Arming Falk, *Psychological Foundations of Incentives*, 46 EUR. ECON. REV. 687, 705 (2002).

[117] The Organisation for Economic Co-operation and Development and the Council of Europe, for example, conduct peer reviews, e.g., for anti-corruption measures. Praising in the reports could enhance the positive reputation. The WTO in its Trade Policy Review Mechanism, although mainly descriptive, also praises countries. *See, e.g.*, Trade Policy Review Northern Macedonia, WT/TPR/S/390 (June 11, 2019).

[118] Especially human rights scholars have researched the various effects of naming and shaming. *See, e.g.*, Jacob Ausderan, *How Naming and Shaming Affects Human Rights Perceptions in the Shamed Country*, 51 J. PEACE RES. 81 (2014); Emilie M. Hafner-Burton, *Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem*, 62 INT'L ORG. 689 (2008); Michelle Giacobbe Allendoerfer, Amanda Murdie & Ryan M. Welch, *The Path of the Boomerang: Human Rights Campaigns, Third-Party Pressure, and Human Rights*, 64 INT'L STUD. Q. 111 (2019). For micro-level evidence (the effect of shaming on domestic opinion), see Dustin Tingley & Michael Tomz, *The Effects of Naming and Shaming on Public Support for Compliance with International Agreements: An Experimental Analysis of the Paris Agreement* (Working Paper, 2019), *available at* https://scholar.harvard.edu/files/dtingley/files/tingley-tomzparis-shame.pdf.

[119] *But see* Margarita H. Petrova, *Naming and Praising in Humanitarian Norm Development*, 71 WORLD POL. 586 (2019).

[120] JUDITH G. KELLEY, SCORECARD DIPLOMACY: GRADING STATES TO INFLUENCE THEIR REPUTATION AND BEHAVIOR (2017); Judith G. Kelley & Beth A. Simmons, *Politics by Number: Indicators as Social Pressure in International Relations*, 59 AM. J. POL. SCI. 551146 (2015); Judith G. Kelley & Beth A. Simmons, *Introduction: The Power of Global Performance Indicators*, 73 INT'L ORG. 491 (2019); Kevin E. Davis, Benedict Kingsbury & Sally E. Merry: *Introduction: Global Governance by Indicators*, *in* GOVERNANCE BY INDICATORS: GLOBAL POWER THROUGH QUANTIFICATION AND RANKINGS 3 (Kevin E. Davis, Angelina Fisher, Benedict Kingsbury & Sally Engle Merry eds., 2012).

[121] BETH A. SIMMONS, MOBILIZING FOR HUMAN RIGHTS: INTERNATIONAL LAW IN DOMESTIC POLITICS (2009).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

and follow-up on implementation of the concluding observations. These concluding observations thus already include "good practices" or positive aspects before turning to areas of concern and recommendations. For example, the concluding observations on the initial report of Greece on the Convention on the Rights of Persons with Disabilities[122] states that it "values the State party's measures to render public transport in Athens and other major cities accessible and the preservation of the nominal level of disability allowances during the economic and financial crisis."[123]

But UN treaty bodies could go further in using naming and praising as rewards. For instance, UN human rights treaty bodies could, under the current legal norms, reward countries for compliance with UN human rights treaties by flagging the best performers in its annual reports, in addition to naming the worst performers as they do now. Article 24 of the Convention against Torture (CAT),[124] for example, gives considerable leeway on how to write the annual report.[125] The Subcommittee on Prevention of Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment decided to identify those states parties whose establishment of their national preventive mechanism was substantially overdue and to record them on a public list.[126] This amounts to naming and shaming. Additionally, it could also name the best performers in submitting timely state reports in the last five or ten years (in addition to those states who have not submitted their due reports) and report on their best practices. This could be backed up with support provided—for example, through the Special Fund established under Article 26(1) of the Optional Protocol to the Convention Against Torture (OPCAT), which is directed toward projects aimed at establishing or strengthening national preventive mechanisms. Similar mechanisms could be established for other UN human rights treaties or regional human rights treaties, adding internal rewards to the external one of reputation.

Another example of rewarding in the human rights sphere can be derived from the Global Alliance of National Human Rights Institutions (GANHRI).[127] Its Subcommittee on Accreditation (SCA)[128] gives letter grades to individual National Human Rights Institutions (NHRIs) indicating compliance with the Paris Principles.[129] The SCA is unique within UN structures, serving as the gatekeeper of these international standards, independent of UN member states. More specifically, NHRIs are given a score of "A" (full compliance),

[122] Convention on the Rights of Persons with Disabilities, CRPD/C/GRC/CO/1 of Oct. 29, 2019 (Advance Unedited Version).

[123] *Id.*, para. II.

[124] Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, Dec. 10, 1984, 1465 UNTS 85, *reprinted in* 23 ILM 1027 (1984).

[125] Manfred Nowak, Moritz Birk & Giuliana Monina, The United Nations Convention Against Torture and Its Optional Protocol. A Commentary, Art. 24 (2019).

[126] *See* United Nations Human Rights Office of the High Commissioner, Optional Protocol to the Convention Against Torture (OPCAT) Subcommittee on Prevention of Torture, *at* http://www.ohchr.org/EN/HRBodies/OPCAT/Pages/Article17.aspx.

[127] The GANHRI is an international association of NHRIs from all parts of the globe, established in 1993, promoting and strengthening NHRIs to be in accordance with the Paris Principles.

[128] The UN Office of the High Commissioner for Human Rights serves as a permanent observer and secretariat to the SCA and the Sub-Committee invites information from third parties to inform its work.

[129] Principles relating to the Status of National Institutions (The Paris Principles), *adopted* by UN GA Res. 48/134 (Dec. 20, 1993). The Principles provide the international benchmarks against which NHRIs can be accredited by GANHRI.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

"B" (partial compliance), "C" (noncompliance), and zero if the NHRI was suspended or accreditation was revoked. This puts countries in the spotlight. Next to the reputational mechanism, NHRIs with an "A" are rewarded intangibly by being allowed to fully participate in the international and regional work, enter meetings of national institutions as voting members, and hold office in the Bureau of the International Coordinating Committee or any subcommittee the Bureau establishes. They can also participate in sessions of the Human Rights Council, take the floor for any agenda item, submit documentation, and take up separate seating. States with "B" status institutions may participate as observers in the international and regional work and meetings of the national human rights institutions. They cannot vote or hold office with the Bureau or its subcommittees. They are not given NHRIs badges, nor may they take the floor for agenda items, or submit documentation to the Human Rights Council.[130] This system shows how reputation and gradual rewarding by intangible benefits works, and it has been deemed very effective.[131]

iii. Side Payments

Side payments on top of the treaty are another (classical) example of external rewards.[132] They can be given for entry or compliance. Side payments can be considered as a reservation price paid to a target country to make it willing to enter a treaty and/or comply.[133] Side payments are commonly used to enhance international cooperation and can further expand the zone of potential agreement.[134] For example, the United States offered substantial economic and military aid to Egypt and Israel to sign a peace treaty.[135] In another instance, in 1990, the Soviet Union agreed to withdraw its troops from East Germany in return for economic aid.[136] A further example, described by Chayes and Chayes, involved funding provided by the United States to assure that successor states to the Soviet Union which contained nuclear weapons within their territories (Ukraine, Belarus, and Kazakhstan) could meet the requirements of the Strategic Arms Reduction Talks (START) to which the former Soviet Union had agreed and join the Non-Proliferation Treaty.[137]

[130] GANHRI Sub-Committee on Accreditation (SCA), *at* https://nhri.ohchr.org/EN/AboutUs/GANHRIAccreditation/Pages/default.aspxhttps://nhri.ohchr.org/EN/AboutUs/GANHRIAccreditation.

[131] Katerina Linos & Tom Pegram, *What Works in Human Rights Institutions?*, 111 AJIL 628 (2017).

[132] Whether those payments are made inside or outside the treaty can be a result of historical accident or of conscious design, especially in bilateral treaties. For analytical and legal purposes, the distinction still is of importance.

[133] A reservation price is the highest price acceptable to a buyer to pay for a good (in our case, the highest reward acceptable to an enforcer to give to the target to induce entry/compliance) and the lowest price acceptable to a seller to sell the good (the lowest reward acceptable to a target to enter/comply).

[134] *See* SCOTT BARRETT, ENVIRONMENT & STATECRAFT: THE STRATEGY OF ENVIRONMENTAL TREATY-MAKING 335–54 (2003).

[135] Drezner, *supra* note 39, at 188.

[136] Randall Newnham, *The Price of German Unity: The Role of Economic Aid in the German-Soviet Negotiations*, 22 GER. STUD. REV. 421 (1999).

[137] Chayes & Chayes, *supra* note 1, at 193; DAVIS, *supra* note 6, at 8; John M. Shields, *Conference Findings on the Nunn-Lugar Cooperative Threat Reduction Program: Donor and Recipient Country Perspectives*, 3 NONPROLIFERATION REV. 66 (1995); Sherman W. Garnett, *Ukraine's Decision to Join the NPT*, 25 ARMS CONTROL TODAY 7 (Jan./Feb. 1995), *at* https://www.armscontrol.org/blog/2014-03-08/ukraine-russia-npt.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

### iv. Rewarding on Top of the Bargain but Within the Treaty

Treaties dealing with public goods can provide rewards, like assistance, within the objectives of the treaty. They can also provide rewards outside the scope of the objective of the treaty but regulated within the treaty. This gives additional incentives for states to enter and comply with the treaty. For instance, in arms control, low-capacity countries have claimed their willingness to engage in new arms control commitments in return for assistance by wealthier countries.[138] Countries already contributing to the public good of arms control have incentives to reward countries for entering the commitment, since otherwise countries continue to engage in their harmful behavior. One example is the International Atomic Energy Agency (IAEA) aiming to "accelerate and enlarge the contribution of atomic energy to peace, health and prosperity."[139] The IAEA verifies through its inspection system that states comply with their commitments to use nuclear facilities only for peaceful purposes.[140] States agree to grant the IAEA access to peaceful nuclear facilities and to allow it to employ various verification systems. It offers its member states assistance in the planning and generation of electricity and facilitates the transfer of technology and knowledge (internal rewards). But it also promotes the achievement of the participating states' development goals concerning issues such as poverty, hunger, health, clean water and energy, and climate change (external rewards) by providing assistance in nuclear science and technology.[141]

Another example is the declaration of a Marine Protected Area (MPA) as an important means for improving biodiversity and fish resources to protect certain species to which that specific area is especially important.[142] Research shows furthermore that strategically expanding the existing global MPA network to protect an additional 5 percent of the ocean could increase future catch by at least 20 percent via spillover.[143] This amounts to internal rewarding within the Biodiversity Convention, but given the severity of the problem and the gains from mitigating it, external rewarding could be added, since there are considerable spill-over benefits to other countries from increased fish-stock by strategically creating a network of MPAs.[144] In both cases, those rewards could be given in the form of positive transfers to or naming and praising of the countries declaring the MPAs by the Biodiversity Secretariat or the Conference of state parties of the Biodiversity Convention.

In summary, the typology provides possibilities for the usage of rewards under different circumstances and underlying problem structures (game theoretically speaking) and their combination. Its toolbox also illuminates how and which rewards can be used if penalizing mechanisms do not work by themselves. In global or regional public good constellations (like

[138] Bernauer & Ruloff, *supra* note 106, at 7.

[139] International Atomic Energy Agency, *The IAEA Mission Statement*, *at* Article II of the IAEA Statute.

[140] International Atomic Energy Agency, *The IAEA Mission Statement*, *at* https://www.iaea.org/about/mission.

[141] Nicole Jawerth & Miklos Gaspar, *How the IAEA Will Contribute to the Sustainable Development Goals*, IAEA (Sept. 25, 2015), *at* https://www.iaea.org/newscenter/news/how-iaea-will-contribute-sustainable-development-goals.

[142] For details, see Sarah Wolf & Jan Asmus Bischoff: *Marine Protected Areas*, *in* Max Planck Encyclopedia of Public International Law (Rüdiger Wolfrum ed., 2013), *at* https://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e2029.

[143] Reniel B. Cabral, et al., *A Global Network of Marine Protected Areas for Food*, 117 PNAS 45, 28134, 28137 (2020) ("MPAs can substantially boost fisheries productivity while simultaneously providing other ecosystem and conservation benefits."). They cover 811 species.

[144] Convention on Biological Diversity, June 5, 1992, 1760 UNTS 79.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

MEAs, arms control treaties, and human rights treaties) coupled with incapacity, assistance as a reward, internal or external, may be needed and sometimes even topped by side payments. Negative reciprocity and complete outcasting would be undesirable in these constellations. In order to prevent negative reciprocity spirals, a shift toward positive reframing of compliance can be achieved by changing narratives of reciprocity. This connects to another way of rewarding: the use of positive reputation highlighting compliance instead of violations. If needed, this can be topped up with other external rewards, such as linkage of treaties or assistance on top of the objective of the treaty.

## IV. Rewards Versus Penalties: A Rationalist Analysis

In a rational choice framework, a target country does not comply if noncompliance is more beneficial than compliance. A rational enforcer can induce the target country to comply by penalizing the target country for its noncompliance or by rewarding the target country conditional on its compliance. The enforcer will penalize/reward if the benefit expected from the target's compliance is higher than the cost of penalizing/rewarding. To induce compliance, the penalty/reward has to offset the target's gains from noncompliance.

In principle, either sticks or carrots should equally lead to the target's compliance because both means produce the same opportunity costs (costs of noncompliance) to the target country. For instance, if the target country receives a benefit from noncompliance of twenty dollars, then a penalty of twenty-one dollars or a reward of twenty-one dollars should make the target equally compliant. If the enforcer applies a penalty, the target country will suffer a penalty of twenty-one dollars if it does not comply (a net loss of one dollar). If the enforcer applies a reward, the target country will suffer a forgone reward of twenty-one dollars if it does not comply (a net loss of one dollar).

The question then is, when does an enforcer use a reward and when a penalty? A rational enforcer will choose the less costly measure. Both rewards and penalties produce costs. Some scholars have claimed that penalties are always superior to rewards when they are credible, because a credible threat does not need to be carried out and is therefore costless. We dispute this assertion as threats produce costs as well: not only is building up the capacity to make threats and keeping the threat costly, but acquiring a reputation for punishing violators is also costly. Because penalties can be very costly—we will discuss those costs in the following sections—the rational enforcer may consider offering a reward. That rewards are a means to elicit cooperation can be traced back to the Coasean theorem. The Coase theorem stresses that—absent transaction costs—it does not matter to whom one initially assigns a right. Another party may (and if rational should and would) pay the right holder to relinquish it, so long as both would be better off.

Since rewards and penalties are conceptually symmetric under rationalist assumptions, less effort has been undertaken to analyze their differences, assuming that all or most generalizations about penalties are applicable to rewards.[145] In this Part, we consider some differences

[145] In international law, to our knowledge, the only paper is Ben-Shahar & Bradford, *supra* note 4, which remains rationalist. For a law and economics approach on the distinction between rewards and penalties, see Giuseppe Dari-Mattiacci & Gerrit De Geest: *Carrots vs. Sticks*, *in* The Oxford Handbook of Law and Economics: Volume 1: Methodology and Concepts 439 (Francesco Parisi ed., 2017). The field of IR offers

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

between rewarding and penalizing that we deem important for the analysis of rewarding in international law.

### *A. Costs*

Rewards and penalties differ in the costs they produce; we will look at the costs that occur from applying penalties before turning to the costs produced by rewards.[146] Costs are mostly discussed in relation to retaliation, such as sanctions. Economic sanctions as described by former UN Secretary-General Kofi Annan "represent more than just verbal condemnation and less than the use of armed force."[147] They are often considered as a "milder" substitute for military confrontation.[148] Substantial research shows a bleak picture on costs and effectiveness of sanctions.[149] Many studies show that economic sanctions increase poverty and income inequality in the target country,[150] negatively affect the availability of food and clean water,[151] and adversely affect life expectancy—especially for women[152]—and infant mortality.[153] Other scholars point out that economic sanctions worsen the targeted government's respect for human rights[154] and have a detrimental impact on the level of democracy.[155] They are also economically painful for the receiving country. For instance, Neuenkirch and Neumeier find that sanctions imposed by the United Nations and the United States reduce the target state's GDP by 25 percent and 13 percent, respectively.[156] A similar picture

the most literature relevant to our analysis, see references in note 4 *supra*. However, also here the analysis of coercion is dominating, and rewards are still undertheorized.

[146] Costs and benefits we analyze here are those directly accruing to the states. We are keenly aware and would like to thank an anonymous reviewer for pointing out that other actors may, for instance, bear the cost or burden of delivering on a promised reward, or the cost or burden of not being similarly rewarded or supported. A reward in the form of a trade treaty may, for example, impact a third country due to trade diversion. Yet, analyzing those indirect and network effects must be left for another paper.

[147] UN Press Release, Secretary-General Reviews Lessons Learned During "Sanctions Decade" in Remarks to International Peace Academy Seminar, Press Release SG/SM/7360 (Apr. 17, 2000), *at* https://www.un.org/press/en/2000/20000417.sgsm7360.doc.html.

[148] *See, e.g.*, David S. Cohen & Zachary K. Goldman, *Like It or Not, Unilateral Sanctions Are Here to Stay*, 113 AJIL Unbound 146 (2019).

[149] For a recent discussion on the detrimental impact of sanctions, see Dursun Peksen, *Political Effectiveness, Negative Externalities, and the Ethics of Economic Sanctions*, 33 Eth. & Int'l Aff. 279 (2019). *See also* Thomas G. Weiss, David Cortright, George A. Lopez & Larry Minearet, Political Gain and Civilian Pain: Humanitarian Impacts of Economic Sanctions (1997).

[150] Sylvanus Kwaku Afesorgbor & Renuka Mahadevan, *The Impact of Economic Sanctions on Income Inequality of Target States*, 83 World Dev. 1 (2016); Matthias Neuenkirch & Florian Neumeier, *The Impact of US Sanctions on Poverty*, 121 J. Dev. Econ. 110 (2016); Seung-Whan Choi & Shali Luo, *Economic Sanctions, Poverty, and International Terrorism: An Empirical Analysis*, 39 Int. Interact. 217 (2013).

[151] David Cortright & George A. Lopez, *Learning from the Sanctions Decade*, 2 Glob. Dialogue 11 (2000); Elizabeth Gibbons & Richard Garfield, *The Impact of Economic Sanctions on Health and Human Rights in Haiti, 1991–1994*, 89 Am. J. Pub. Health 1499 (1999).

[152] Jerg Gutmann, Matthias Neuenkirch & Florian Neumeier, *Sanctioned to Death? The Impact of Economic Sanctions on Life Expectancy and its Gender Gap*, 57 J. Dev. Stud. 139 (2021).

[153] Beth Osbourne Daponte & Richard Garfield, *The Effect of Economic Sanctions on the Mortality of Iraqi Children Prior to the 1991 Persian Gulf War*, 90 Am. J. Pub. Health 546 (2000).

[154] Dursun Peksen, *Better or Worse? The Effect of Economic Sanctions on Human Rights*, 46 J. Peace Res. 59 (2009).

[155] Dursun Peksen & A. Cooper Drury, *Coercive or Corrosive: The Negative Impact of Economic Sanctions on Democracy*, 36 Int'l Interactions 240 (2010).

[156] Matthias Neuenkirch & Florian Neumeier, *The Impact of UN and US Economic Sanctions on GDP Growth*, 40 Eur. J. Pol. Econ. 110 (2015).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

emerges for so-called targeted or smart sanctions.[157] These also have unintended consequences, including increases in corruption and criminality, strengthening of authoritarian rule, burdens on neighboring states, strengthening of political factions, resource diversion, and humanitarian impacts.[158]

Furthermore, penalties are also costly to the imposing country. Economic sanctions significantly reduce the volume of bilateral trade between the imposing and the target state.[159] The imposition of economic sanctions may interrupt trade and financial contacts of domestic firms with the sanctioned counterpart, generating deadweight losses because import substitutes have to be produced at home at higher costs and the demand abroad shrinks, making economic operators in the sanctioning country economically worse off.[160]

Thus, there is a broad consensus among scholars that sanctions produce substantial costs and in many cases fail.[161] The same holds true for targeted sanctions.[162]

The costs of penalties are not only a question of how many potential violators there are, and what a specific act of enforcement costs, but how costly maintaining the *threat* of the sanctioning mechanism is. One may argue that the *threat* of negative reciprocity is inefficient only if it is unsuccessful; if it induces compliance then it is efficient since it ensures a compliance

[157] Smart sanctions have increasingly replaced conventional sanctions. They include actions such as asset freezes, restrictions on luxury goods sales, travel and financial restrictions that should directly affect parties that have some leverage on the regime (e.g., political elites). In theory, targeted sanctions should be more effective as they pressure political elites and decrease the negative effects in the civil population. However, current literature suggests that comprehensive and targeted sanctions both result in civilian pain while political elites remain unharmed. *See* Peksen, *supra* note 149, at 280 (reviewing some of the most up-to-date empirical research on sanctions and stating "that both comprehensive and targeted sanctions remain morally impermissible tools due to their substantial negative externalities and low success rate").

[158] Thomas Bierstecker, Sue E. Eckert, Marcos Tourinho & Zuzana Hudáková, The Effectiveness of United Nations Targeted Sanctions. Findings from the Targeted Sanctions Consortium (TSC) 8 (2013). For UN sanctions, they warn that another consequence of ineffective efforts to constrain is the impact they can have on the credibility of the UN itself (possibly in part due to overuse of sanctions for ineffective purposes).

[159] Gary Clyde Hufbauer, Jeffrey J. Schott, Kimberly Ann Elliott & Barbara Oegg, Economic Sanctions Reconsidered: History and Current Policy (2009). In a recent study covering Iran for the years 1950–2015, it was shown that "all else equal, on average complete bilateral sanctions have the potential to reduce trade among participants by 85% to 86%." *See* Gabriel Felbermayr, Constantinos Syropoulos, Erdal Yalcin & Yoto V. Yotov, *On the Effects of Sanctions on Trade and Welfare: New Evidence Based on Structural Gravity and a New Database*, at 42 (CESifo Working Paper No. 7728, 2019).

[160] *See, e.g.*, Mary Amiti, Stephen J. Redding & David E. Weinstein, *The Impact of the 2018 Tariffs on Prices and Welfare*, 33 J. Econ. Perspec. 187, 188–89 (2019) (they found that "by December 2018, import tariffs [introduced by the Trump administration] were costing US consumers and the firms that import foreign goods an additional $3.2 billion per month in added tax costs and another $1.4 billion per month in deadweight welfare (efficiency) losses"). This must hold more forcefully for embargos using quantitative trade restrictions. *See also* Ben-Shahar & Bradford, *supra* note 4, at 386–89 on domestic firms' costs.

[161] *See* Hufbauer, Schott, Elliott & Oegg, *supra* note 159 (finding that economic sanctions are partially successful in 34% of the cases). Robert Pape found that economic sanctions are effective in around 5% of the cases. Robert A. Pape, *Why Economic Sanctions Do Not Work*, 22 Int'l Sec. 90 (1997); Robert A. Pape, *Why Economic Sanctions Still Do Not Work*, 23 Int'l Sec. 66 (1998). However, it should be noted that it is difficult to measure success, as one component is the deterrence of future noncompliance and this cannot be observed.

[162] Bierstecker, et al., *supra* note 158. One can safely assume that UN targeted sanctions should be more effective than unilateral (or regional) ones, given the number of states implementing them even if considering the practice of secondary sanctions. Assessing their effectiveness is complex, since those sanctions have different goals, they may be designed to (1) change behavior (to coerce), (2) constrain behavior (to constrain), and/or (3) send a signal (to signal). Based on an analysis of twenty-two UN targeted sanctions regimes, the authors conclude that UN targeted sanctions are effective in achieving at least one of the three purposes of sanctions 22% of the time with a higher rate for constraining and signaling but effective in coercing only about 10% of the time.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

equilibrium. But the sanctioner can already incur costs at the threatening stage. For example, threats of sanctions may trigger uncertainties concerning trade and investment policies.[163] It is clearly costly to maintain a large military for contingencies. This aspect of the costs of penalties is often neglected. In comparison, rewards may not have such negative consequences on the target country's humanitarian situation but rewards must be paid, and this requires the necessary capacity of the sending state.[164] In other words, the credibility of a prospective reward depends on whether the rewarding entity has sufficient resources to provide them. The wealth of the sender is an upper constraint on the use of rewards as the sender must be able to pay the reward (similarly, the wealth of the receiver is the upper constraint on the use of some penalties).[165]

One problem arises from how many countries can be rewarded. That depends on the type of reward. Internal rewards can be used for a large number of countries and may even produce "economies of scale."[166] This is different for external rewards such as payouts and assistance since funds are especially scarce on the international level. Those can be costly when many states have to be incentivized to enter and comply but are feasible when only few countries have to be incentivized, e.g., to provide a public good.[167] However, even if the number of states that require rewards appears limited, that number may not be static, and could multiply. If a state can reap not only the intrinsic rewards of participating in a treaty, but also extract a side payment for full compliance, presumably more states will insist on the side payment as a necessary part of the bargain.[168] Not only may they feel fully justified in doing so, as with treaties that assist lower income states, but the rewards may after some period of time also been seen as an entitlement.[169] Intangible rewards such as praising, in contrast, provide another type of reward that can be multiplied to more compliant states at lower costs.

[163] The IMF Global Uncertainty Index measures this. It is deemed relevant since the "index is associated with greater economic policy uncertainty, stock market volatility, risk, and lower GDP growth." *See* Hites Ahir, Nicholas Bloom & Davide Furceri, *60 Years of Uncertainty*, 57 FIN. DEV. 58, 59 (2020); Clayton Webb, *Re-examining the Costs of Sanctions and Sanctions Threats Using Stock Market Data*, 46 INT'L INTERACTIONS 749, 771 (2020) (Webb concludes that "sanctions threats are not costless. The results show that sanctions threats create stock volatility, even when sanctions have not been imposed. This volatility imposes costs on firms.").

[164] The question of initial endowments and capacity is connected to the Coase theorem which is generally overlooked in scholarship. In order to be able to bargain, a minimal capacity is necessary. This is not the same as transaction costs. This problem also applies to penalties since imposing the penalty requires resources and capacity. In addition, if we consider monetary fines, the receiver has to be able to pay the fine, thus penalties may incur capacity problems on both sides, receiver and sender, which is worse in comparison with rewards facing capacity problems only on the sender side.

[165] *See* Dari-Mattiacci & De Geest, *supra* note 145, at 448.

[166] Economies of scale describe the lower production costs of goods with increasing production output. In our case, additional members reduce the average costs of providing the (public) good. At the same time, the larger the number of participants in a treaty the higher is the collective benefit of the treaty.

[167] *See, e.g.*, Pamela Oliver, *Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations*, 85 AM. J. SOC. 1356 (1980) (pointing out that the relative costs of using rewards or punishments to produce a public good depend on the fraction of cooperators out of potential cooperators required to produce that good). On the different types of global public goods and the necessary mode of providing them with a view on the number of states, see SCOTT BARRETT, WHY COOPERATE?: THE INCENTIVE TO SUPPLY GLOBAL PUBLIC GOODS (2007).

[168] On the moral hazard problem, see Section VI.A.1 *infra*.

[169] This may also be positive: if an endowment effect plays a role, states will not be willing to give up the reward and thus may be further pushed toward compliance. *See* note 69 *supra*.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

### *B. Rewarders' and Volunteers' Dilemmas*

Another problem connected to costs arises from how to incentivize countries to contribute to the reward (rewarders' dilemma) and/or to volunteer to provide a public good (volunteers' dilemma). The rewarders' dilemma is that states would rather free ride on other countries providing the reward. The rewarders' dilemma is mitigated if rewarding generates gains for the giver(s), e.g., when states have a strong preference for the public good to be provided. Benefits vary according to the nature of the treaty or the global public good dealt with in the treaty. Rewarding may generate a positive reputation at the enforcement level (as we will see in the next Section) that may alleviate the rewarders' dilemma. Especially costly rewards are prone to the rewarders' dilemma; intangible rewards can be less costly and therefore less subject to the dilemma. Even with this dilemma, experiments demonstrate that individuals typically choose to reward, regardless of cost, and that rewards increase other individuals' contributions to public goods.[170]

The volunteers' dilemma captures the expectation that one prefers to free ride on the effort of other volunteers.[171] A classic example involves bystanders observing a person in danger. One bystander is necessary to help the person but if no bystander volunteers the person suffers harm. The optimal solution to the volunteers' dilemma does not require each individual to fully volunteer; rather, coordination may be needed for assigning the volunteer. Leshem and Tabbach show that for nearly all numbers of volunteers, rewards are more efficient than penalties.[172] There are many instances where a single country can produce alone or contribute to a (global) public good, i.e., volunteer. While tangible rewards can be costly when they have to be multiplied for many countries, they may be especially useful for incentivizing only one volunteering country.[173] Penalties are hard to imagine in those instances. Costs are always smaller with rewarding than with penalizing with regard to the provision of a single shot global public good.[174] For instance, saving the planet from an asteroid does not require all states to contribute to the public good. Another example would be rewarding (tangibly or intangibly) the declaration of an MPA as an important means for improving biodiversity

[170] *See, e.g.*, David G. Rand, Anna Dreber, Tore Ellingsen, Drew Fudenberg & Martin A. Nowak, *Positive Interactions Promote Public Cooperation*, 325 SCI. 1272 (2009)("We show that reward is as effective as punishment for maintaining public cooperation and leads to higher total earnings. Moreover, when both options are available, reward leads to increased contributions and payoff, whereas punishment has no effect on contributions and leads to lower payoff. We conclude that reward outperforms punishment in repeated public goods games and that human cooperation in such repeated settings is best supported by positive interactions with others."); Matthias Sutter, Stefan Haigner & Martin G. Kocher, *Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations*, 77 REV. ECON. STUD. 1540 (2010) ( "groups typically vote for the reward option"). *See also* Martin Sefton, Robert Shupp & James M. Walker, *The Effect of Rewards and Sanctions in Provision of Public Goods*, 45 ECON. INQ. 671, 684 (2007). For a summary of literature of reward and punishment in social dilemmas constellations, see REWARD AND PUNISHMENT IN SOCIAL DILEMMAS (Paul A. M. van Lange, Bettina Rockenbach & Toshio Yamagishi eds., 2014).

[171] Andreas Diekmann, *Volunteer's Dilemma*, 29 J. CONFLICT RES. 605 (1985).

[172] Shmuel Leshem & Avraham Tabbach, *Solving the Volunteer's Dilemma: The Efficiency of Rewards Versus Punishments*, 18 AM. L. ECON. REV. 1 (2016). This is in line with Gerrit De Geest & Giuseppe Dari-Mattiacci, *The Rise of Carrots and the Decline of Sticks*, U. CHI. L. REV. 341 (2013) (pointing out that rewards can be superior to penalties when the lawmaker requires higher efforts from some citizens than from others, for instance, when only some families need to sacrifice land for a highway project).

[173] *See* Giuseppe Dari-Mattiacci & Gerrit De Geest, *Carrots, Sticks, and the Multiplication Effect*, 26 J. L. ECON. & ORG. 365 (2010).

[174] BARRETT, *supra* note 167, ch. 4.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

TABLE 2.
REWARDERS' AND VOLUNTEERS' DILEMMAS

| | Rewarding state = 1 | Rewarding state = N |
|---|---|---|
| Rewarded state = 1 | No volunteers' dilemma / No rewarders' dilemma | Rewarders' dilemma |
| Rewarded state = N | Volunteers' dilemma | Volunteers' dilemma / Rewarders' dilemma |

in an exclusive economic zone to protect certain species to which that area is especially important or if that area is of special importance for a network of MPAs.[175] The compliance stage of CITES could yet be another example,[176] if African states would be rewarded for keeping the numbers of endangered species at a sustainable rate. Given that ever more private-public partnerships are set up for the management of protected parks[177] and these generate income from tourism, intangible rewards from praising[178] by states would be followed by tangible ones from third parties (like income from tourism).[179]

Both dilemmas are captured in Table 2. If we have one possible rewarding state and one possible rewarded state, neither of the two dilemmas occurs. When there is one possible rewarding state but *N* states that could be rewarded, the question arises of who should volunteer and thus be rewarded (volunteers' dilemma). If there are *N* prospective rewarding states and one prospective state to be rewarded, the question arises of who will provide the reward (rewarders' dilemma). When there are *N* prospective rewarding and *N* prospective rewarded states, both questions arise: who should be rewarded and who is rewarding.

### *C. Pareto Efficiency*

When the enforcing country expects a higher loss from the target's noncompliance than the target country gains, there is room for Coasean bargaining. The enforcing country can structure the reward to be conditional upon the target country's compliance. Two assumptions are crucial for the pareto efficiency of rewards: expectation and conditionality.

A rational enforcer will only offer a reward if he or she expects a gain from the target's compliance that is higher than the cost of rewarding. How high must the reward be? Recall that in a rational choice framework, the target country would not comply if noncompliance is more beneficial than compliance. Thus, the reward has to offset the target's gains from noncompliance to induce compliance. In other words, a reward has to compensate the target for its compliance. A reward therefore has to make the target country better off compared to the status quo wherein the target does not comply and rewards are not considered. Rewards have a built-in compensation mechanism, which means that rewards always allow the target country to opt for the status quo.[180]

[175] *See* note 142 *supra* and accompanying text.
[176] Note 48 *supra*.
[177] *See, e.g.*, African Parks, *at* https://www.africanparks.org/about-us/our-story.
[178] CITES uses praising already. *See, e.g.*, CITES, *New National CITES Enforcement Coordinating Body Shows Positive Results*, *at* https://www.cites.org/eng/news/pr/2012/20120509_certificate_cn.php.
[179] Peter Lindsey, et al., *Conserving Africa's Wildlife and Wildlands Through the COVID-19 Crisis and Beyond*, 4 NATURE ECOL. EVOL. 1300 (2020).
[180] *See* Dari-Mattiacci & De Geest, *supra* note 145, at 453f.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

Another important assumption for pareto efficiency is conditionality, which conditions provision of the reward upon compliance by the target country. If a reward is paid before the target complies it can lead to opportunistic behavior, that is, after receiving the reward the target country may decide not to comply. In that case, the rewarding country is worse off—the target country did not comply while the enforcer paid the reward, and thus the reward is not pareto efficient. This problem is alleviated when rewards are paid conditional upon the target's compliance. If the target fails to comply, the enforcer will not pay the reward and the enforcer is not worse off. Note that in a repeated game with reputation, it can be expected that the target does not behave opportunistically when receiving the reward *ex ante*. A reputation for complying makes the target country likely to receive further rewards. In that case, even though rewards are given *ex ante* they still lead to compliance and thus remain pareto efficient.

In summary, when rewards are not conditional they can lead to pareto inefficiency but if the (prepaid) reward is followed by compliance the reward is pareto efficient. When a reward is beneficial in expectation and conditional it is always pareto efficient because it makes no country worse off while at least one country is better off.

An enforcing country can also penalize the violator for noncompliance. In a rational choice framework, one difference between rewards and penalties is that penalties do not lead to pareto efficiency. Why is that the case? In the status quo of no penalty, the target country complies if the gain from compliance is higher than the gain from noncompliance and there would thus be no need for a penalty because the target would comply anyway. The pareto inefficiency results from the fact that penalties always make the target country worse off compared to the status quo in which the target country does not comply and does not receive a penalty.[181] If the target complies, it forgoes gains from noncompliance (gains under the status quo) and it may also incur some effort costs of compliance (e.g., destroying weapons requires effort and thus is costly). Note that this also applies to threats that are deterrent.[182] Because the target country is always worse off compared to the status quo of noncompliance and no penalty, penalties cannot meet the requirement of pareto efficiency. Penalties can at best be Kaldor-Hicks efficient if the overall benefit of the punishing countries outweighs the loss of the punished states.[183]

For instance, in arms control, if country A provides a reward to country B for giving up weapons it would acquire otherwise, then the acceptance of the reward by country B increases the welfare on both sides—given that the reward is conditioned and beneficial in expectation to the enforcing country. If, in contrast, country B is subject to penalties for acquiring weapons, then country B either suffers from the effort costs of giving up certain weapons or it suffers from the penalty if it does not give up the weapons. Therefore, under a penalty mechanism, the target country is always worse off compared to the status quo.

[181] *Id.*

[182] Also as highlighted in the Section Costs, threats incur costs to the target (and the enforcer) already in the threatening stage.

[183] The Kaldor-Hicks efficiency criterion is less stringent than pareto efficiency in that if those that are made better off could in principle compensate those that are made worse off, so that a pareto-improving outcome could (though does not have to) be achieved. Thus, an allocation of resources is said to be Kaldor-Hicks efficient when it produces in sum more benefits than costs.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

### *D. Reputation of the Enforcing Country*

Reputation acts negatively and positively and on two levels. The first level concerns the receiving country and how reputation affects its decision to enter or comply with a base treaty.[184] The second level, which is the focus here, regards the enforcement level, namely the reputation of the enforcing (rewarding/penalizing) country. The sanctioning dilemma has been at the forefront of discussion on compliance, leading to the prognosis that costly retaliation will seldom take place. However, states may still refer to external penalties in order to build up a reputation for penalizing violators, so that other states will be less likely to breach their obligations.[185] Imposing external penalties may hurt ties with the receiving country and is prone to generate feelings of hostility.[186] Penalties may also decrease the willingness of other countries (third countries that are not subject to the penalty) to cooperate with the penalizing country, especially when the penalty is perceived as unfair. States then could be deterred from entering a treaty with the penalizing country in the first place.

Similar to penalties but different in direction, rewards can generate a reputation for appreciating those who honor their obligations. The reliance on rewards to appreciate countries' compliance may generate a reputation of goodwill.[187] A reputation of goodwill in turn facilitates future cooperation with the target and other countries.[188] Additionally, rewards may be used to keep a reputation of good intentions, especially in relations considered as friendly. Furthermore, the more dependent the relationship, the more important obtaining the approval of the other nation will be and the higher the incentive to grant (and reciprocate) rewards.[189]

Thus, even though costly today, states may refer to rewards to build up a reputation of good intentions that in turn eases (future) cooperation. However, a reputation for not penalizing may signal a permissive attitude and allow more violations to take place. Rewarding in a relationship that is considered as adversarial may not be accepted by the sender's citizens; instead penalties are used to demonstrate firmness.[190] At the same time, penalties fulfill an important signaling function: disapproval. Nevertheless, penalties that incorporate some rewarding may

[184] *See* Section III.B.2.ii *supra*.

[185] *See* Sykes & Guzman, *supra* note 35, at 443 ("The decision to bear the costs of retaliation, then, is justified by a desire to persuade others that a state will punish violations. If a state is successful in building this reputation for punishing violators, other states will be less likely to breach their obligations.").

[186] *See also* Section V.B *infra*, dealing with difference in perception.

[187] David Cortright, The Price of Peace: Incentives and International Conflict Prevention 10–11 (1997) ("Perhaps the greatest difference between sanctions and incentives lies in their impact on human behavior. Drawing on the insights of behavioral psychology, Baldwin has identified key distinctions between the two approaches. Incentives foster cooperation and goodwill, while sanctions create hostility and separation."). And further, at 11 ("Punitive measures may be effective in expressing disapproval of a particular policy, but they are not conducive to constructive dialogue. Where sanctions generate communications gridlock, incentives open the door to greater interaction and understanding.").

[188] *See also* Section V.4. dealing with difference in stability.

[189] Martin Patchen, Resolving Disputes Between Nations: Coercion or Conciliation? 266 (1988).

[190] *See, e.g.*, Ben-Shahar & Bradford, *supra* note 4, at 385 ("Rewarding rogue regimes could be particularly difficult to justify to the domestic audience that contests the moral rationale for bribing belligerent countries."). Similarly, in arms control, see Bernauer & Ruloff, *supra* note 106, at 6 ("The strong focus on negative incentives may also be attributable to the fact that many academics and practitioners tend to dislike the idea of rewarding hold-out or laggard countries for not collaborating voluntarily in arms control. They would rather bully even bomb reticent countries into line than bribe them.").

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

absorb some of its positive characteristics (e.g., reduced hostility, good will) and may be a more efficient incentive compared to pure penalizing.[191]

### *E. Monitoring*

Differences between penalties and rewards are further reflected in the monitoring of compliance. The problem of monitoring is that the target state has incentives to hide or misrepresent information to avoid the prospect of a penalty or to attract a reward by falsely claiming compliance.[192] There are an array of monitoring and verification mechanisms within international law that parties are able to "cheat."[193] For instance, a test-ban treaty (that forbids nuclear weapons testing) cannot be perfectly monitored because nuclear tests are detected by observing seismic activity (e.g., an earthquake).[194] This incentivizes states to cheat by exploding nuclear devices while claiming that the outcome was caused by seismic activity.[195]

The question thus is which of the two mechanisms better incentivizes states to reveal truthful information. Actors with favorable information usually disclose it whereas parties with unfavorable information keep silent.[196] With rewarding, the incentive to provide information by the relevant state is higher, especially "to provide information on problems they encounter in implementing international commitments—if a country has no such problems—it will not receive assistance."[197] This effect is enhanced if the rewarded state is the one with the duty to provide the information in order to receive the reward. With threats, the common strategy is to act as if it did not hear or understand the threat.[198] The target's incentives with rewards are different:

> It is likely to make every effort to show it has heard the source and is behaving accordingly in order to reap the reward. Threats promote deceptive behavior on the part of targets, distrust on the part of the source. In contrast, promises can promote open and honest action on the part of targets and may promote greater trust in the overall relationship.[199]

While the misrepresentation of information may be enough to avoid a penalty when noncompliance cannot be directly observed, it may not be sufficient to receive a reward when the burden of proof of compliance is on the recipient's side and demands more information sharing.

[191] Eileen Filmus, *Sticks and Carrots in Coercive Diplomacy: Toward a Theory of Inducements* (Master's Thesis, U. Chi., 2015), *available at* https://www.academia.edu/21783586/Sticks_and_Carrots_in_Coercive_Diplomacy_Toward_a_Theory_of_Inducements.

[192] Although CHAYES & CHAYES, *supra* note 1, chapters 7 and 8 treat reporting, monitoring, and verification, they do not systematically analyze the incentives of states with a view on the consequences (rewarding or penalty).

[193] On monitoring and verification provisions and their variations, see: *id.*; KOREMENOS, *supra* note 20, ch. 9; Winfried Lang, *Compliance with Disarmament Obligations*, 55 HEIDELB. J. INT'L L. 69 (1995); Stefan Oeter, *Inspection in International Law: Monitoring Compliance and the Problem of Implementation in International Law*, 28 NETH. Y.B. INT'L L. 101(1997).

[194] Drezner, *supra* note 39, at 192.

[195] *Id.*

[196] Steven Shavell, *A Note on the Incentive to Reveal Information*, 14 GENEVA PAPERS RISK INSUR. 66 (1989).

[197] Bernauer & Ruloff, *supra* note 106, at 21.

[198] DAVIS, *supra* note 6, at 19.

[199] *Id.* at 19–20.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

Monitoring is often hampered by technical as well as political problems. Rewards in the form of technical support, for instance, can make monitoring compliance cheaper. Regarding the political problems, monitoring can interfere with sovereign rights of a country, such as onsite inspection of military installations in arms control, prisons under IHL, or international human rights law. Inspections may be perceived as more legitimate if rewards instead of penalties are promised and thus states will allow inspections to take place, easing monitoring. Rewards may also decrease the inclination of cheating by fostering cooperative relations. Lazear points out that rewards are superior when it is not clear what the maximum level of performance is, whereas penalties are superior when the minimum level of performance is unclear.[200]

## V. Rewarding: The (Behavioral) Difference It Makes

In the previous Part, we looked at differences between rewards and penalties from a rational choice perspective. In this Part, we deal with behavioral differences between rewards and penalties. As Baldwin expresses it, "When B reacts one way to a promise of $100 if he will do X, and another way to a threat to deprive him of $100 if he fails to do X, the concept of opportunity costs makes it difficult to explain why."[201] The importance of studying compliance theory from a psychological perspective arises from the different effects rewards and penalties have on human behavior. In the field of psychology, where rewards have been studied for a long time,[202] the literature generally seems to be more favorable toward rewards than toward penalties with respect to human behavior.[203] In international politics, two psychologists, Milburn and Christie, summarize rewarding as "an alternative without the major disadvantages of threat with its potential implications for instability, distrust, and mutual dislike."[204] That is, psychologists assume asymmetrical effects between rewards and penalties. Perhaps the single most important insight of cognitive psychology derives from Prospect Theory.[205] Prospect Theory questions the validity of the rationalist Coase theorem, the latter neutralizing the psychological contexts of human interactions whereas the former stresses the importance of the difference of perceived gains and losses for behavior. But can those insights be applied to states? We discuss this issue before turning to the psychological differences of rewards and penalties. As for the differences, penalties and rewards first differ in the receiver's perception: penalties are likely to be perceived as negative, rewards are evaluated positively. Second, penalties and rewards differ in the receiver's response: penalties are more likely to cause resistances or even counter-threats, rewards are more likely to be reciprocated.

[200] Edward P. Lazear, *Labor Economics and the Psychology of Organizations*, 5 J. Econ. Perspec. 89 (1991).

[201] Baldwin, *The Power of Positive Sanctions*, *supra* note 4, at 37.

[202] *See, e.g.*, Joseph Nuttin & Anthony G. Greenwald, Reward and Punishment in Human Learning (1968); Edward L. Thorndike, Educational Psychology (1913); Edward L. Thorndike, The Fundamentals of Learning (1932). Findings suggest that rewards result in more rapid learning than penalties. *See also* Bernard Berelson & Gary A. Steiner, Human Behavior: An Inventory of Scientific Findings (1964) (pointing out that regular penalties are deterrent rather than a stimulus to learning).

[203] For a critical discussion of the favorable tendency toward rewards in psychology, see Alfie Kohn, Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes (1993).

[204] Milburn & Christie, *supra* note 4, at 625.

[205] *See* text at note 12 *supra*.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

Third, rewards and penalties differ in the impact on international relations, and thus their stability: penalties are likely to increase conflicts, rewards to decrease them.

### *A. Applying Behavioral Insights to States*

The rational choice paradigm as employed in economics and IR theory informing international law has been challenged since the 1970s by psychological experimental research, with a revolutionary impact for economics and law and economics. This research shows that in contrast to the expected utility model, actors are only boundedly rational, and systematically have other-regarding preferences (both positive and negative). But is this even relevant for international law given that states are complex organizations? Political psychology in international relations has a long and strong tradition using psychological insights[206] and international law scholars have been taking up those insights more recently; we are thus not in uncharted waters.[207] There are two major challenges to applying psychological insights, especially when based on experiments, to international law.

The first challenge is the *relevant unit of analysis*. Whose behavior is at issue? Is it the state as a "black box" or is it individual actors, such as judges, political leaders, military commanders, trade negotiators, or other individuals, whose actions and decisions are attributable to the state under international law? Are we concerned with "elite" decisionmakers, experts, or the public? There is no methodological challenge if individual behavior is attributed to the state under international law. Or is it small decision-making groups, acknowledging that many decisions regarding international law-related conduct are made by such groups, and that group psychology is often different from individual decision making? If so, the research becomes more complex but we also know that group behavior deviates from rational choice assumptions—groups do not necessarily make a decision more rational.[208] Or do we take the state as unit of analysis? States are multi-sectoral, multi-agent entities and as such are complex organizations. There are three possible approaches to this problem. The first views the state as an organization. Psychology and behavioral economics are already being successfully applied to organizations, albeit mainly business organizations.[209] The second approach looks at the relationship between citizens and politicians. Whereas political economy has long explored domestic political processes and interactions between national and international politics (the

[206] *See, e.g.*, ROSE MCDERMOTT, POLITICAL PSYCHOLOGY IN INTERNATIONAL RELATIONS (2004); JAMES DAVIS, PSYCHOLOGY, STRATEGY AND CONFLICT: PERCEPTIONS OF INSECURITY IN INTERNATIONAL RELATIONS (2013). IR scholarship has revisited the incorporation of behavioral insights, but without addressing international law. *See, e.g.*, Emilie M. Hafner-Burton, Stephan Haggard, David A. Lake & David G. Victor, *The Behavioral Revolution and International Relations,* 71 INT'L ORG. 1 (2017) (special issue); James W. Davis & Rose McDermott, *The Past, Present, and Future of Behavioral IR*, 75 INT'L ORG. 147 (2021); John M. Gowdy, *Behavioral Economics and Climate Change Policy*, 68 J. ECON. BEHAV. & ORG. 632, 632 (2008) ("[Behavioral economics] suggests that the standard economic approach to climate change policy, with its focus on narrowly rational, self-regarding responses to monetary incentives, is seriously flawed.").

[207] Anne van Aaken, *Behavioral International Law and Economics*, 55 HARV. INT'L L.J. 421 (2014); Tomer Broude, *Behavioral International Law*, 163 U. PENN. L. REV. 1099 (2015), both with further references; Anne van Aaken & Tomer Broude, *The Psychology of International Law: An Introduction*, 30 EUR. J. INT'L L. 1225 (2019).

[208] Norbert L. Kerr, Geoffrey P. Kramer & Robert J. MacCoun, *Bias in Judgment: Comparing Individuals and Groups*, 103 PSYCHOL. REV. 687 (1996).

[209] Colin F. Camerer & Ulrike Malmendier: *Behavioral Economics of Organizations*, *in* BEHAVIORAL ECONOMICS AND ITS APPLICATIONS 235, 235 (Peter Diamond & Hannu Vartiainen eds., 2007).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

"two level game"),[210] behavioral political economy is still in its early stages, but is gaining ground.[211] The third approach simply attributes nonstandard preferences, beliefs, and decision making directly to states (or the individuals acting on its behalf); it is this approach we follow here for simplicity reasons. This can be defended since much of international law decision making is in essence made by individuals or small decision-making groups; the very term "state conduct" implies that states are regularly assimilated to individual actors. The rational choice approach is no different in this—in order to reduce complexity, the same behavioral assumptions are used on the individual and the state level.

The second challenge derives from the *experimental basis* of much of behavioral research (but not all psychological research) used in our context.[212] Applying experimental psychology and its methods to international law is feasible.[213] Some experimental results with intuitive appeal were confirmed by field studies (e.g., in the realm of commons)[214] and were empirically tested in the context of international law using the state as unit of analysis.[215] Applying experimental insights to individual decision makers whose acts are in turn attributed to the state (e.g., treaty negotiators, diplomats, or state officials) or to international judges is no major problem—the unit of analysis is the individual (as in most experiments). One limit of external validity is that most experiments are conducted with students. But experiments that are conducted with experts mostly show that experts exhibit similar deviations from rationality.[216] Applying experimental insights to the state directly is more problematic since aggregation problems arise. But rational choice theory faces the same criticism when applied to the state as such—reverting to the "unit of analysis" problem. The rationalist approach also needs to justify why and how states act rationally given that individual actors evidently show bounded rationality, since there is currently a disconnect between behavioral insights for individuals and states in the rationalist approach. Of course, material interests and strategic interaction remain of cardinal importance, as posited by rational choice theory. But psychological realities have been underappreciated even though its experiments yield more factors to consider—more, perhaps better, tools in the toolbox—that may enable sustained cooperation in the international realm.

[210] Robert Putnam, *Diplomacy and Domestic Politics: The Logic of Two-Level Games*, 42 INT'L ORG. 427 (1988). This is a fundamental paradigm explaining international trade treaties. Gene M. Grossman & Elhanan Helpman, *Protection for Sale*, 84 AM. ECON. REV. 833 (1994); Gene M. Grossman & Elhanan Helpman, *Trade Wars and Trade Talks*, 103 J. POL. ECON. 675 (1995). They also have shifted recently to a more behavioral approach. *See* Gene M. Grossman & Elhanan Helpman, *Identity Politics and Trade Policy*, REV. ECON. STUD. (forthcoming).

[211] Jan Schnellenbach & Christian Schubert, *Behavioral Political Economy: A Survey*, 40 EUR. J. POL. ECON. 395 (2015); Rick K. Wilson, *The Contribution of Behavioral Economics to Political Science*, 14 ANN. REV. POL. SCI. 201 (2011).

[212] A discussion on the value of experiments in international law is found in Jeffrey L. Dunoff & Mark A. Pollock, *Experimenting with International Law*, 28 EUR. J. INT'L L. 1317 (2017).

[213] *Id.* For a discussion in IR, see Robert Powell, *Research Bets and Behavioral IR*, 71 INT'L ORG. 265 (2017).

[214] Elinor Ostrom, *A Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address*, 92 AM. POL. SCI. REV. 1 (1998).

[215] Jean Galbraith, *Treaty Options: Towards a Behavioral Understanding of Treaty Design*, 53 VA. J. INT'L L 309 (2013) (who finds by an empirical study that framing of treaty options matters powerfully in ways inconsistent with rational choice theory, but consistent with insights from behavioral economics, based in this case on Prospect Theory).

[216] For some experiments conducted with experts, see, e.g., Susan D. Franck, Anne van Aaken, James Freda, Chris Guthrie & Jeffrey J. Rachlinski, *Inside the Arbitrator's Mind*, 66 EMORY L.J. 1115 (2017).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

### *B. The Difference in Perception*

Scholars of international relations have long understood that threats and sanctions in the international arena may generate feelings of hostility toward the source country. They may generate perceptions of "out-and-in-groups" and may result in stigmatizing effects, where the threatening part positions itself as complying with norms and stigmatizes the other country as the deviant.[217] The target's interpretation of the sender's intentions matters. Threats and sanctions are often exploited by the target government to generate an image of a hostile foreigner that holds malevolent intentions and is blamed for the economic difficulties faced by the receiving country. Positive inducements make it difficult for the regime to stigmatize the foreign country's behavior as hostile, and as a result undermines the regime's ability to mobilize support.[218] Rewards are also considered to be less confrontational as compared to penalties, and therefore lead to fewer problems related to sovereignty issues and interference in domestic politics.[219]

Perceptions may differ depending on who the sender is. Threats and promises coming from adversaries are most likely perceived differently than ones from an ally.[220] Rewarding in a relation considered as friendly is more likely to be perceived as appropriate and received with sympathy, while rewards in adversarial relations are probably perceived as suspect.[221] Penalties in close relations, e.g. relations between the United States and Canada, or France and Germany, might be considered as inappropriate. But also when dealing with adversaries, negative incentives might be less effective in extracting meaningful concessions.[222] Scholars have highlighted the expectation of conflicts: when a conflict is expected, concessions to a threat will only weaken the bargaining strength, but accepting a reward would strengthen the target's position; thus rewards are more likely to be accepted.[223]

Moreover, penalties may be perceived as unfair, e.g., when a country unsuccessfully tries to explain why it could not avoid breaching the agreement or when a penalty falls on low-income countries with capacity restrictions.[224] Penalties may also be perceived as illegitimate when they undermine sovereign rights of states. Rewards can help to reach international agreements to be perceived as fair. For instance, agreements that reward poor countries for their compliance may be perceived as more appropriate and fair than penalties.[225] In practice, the

[217] *See* references in note 4 *supra*. On stigmatizing effects of sanctions, see Alexandra Hofer, *The Efficacy of Targeted Sanctions in Enforcing Compliance with International Law*, 113 AJIL UNBOUND 163 (2019).

[218] Han Dorussen, *Mixing Carrots with Sticks: Evaluating the Effectiveness of Positive Incentives*, 38 J. PEACE RES. 251 (2001); WILLIAM LONG, ECONOMIC INCENTIVES AND BILATERAL COOPERATION (1996).

[219] Bernauer & Ruloff, *supra* note 106, at 21.

[220] DREZNER, *supra* note 39.

[221] *See* Milburn & Christie, *supra* note 4, at 631 ("The arms race could be seen as a perceptual dilemma in which each side professes to desire near parity of forces or disarmament but believes that the other side secretly harbors a motive for superiority.").

[222] Drezner, *supra* note 39, at 201.

[223] *Id.* (This leads to the following paradox: "In the case of incentives, receivers will be the most eager to accept a carrot when they anticipate frequent conflicts with the sender, which is precisely the situation where senders are the most reluctant to proffer the carrot. . . . Thus, senders will prefer to use sanctions over inducements against adversaries because they anticipate frequent conflicts, but those expectations also make sanctions less effective and inducements more so.").

[224] Bernauer & Ruloff, *supra* note 106, at 5.

[225] For an analogue's argument to national law, see Brian Galle, *The Tragedy of the Carrots: Economics & Politics in the Choice of Price Instruments*, 64 STAN. L. REV. 797, 817–18 (2012) ("Another important aspect of the distributive question is that sticks may be undesirable when they fall on households that are poorer than average.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

international climate policy, for example, tackles fairness considerations by relying on rewards in the form of funds and attaching a stronger burden on developed countries.[226] Scholars of international law have pointed out the importance of fairness perception to effectively implement agreed measures.[227] Rewards that support fairness perceptions can increase compliance.

### *C. The Difference in Response*

There are several reasons why responses to rewards and penalties may differ. One reason is linked to the emotions produced. Threats trigger negative emotions such as fear, anxiety, or anger, and cause a subject to feel stress.[228] Stress is supposed to reduce cognitive abilities of decision making and may result in irrational evaluations of the benefits and costs of compliance versus noncompliance.[229] Threats may also provoke a perception of conflict. People tend to take hawkish decisions in conflict situations, including those described by Prospect Theory.[230] The term "hawkish" denotes a propensity for suspicion, hostility, and aggression as well as for less cooperation and less trust in the resolution of the conflict.[231] Actors who are susceptible to hawkish biases are not only more likely to see threats as more severe than an objective observer would perceive them, but are also likely to act in a way that will lead to unnecessary conflict. Thus, the response may be to resort to noncompliant behavior.

Another reason why threats may be less effective than rewards is linked to psychological costs.[232] When threats are perceived as hostile or even as insulting by the leader (or by an audience that has some leverage over the leader, e.g., elites, citizens, foreign allies), noncompliance is motivated by the avoidance of looking weak (or to lose the approval of the group). Compliance under threats is then considered as damaging to a government's reputation of

Whether as a matter of efficiency or some other basis of social justice, we tend to want government programs to transfer wealth overall from those with more to those with less. When the opposite happens, the social benefits from curtailing negative externalities stand in tension with our preference for distributive fairness.").

[226] Lasse Ringius, Asbjørn Torvanger & Arild Underdal, *Burden Sharing and Fairness Principles in International Climate Policy*, 2 INT'L ENVTL. AGREEMENTS 1 (2002). *See also generally* Anne van Aaken, *Behavioral Aspects of the International Law of Global Public Goods and Common Pool Resources*, 112 AJIL 67 (2018).

[227] The importance of the perception of fairness has long been discussed in international law scholarship, see Thomas M. Franck, FAIRNESS IN INTERNATIONAL LAW AND INSTITUTIONS (1998). It has also been well researched, via empirical studies, in national law, see TOM R. TYLER, WHY PEOPLE OBEY THE LAW (1990), and has been experimentally researched, see Armin Falk, Ernst Fehr & Urs Fischbacher, *Testing Theories of Fairness—Intentions Matter*, 62 GAMES & ECON. BEHAV. 287 (2008).

[228] *See* PATCHEN, *supra* note 189, at 190–91.

[229] *Id. See also* ROBERT JERVIS, RICHARD NED LEBOW & JANICE GROSS STEIN, PSYCHOLOGY & DETERRENCE 5 (1985) ("Early critics of deterrence argued that fear would lead not to compliance but to rage, increased conflict, and miscalculation.").

[230] Kahneman & Tversky, *supra* note 12; Kahneman, et al., *supra* note 69.

[231] Jonathan Renshon & Daniel Kahneman: *Hawkish Biases and the Interdisciplinary Study of Conflict Decision-Making*, *in* AMERICAN FOREIGN POLICY AND THE POLITICS OF FEAR: THREAT INFLATION SINCE 9/11 (A. Trevor Thrall & Jane K. Cramer eds., 2009). For example, actors subject to hawkish biases are mostly also overconfident in being able to "win" the conflict, and they are in turn risk seeking. They also commit fundamental attribution errors (overemphasizing dispositional factors and downplay situational ones) and are under an "illusion of control" (which is an exaggerated perception of the extent to which outcomes depend on one's actions). In short, they tend to make more hawkish decisions in international conflict situations.

[232] *See* PATCHEN, *supra* note 189, at 180–85.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

firmness.[233] As penalties are almost always public (and they should be to achieve deterrence), it makes it much harder for the government, if it complies with demands, to maintain that it did so voluntarily.[234] The fear of losing face therefore can result in escalation.[235] Rewards can produce a more neutral setting, e.g., by highlighting mutual benefits.[236] Adding a reward might, for example, be very effective when a government considers cooperating due to the pressure from sanctions but would not do so for fear of humiliation. A substantial reward might allow a government to sell their eventual concession as a mutually beneficial deal.

Another reason for differences in responses is linked to reciprocity. Receivers may behave noncompliant to penalties and compliant to rewards because of reciprocity that calls for returning bad for bad as well as good for good.[237] As mentioned, reciprocity has long been known to be a crucial building block of international law.[238] Laboratory experiments show that a significant number of subjects are willing to reward cooperative behavior, referred to as "strong positive reciprocity," and to punish the uncooperative behavior of opponents, referred to as "strong negative reciprocity."[239] This finding even holds true in one-shot interactions where reciprocity is costly and does not maximize the tangible payoff. More recently, in an experiment, Chilton, Milner, and Tingley examined how reciprocity influences public opposition to foreign direct investment.[240] They showed that individuals care about rewarding or penalizing foreign countries for their policies. When a foreign firm's home country restricts investments from the respondents' country, the respondents are more likely to oppose potential transactions. Other empirical findings confirm that nations reciprocate each other's behavior.[241]

[233] *Id.* at 180–81 ("The target of a threat may defy the threatener not because the immediate tangible costs of compliance are too high but, rather, because he views compliance as humiliating or as damaging to his long-term relationships with adversaries by creating an impression of weakness under pressure.").

[234] *Id.*

[235] DAVIS, *supra* note 6, at 22.

[236] Milburn & Christie, *supra* note 4, at 633.

[237] PATCHEN, *supra* note 189, at 264 ("When leaders of one nation receive a reward or concession from another or a promise of such reward, they may believe that it is right and appropriate to reciprocate."). *See* also DAVIS, *supra* note 6, at 19 ("To the extent that promises of shared rewards promote the norm of reciprocity between actors, they hold greater prospect for transforming relations from conflictual to cooperative over a range of issues.").

[238] *See* note 33 *supra*.

[239] *See, e.g.*, Joyce Berg, John Dickhaut & Kevin McCabe, *Trust, Reciprocity, and Social History*, 10 GAMES & ECON. BEHAV. 122 (1995). However, there is also evidence for antisocial punishment where noncooperators punish cooperators. *See* Benedikt Herrmann, Christian Thöni & Simon Gächter, *Antisocial Punishment Across Societies*, 319 SCI. 1362 (2008). One example in international law is the punishment of International Criminal Court officials by the United States under the Trump administration. *See US Sanctions Against International Court Staff a "Direct Attack" on Judicial Independence*, UN NEWS (June 25, 2020), *at* https://news.un.org/en/story/2020/06/1067142.

[240] Adam S. Chilton, Helen V. Milner & Dustin Tingley, *Reciprocity and Public Opposition to Foreign Direct Investment*, 50 BRIT. J. POL. SCI. 129 (2020).

[241] *See* PATCHEN, *supra* note 189, at 262–63 (*citing* WILLIAM A. GAMSON & ANDRE MODIGLIANI, UNTANGLING THE COLD WAR: A STRATEGY FOR TESTING RIVAL THEORIES (1971) (who studied the interaction between the Western Allies and the Soviet Union from 1946 to 1963, showing that each of the sides was likely to respond in kind to conciliatory actions by the other); Russell J. Leng: *Influence Strategies and Interstate Conflict*, *in* THE CORRELATES OF WAR (J. D. Singer ed., 1980) (who studied the influence attempts in fourteen disputes between nations, finding that actions taken to increase the magnitude and credibility of threats tend to be associated with extreme responses by the target—either outright compliance or defiance in the form of counter-threats and penalties)).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

Penalties and rewards change individual perception of others' expected conduct.[242] Sanctioning increases the salience of unlawful behavior that may harm the general perception.[243] They might be seen as a cue that others are not cooperating and this can in turn trigger negative reciprocal behavior. Rewards incorporate an important signaling function that increases the perception of respecting law in the international arena. For instance, research in tax compliance has shown that the perception of other individual's tax compliance is crucial for the own tax compliance.[244] As rewards increase the salience of law-abiding behavior, they can encourage conditional cooperators to invest trust, complying if they expect enough others do so, too. In other words, individuals are willing to contribute, trusting or knowing that others are contributing as well. The expectation is that if there are enough players "in" and there is a reasonable expectation that other states will comply, those that are conditional cooperators and who are willing to invest trust will indeed cooperate. Most individuals are conditional cooperators and the same has been diagnosed for states in climate change law.[245] The illustration in IHL on positive reciprocity from above is another intriguing example.[246] Rewarding thus has a third-party effect, especially in multilateral treaties.

Yet another reason for differences in responses is linked to Prospect Theory. Prospect Theory revealed that people are very sensitive to changes in their endowment and that choice is driven by an overwhelming psychological desire to avoid loss.[247] In our analysis of compliance with international law, Prospect Theory leads to an important hypothesis: in the domain of loss,[248] rewarding is more effective than penalizing.[249]

Not all noncompliance is motivated by the desire to make gains. Sometimes it is fear that is driving behavior. If noncompliance is motivated by the fear of loss, threatening with more loss only enhances the motivation that gave rise to the problem in the first place. Actors in the domain of loss are risk seeking, thus, *ceteris paribus*, less sensitive to the risk associated with escalation.[250] For instance, not all states have an interest to comply with human rights (even when entering human rights treaties). If a leader fears losing political power when complying with human rights, threats of further losses (e.g., sanctions) may not have a deterrent effect.[251] Rather, the avoidance of losses may lead to higher degrees of repression. However,

[242] Dan M. Kahan, *The Logic of Reciprocity: Trust, Collective Action, and Law*, 102 MICH. L. REV. 71, 79 (2003) ("Incentives do more than affect individuals' calculations of the costs and benefits of particular forms of conduct; they also shape their impressions of the attitudes and intentions of those around them.").

[243] In analogy of the broken window theory; seminal: James Q. Wilson & George L. Kelling, *Broken Windows*, ATLANTIC (Mar. 1982), *at* https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465.

[244] *See, e.g.*, Donna Bobek, Robin Roberts & John Sweeney, *The Social Norms of Tax Compliance: Evidence from Australia, Singapore, and the United States*, 74 J. BUS. ETH. 49 (2007) (pointing out that own moral personal beliefs and beliefs about others close to them were significant in explaining tax compliance).

[245] Elinor Ostrom, Joanna Burger, Christopher B. Field, Richard B. Norgaard & David Policansky, *Revisiting the Commons: Local Lessons, Global Challenges*, 284 SCI. 278, 279 (1999); Mitchell, *supra* note 96 (for states).

[246] Text at notes 101–102 *supra*.

[247] *See* text at notes 12, 69 *supra*.

[248] *Id.*

[249] DAVIS, *supra* note 6, at 32–43.

[250] Text at notes 260–261 *infra*.

[251] We do not doubt the importance of penalties with respect to human rights violations, setting a signal of disapproval. Penalties can change political outcome by mobilizing collective action against governments. *See, e.g.*, Julia Grauvogel, Amanda A. Licht & Christian von Soest, *Sanctions and Signals: How International Sanction Threats Trigger Domestic Protest in Targeted Regimes*, 61 INT'L STUD. Q. 86 (2017). What we do is to

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

according to Prospect Theory, decision makers should be highly receptive to promised rewards when noncompliance is motivated by the fear of loss. Hafner-Burton shows that when compliance with human rights treaties is tied to tangible benefits (in the case examined, the benefits were captured via preferential trade agreements), it improved states' human rights records.[252]

### *D. The Difference in Stability*

The necessity to analyze the behavioral differences between rewards and penalties in the field of international law further arises from their effects on interstate relations and thus on political stability: "[P]romises can transform relations among adversaries in a way that threats cannot."[253] Rewards are more likely to please the receiver and tend to invite future cooperation. Threats are more likely to reduce the receiver's willingness to have any future contact with the penalizing state. As noted by Baldwin, "If A uses positive sanctions today, B will tend to be more willing to cooperate with A in the future, but if A uses negative sanctions today, B will tend to be less willing to cooperate with A in the future."[254] Rewards are likely to spill over to the target's willingness to cooperate on other issues while penalties are likely to impede such cooperation.[255] For instance, sanctions by the United States against Cuba, Iran, and North Korea since the Cold War era have also hindered their willingness to cooperate on other foreign policy issues.[256]

Rewards in the form of (economic) integration support cooperation and communication, while penalties lead to isolation. For instance, Hellquist argues that one possible reason why the EU favors sanctions abroad but not at home is that sanctions are instruments of exclusion and ostracism—a tool that does not fit at home where disagreements are resolved through dialogue.[257] Integration is considered to be essential in promoting compliance with international law.[258] There is strong evidence from experimental research that communication and personal contacts between players increases cooperation. A meta-analysis of social-dilemma experiments concludes that discussion has an extremely positive effect on subjects' willingness to cooperate.[259]

provide an additional explanation of why leaders may resist pressure (which can be linked to the domain of losses) when they violate human rights.

[252] Emilie M. Hafner-Burton, *Trading Human Rights: How Preferential Trade Agreements Influence Government Repression*, 59 INT'L ORG. 593, 623–24 (2005) ("Human rights regimes alone rarely create the conditions necessary for state compliance with human rights because they are almost always soft . . . material and political rewards are often a more effective (and compatible) incentive structure to support the initial stages of compliance.").

[253] DAVIS, *supra* note 6, at 19.

[254] Baldwin, *The Power of Positive Sanctions*, *supra* note 4, at 33 (This effect is also referred to as the scar effect, where today's use of positive inducements has a positive effect on future cooperation).

[255] *Id.* at 32.

[256] Peksen, *supra* note 149, at 285.

[257] Ellin Hellquist, *Ostracism and the EU's Contradictory Approach to Sanctions at Home and Abroad*, 25 CONTEMP. POL. 393 (2019).

[258] For instance, scholars show that international trade and foreign investment encourage government respect of human rights. *See, e.g.*, Emilie M. Hafner-Burton, *Right or Robust? The Sensitive Nature of Repression to Globalization*, 42 J. PEACE RES. 679 (2005); David L. Richards, Ronald D. Gelleny & David H. Sacko, *Money with a Mean Streak? Foreign Economic Penetration and Government Respect for Human Rights in Developing Countries*, 45 INT'L STUD. Q. 219 (2001).

[259] David Sally, *Conversation and Cooperation in Social Dilemmas: A Meta-analysis of Experiments from 1958 to 1992*, 7 RATIONALITY & SOC'Y 58 (1995).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

The tendency of penalties to spawn more penalties may escalate in conflicts, sometimes referred to as "conflict spirals."[260] What has been less of a focus is that rewards can produce de-escalatory behavior.[261] This insight is again linked to Prospect Theory. People generally place a higher value (usually twice as much) on what they stand to lose than on what they may gain of objective equivalent size. The use of penalties increases the value of winning because the receiver of penalties is more likely to face a loss with respect to the status quo. Fighting in a dispute which grows in magnitude, makes winning more important than it was initially. Each side becomes less willing to concede and more willing to suffer costs and take the risk of escalating the fight. In contrast, rewards reduce the value of winning in a dispute and the question of who will win becomes less salient. At the same time, the prospect of cooperation and mutual benefits increases in salience.

Another difference is built on trust. It was only relatively recently that IR scholars began to probe what trust really is, how it can be studied, and how it affects state relations, be it from the rationalist perspective[262] or from a more constructivist or psychological perspective.[263] Rewards are more likely than penalties to create trust, be it inter-personal trust[264] or strategic trust.[265] Intangible rewards, e.g. visits, social approvals, and praising, are important instruments to build up trust. There is an agreement in the literature that signals of uncooperative behavior or sanctioning do not help to develop trust.[266] Missing trust, in turn, affects the stability of international relations and impacts the behavior of conditional cooperators.

### *E. Summary*

Contrary to rational choice assumptions, rewards and penalties are not equally incentivizing means—they are not two sides of the same coin but are in fact two different currencies. Penalties and rewards communicate two different principles: the bad of breaking law versus the good of complying; disapproval versus approval; the willingness to punish noncooperative behavior versus the willingness to reward cooperate behavior. All other things being equal, these two principles do not lead to the same impact on behavior. Table 3 summarizes our main findings regarding the behavioral differences between rewards and sanctions.

## VI. When Will Rewards Work Best?

In this Part, we describe the limitations to rewarding as well as the conditions under which rewards (internal and external) are likely to be successful, taking into account also rewards' behavioral impact on a state's decision to comply.

[260] *See, e.g.*, Dean G. Pruitt, Jeffrey Z. Rubin, Social Conflict: Escalation, Stalemate, and Settlement 90 (1986). Thomas C. Schelling, Arms and Influence (1966) pointed out that those increases in coercion may not only increase in quantity but also in quality.

[261] *See* Patchen, *supra* note 189, at 270.

[262] Andrew H. Kydd, Trust and Mistrust in International Relations (2005).

[263] Brian C. Rathbun, *It Takes All Types: Social Psychology, Trust, and the International Relations Paradigm in Our Minds*, 1 Int'l Theory 345 (2009).

[264] Nicholas J. Wheeler, Trusting Enemies: Interpersonal Relationships in International Conflict (2018); Brian Rathbun, Trust in International Cooperation: International Security Institutions, Domestic Politics and American Multilateralism (2012).

[265] Kydd, *supra* note 262.

[266] Hiski Haukkala, Carina van de Wetering & Johanna Vuorelma, Trust in International Relations: Rationalist, Constructivist, and Psychological Approaches (2018).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

TABLE 3.
BEHAVIORAL DIFFERENCES BETWEEN PENALTIES AND REWARDS

| | Penalties | Rewards |
|---|---|---|
| Perception | Perceived more negative | Perceived more positive |
| Response | In order to avoid the penalty, threats can result in compliance. Negative emotions and reciprocity may call for counter-threats and resistances. Hawkish biases may aggravate conflict. | Rewards are more likely to be reciprocated by compliance. |
| Stability | Penalties may increase mutual dislike. Reciprocated in kind penalties may deepen conflicts. | Rewards (when reciprocated) are more likely to foster cooperation and trust. De-escalatory mechanism. |

### *A. Limitations to Rewarding*

Rewarding is no panacea; its limitations need to be carefully considered and they show that rewarding cannot be used indiscriminately in international law. But penalties are also no panacea and can only be used in certain circumstances.[267] We have already discussed some issues, like costs, when comparing rewards and penalties. Here, we focus on limitations and objections specific to rewarding.

#### *1. Moral Hazard and Adverse Selection*

The most important objection to rewarding is that rewards can cause moral hazard. Moral hazard describes a hidden action that results from a transaction and can occur when the party with more information about its actions or intentions has a tendency or incentive to behave inappropriately from the perspective of the party with less information. For example, the concept of moral hazard suggests that customers who have insurance may be more likely to behave recklessly than those who do not have insurance. Transposed to international law, the prospect of obtaining a reward when returning to lawful behavior may incentivize states to exhibit undesirable behavior in the first place. Especially in global collective action problems, states may seek to present themselves as potential targets in need of assistance in order to reap a reward.[268] Moral hazard can thus be a serious problem.

However, sending states have different means to reduce such problems, for instance, the joint financing of activities (including the target side) that also creates burdens for the target, stepwise cooperation, the use of hostages (in the political science sense, not literally), and the option of senders to withdraw rewards when moral hazard is suspected.[269] Sending states may also signal that the rewarding act is limited to the respective target and does not entitle all countries to a reward.[270] Rewards can be less prone to moral hazard when a previous sanction has been applied.[271] For instance, it would be difficult to argue that Russia should be

[267] *See* Section III.B *supra*.
[268] Ben-Shahar & Bradford, *supra* note 4, at 422.
[269] Bernauer & Ruloff, *supra* note 106, at 30.
[270] Ben-Shahar & Bradford, *supra* note 4, at 423.
[271] Previous negative incentives also make positive incentives appear more attractive to the potential recipient. *See id.* at 186.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

rewarded when handing Crimea back over to Ukraine, since this may give incentives to other states to annex foreign territory. But if it is rewarding in the form of redemption after outcasting (e.g., readmission of Russia to the G8), then the moral hazard problem is alleviated since a negative sanction was still handed out previously. Thus, in order to avoid moral hazard, in constellations of bad-faith violations of international law, it is necessary to introduce a negative sanction and only then use rewards.

Hold up problems are also commonly known in international negotiations. They are part of negotiating techniques, not specific to rewards. It is a situation where two or more parties may be able to work most efficiently by cooperating but refrain from doing so because of concerns that they may give the other party increased bargaining power. It also occurs in situations where rewards in the form of technical and financial assistance are part of the treaty. For example, in a number of initiatives coming from the 1992 Rio Summit, namely in negotiations of the MF associated with the Montreal Protocol, much resistance to the Fund, especially its administrative structure and funding and other forms of technical assistance, was voiced.[272] Countries in transition and South Africa wanted special exemptions from their contributions to the Fund as well as from meeting control targets of ozone-depleting substances.[273] In sum, rewards do not eliminate the hold up problem in negotiations, and they may exacerbate moral hazard in the compliance stage but problems can be mitigated as described above, e.g., via conditionality and reversed burden of proof monitoring.

### *2. Crowding-Out*

Another critique of rewarding is that it may crowd-out the receiving country's own motivation to comply with international law.[274] In the literature on development aid, this problem is captured by the Samaritan's dilemma.[275] In this concept, foreign aid decreases the recipient country's effort in development goals.[276] For instance, if the developing country knows that the donor country will provide food relief, the developing country may consider the relief as substitute for domestic food production and decrease its efforts to improve self-support.

Extrinsic monetary incentives may crowd-out intrinsic motivations.[277] This effect has been attributed to two psychological processes: impaired self-determination and impaired

[272] *See* text at note 47 *supra*.

[273] Ian H. Rowlands, *The Fourth Meeting of the Parties to the Montreal Protocol: Report and Reflection*, 35 ENV'T 25, 29–30 (1993).

[274] For a detailed discussion on crowding-out effects of carrots and sticks in national law, see Kirsten Underhill, *When Extrinsic Incentives Displace Intrinsic Motivation: Designing Legal Carrots and Sticks to Confront the Challenge of Motivational Crowding-Out*, 33 YALE J. REG. 213 (2016).

[275] The Samaritan's Dilemma was first introduced by James M. Buchanan: *The Samaritan's Dilemma*, *in* ALTRUISM, MORALITY, AND ECONOMIC THEORY (Edmund S. Phelps ed., 1975). Initial help by the Samaritan (the sender) causes motivation problems in decreasing incentives of the receiver. One main reason for a Samaritan's dilemma is the inability of the Samaritan to penalize the recipient because a penalty leads to further reductions in the recipient's welfare.

[276] However, comparisons of development aid to compliance theory should be taken with caution, because development aid is not provided with the primary aim of eliciting compliance by the recipient. Bernauer & Ruloff, *supra* note 106, at 25–26.

[277] *See, e.g.*, Bruno S. Frey & Reto Jegen, *Motivation Crowding Theory*, 15 J. ECON. SURV. 589 (2001). One is said to be intrinsically motivated when one receives no apparent reward except of the activity itself, e.g., unpaid voluntary work or giving money to charities. *See also* Fehr & Falk, *supra* note 116; Uri Gneezy & Aldo Rustichini,

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

self-esteem.[278] Rewards in form of development aid may act as one example of reduced self-determination in the international arena. While developed countries may subjectively define development aid as rewarding, developing countries may perceive it as controlling and dependency enhancing. This dependency increases as receiving countries become subject to threats of permanent aid cutoffs. Nevertheless, external interventions crowd-in intrinsic motivation if it is perceived as supportive.

Intangible rewards, in contrast, such as praising or approval have been proven to crowd-in motivation.[279] In that case, self-esteem is fostered and individuals feel they are given more freedom to act, thus enlarging self-determination. If these insights can be translated to the international arena, social approval, recognition, praising, and inclusion could foster motivation to comply. This form of crowding-in effect is important to the analysis of international law, if it is assumed that governments comply also out of intrinsic motivation.

### *B. Conditions Conducive to Rewarding*

Several conditions are conducive for the effectiveness of rewarding.[280] First, for a reward to be effective, it needs to match with the receiver's perception of what is conceived as rewarding and what is valued.[281] While the giver may subjectively define the action as rewarding, the receiver may not: "A may perceive himself as employing carrots, while B may perceive A as using sticks."[282] Promises are simply ineffective in securing concessions if the rewards promised are perceived as inappropriate or even insulting. Rewards are then perceived as coercive and operate similar to penalties. With coercive rewards, the psychological cost of giving in increases.[283] Rewards may be considered imposing and seen as undermining the integrity of the state/political community, especially if they are conditioned, e.g., EU foreign assistance and aid in some states such as Turkey. Therefore, the value of a reward will depend on the receiver's need and on the objectives it pursues. While for some national leaders tangible rewards are of great value, e.g., grants or loans of money to their nation, for others the

*Pay Enough or Don't Pay at All*, 115 Q. J. ECON. 791 (2000) (concluding that monetary rewards have a negative effect on intrinsic motivation. But if subjects were paid a fixed positive amount, independent of their performance, intrinsic motivation was not reduced.).

[278] Frey & Jegen, *supra* note 277 (e.g., when an external intervention carries the notion that the actor's motivation is not acknowledged).

[279] *See* BRUNO S. FREY & JANA GALLUS, HONOURS VERSUS MONEY. THE ECONOMICS OF AWARDS (2017) (pointing out that in contrast to tangible incentives, intangible rewards support intrinsic motivations as they enhance feelings of relatedness and social recognition). *See also* Robert Eisenberger & Judy Cameron, *Detrimental Effects of Rewards: Reality or Myth?*, 51 AM. PSYCHOLOGIST 1153, 1162 (1996) (who provide a meta-study evaluating more than a quarter century of research on the effect of rewards on motivation, concluding that "our analysis of a quarter century of accumulated research provides little evidence that reward reduces intrinsic task interest"; the authors also highlight the positive effect of verbal rewards).

[280] This section is partially based on PATCHEN, *supra* note 189, at 262–71.

[281] *See* Milburn & Christie, *supra* note 4, at 635–36.

[282] Baldwin, *The Power of Positive Sanctions, supra* note 4, at 24.

[283] PATCHEN, *supra* note 189, at 265 ("Offers of reward in return for concessions may sometimes be seen as coercive or demeaning. . . . In such cases, making concessions to the other side may involve some loss of personal and/or national status. Reluctance to accept such a humiliation was probably one factor that influenced North Vietnam to refuse to give up its military efforts in return for money from the United States, since the American offer was combined with threats and coercion. To do so would have been seen by the Vietnamese as a serious blow to their self-esteem and their esteem in the eyes of others.").

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

need for symbols of high status are more important, e.g., social approval, inclusion in international conferences, and invitations to state visits.

Furthermore, the clearer and more specific a promised reward is, the more it is conducive to compliance. Leng showed that national leaders were more likely to respond with compliance to clearly specified, rather than unspecified, promises.[284] In the same line, Snyder and Diesing found that clear, explicit offers of concessions were more likely than vague offers to facilitate settlements of serious disputes between nations.[285] Deutsch suggests that vagueness about the rewards may lead the target to conclude that the promisor has little power to deliver.[286] Thus, the explicit communication of reward contingency is essential for its success; this argues for explicit internal rewards in treaty design.

The timing of delivery is important as well. Rewards are expected to be more successful if the promised reward is timely delivered after compliance by the receiver. As Deutsch states, promises that are to be fulfilled far in the future will have a lesser effect on compliance than promises of rewards closer in time.[287] This holds even more for short-sighted leaders who value the present more than the future.[288] One example that underlines the importance of timely delivery of rewards is the revelation of North Korea's secret uranium enrichment program seven years after the so-called Agreed Framework was signed between the United States and North Korea on October 21, 1994. North Korea complained that the United States fell short in fulfilling its promises by failing to lift economic sanctions and failing to provide it with promised light-water reactors.[289] Invitations to bid for construction of reactors were not issued before 1998, and the construction of the first reactor began only in 2002.[290] North Korea's revelation was allegedly based on a failure by the United States to deliver on its promises.[291]

Furthermore, to be effective, a reward must be credible. The target should perceive that the promise is within the enforcer's control.[292] Promises, like threats, according to Deutsch, will be more credible when the rewarder is perceived as determined to influence and has the capability to implement the promise.[293] Rewards that are seen as excessive may lack credibility.[294] Thus, any promises made should be clearly within the rewarder's capability to fulfill and should not be beyond budget capacity. Public statements as commitment devices can help increase the credibility of promises just as explicit provisions for trust funds in treaties. The credibility of promises will be affected by past behavior of the rewarder as well, i.e.,

[284] Leng, *supra* note 241.

[285] Glenn Herald Snyder & Paul Diesing, Conflict Among Nations: Bargaining, Decision Making, and System Structure in International Crises (1977).

[286] Morton Deutsch, The Resolution of Conflict: Constructive and Destructive Processes (1973).

[287] *Id.*

[288] Galle, *supra* note 5, at 43.

[289] Miroslav Nincic, *The Logic of Positive Engagement: Dealing with Renegade Regimes*, 7 Int'l Stud. 321, 336 (2006).

[290] Miroslav Nincic, *Getting What You Want: Positive Inducements in International Relations*, 35 Int'l Sec. 138, 153 (2010).

[291] *Id.*

[292] Davis, *supra* note 6, at 11.

[293] Deutsch, *supra* note 286.

[294] However, according to Baldwin, *The Power of Positive Sanction*, *supra* note 4, at 28–29, promises are more likely to be scaled down, while threats are likely to be scaled up.

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

how well it has kept its past promises. Establishing a record of not exploiting others and of keeping past promises makes current promises more credible.[295]

Rewards may also be more effective toward countries that already suffer from substantial penalties.[296] The marginal impact of an added penalty diminishes with increasing penalties and countries' sensitivity toward penalties decreases. Lastly, the effectiveness of rewards is influenced by the value the receiver attaches to future cooperation with the rewarder.[297] Considerable economic or political interdependencies increase the value the receiver puts on future cooperation and thus the more likely a positive response will be. This also applies to relations perceived as generally friendly. By complying, the receiver will be viewed more favorably, while rejecting an offer of a reward would make the receiving state be perceived less favorably. Whether a leader will comply in response to a reward thus depends on the advantage of maintaining the goodwill of the rewarding state.

Thus, rewards work best, when: (1) the reward matches with the receiver's perception of what a reward is; (2) the receiver values the reward; (3) the receiver depends on the rewarder to provide the reward; (4) the reward is not combined with coercive measures that would cause increasing psychological costs for accepting the reward (unless to deter moral hazard); (5) the reward is clearly specified; (6) the reward will be timely delivered; (7) the reward is credible; (8) the rewarder has the capability to implement the reward; (9) the rewarder has a record of holding promises; (10) the reward follows substantial penalties in grave cases of violation; and (11) the receiver appreciates the cooperation with the rewarder and depends on its goodwill.

## VII. Conclusion

Rewarding is an important mechanism for compliance with international law. Although the problem of compliance has loomed large in the IR literature, IR scholars have largely focused on penalties (sticks). Only recently have they started to discuss rewards (carrots) as well as the interaction between the two. Mainly, they focus on security constellations and stay on the diplomatic policy level. The rationalist international law scholarship on compliance has also been more focused on penalties, and rewarding has been undertheorized in that discussion. In fact, we are not aware of one international law article dealing with rewarding in international law more generally, let alone elaborating a typology.[298] This is surprising given that rewarding is, as we have shown, inherent in compliance mechanisms like reciprocity, reputation, and outcasting. By defining and elaborating the "classical" mechanisms more precisely and providing a toolbox of options, we hope to illuminate the compliance discussion, and, with it, the rational design literature as well. External penalties, like countermeasures and retorsions (retaliation), have advantages and disadvantages that differ from internal penalties and can be employed in other constellations. The same applies for internal and external

[295] *See* Section VI.D (rewards can be used to generate a reputation of good will).

[296] *See* Ben-Shahar & Bradford, *supra* note 4, at 427 ("Sanctions are similarly ineffective in situations where the Sender has already employed them unsuccessfully against a Target in the past and where further sanctions can only inflict marginal additional pain on the Target. . . . In such situations, the promise of lifting existing sanctions is often the most attractive reward for the Target."). In contrast, because rewards increase the target's wealth they can lead to a saturation effect: at some wealth level a reward does no longer incentivize the target. *See* Dari-Mattiacci & De Geest, *supra* note 145, at 440. In that case, a penalty may be more efficient.

[297] Patchen, *supra* note 189, at 269.

[298] *But see* Ben-Shahar & Bradford, *supra* note 4 (focusing more narrowly on reverse rewards).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press

rewards. Under a rational choice approach, the selection between rewarding and penalizing is made according to the benefits and costs of each means. Their differences, even if both means are considered to be symmetric, are manifold. Even though rewards can be costly when applied to many countries, rewards in general are pareto efficient and bring about mutual benefits. Rewards can generate a reputation of goodwill that can increase cooperation. Monitoring and fact-finding, an approach that is gaining ground in international law, could be more successful when using rewards rather than penalties.[299] Penalties, when applied frequently, also create substantial costs to the penalizing country. However, when a reputation is built up to penalize countries not complying with their commitments, it can deter countries from misbehavior.

The insights we provide into compliance mechanisms help not only to better understand the mechanisms as already discussed in the literature but also shift the focus from a penalty-oriented system to governance mechanisms between states. A focus on rewards can reframe the compliance debate toward positive inducements that have often been overlooked. We illuminate where rewards are already used (but were not discussed as such) and can be used in a more targeted fashion. Rewards can be used not only in international diplomacy (which may end up with treaties) but in treaties more generally. They are thus of practical relevance. They can be used in bilateral as well as in multilateral constellations. Rewarding can be applied in treaties dealing with global public goods and commons just as in reciprocal treaties. They can also be used in soft law. We also show the limits to rewarding and the conditions under which rewards can best be used.

Moreover, theorizing only from a rationalist perspective, in which rewards and penalties are two sides of the same coin, may overlook important differences between penalty and reward. Psychological literature (and IR scholarship) has long emphasized these differences and has reported empirical evidence on the individual and the state level. Psychological insights show that rewarding and penalizing differ significantly in their effect on an individual's behavior. Laboratory experiments show that, even when transaction costs are low, the Coasean equivalence (assuming that penalties and rewards lead to the same result) does not hold true in reality. The behavioral analysis is therefore an important addition to the use of rewards in international law. The behavioral perspective allows evaluating penalties and rewards from a perception, a reaction, and a stability perspective. This leads to additional arguments why theorizing rewarding can fill a gap in the literature and in practice.

Although their costliness and ineffectiveness have been thoroughly discussed, penalties remain at the forefront of academic discussions and policy. With this Article, we submit that compliance theory needs reframing in order to realize the array of positive inducements already existing and their potential in international law. The framework we elaborate can be used for more doctrinal research on specific treaties or issue areas of international law as well as comparatively between them *de lege lata* and *de lege ferenda*.

[299] *See* Symposium: International Commissions of Inquiry, 30 EUR. J. INT'L L. (2019).

https://doi.org/10.1017/ajil.2021.2 Published online by Cambridge University Press